UNIVERSITÀ DEGLI STUDI DI VERONA

DEPARTMENT OF BIOTECHNOLOGY

GRADUATE SCHOOL OF SCIENCE,

ENGINEERING AND MEDICINE

DOCTORAL PROGRAM IN APPLIED BIOTECHNOLOGIES

Cycle XXVII / 2012

WITH THE FINANCIAL CONTRIBUTION OF

CARIVERONA Bank Foundation and the Valorizzazione dei Principali Vitigni Autoctoni Italiani e dei loro Terroir (Vigneto) project funded by the Italian Ministry of Agricultural and Forestry Policies

BERRY TRANSCRIPTOME COMPARISON OF TEN ITALIAN GRAPEVINE VARIETIES

S.S.D. AGR/07

Coordinator: Prof.ssa Paola Dominici

Tutor: Prof. Mario Pezzotti

Doctoral Student: Mélanie Massonnet

ABSTRACT

Grape berry development can be described as a succession of physiological and biochemical changes reflecting the transcriptional modulation of many genes. In the last decade, many transcriptomic studies have been carried out to deeper describe this dynamic and complex development. Nonetheless, most of those transcriptomic studies focused on one single variety at a time and then there is still a lack of resources for comparing berry development in different grape varieties.

This thesis describes the first berry transcriptome comparison carried out by RNA sequencing of 120 RNA samples, corresponding to 10-variety berries collected at four phenological growth stages, two pre- and two post-véraison, in biological triplication.

This RNA-Seq analysis showed an evident deep green-to-maturation transcriptome shift occurring at véraison independently on skin colour and variety, which involves the suppression of diverse metabolic processes related to vegetative growth, and the induction of only a few pathways, such as secondary metabolic processes and responses to biotic stimuli. This fundamental transcriptome reprogramming during ripening was highlighted by distinct approaches: Pearson's correlation distance, PCA, O2PLS-DA, biomarker discovery, clustering analysis and correlation network method. The establishment of the first grape berry development transcriptomic route, corresponding to the genes having similar patterns of expression during whole development independently on the variety, allowed identifying genes involved in the main biological processes occurring during berry development. Finally, the expression of phenylpropanoid/flavonoid biosynthetic pathway-related genes was found to be insufficient by itself to explain the differences between red- and white-grape transcriptomes, however it was supposed to influence – supposedly by the effect of anthocyanins accumulation in berry skin since the onset of ripening – maturation-phase transcriptional program, determining the recruitment of genes belonging to other biological processes.

TABLE OF CONTENTS

Chapter 1

Introduction and outline of the thesis

INTRODUCTION

Worldwide wine production, economic relevance and current objectives of research

Grapevine (*Vitis vinifera* L.) is a woody perennial plant. Worldwide, grapevine is one of the most widely cultivated fruit crop, encompassing 7.5 million hectares of arable land, and representing 75.1 million metric tons of grape berries in 2013 (http://www.oiv.int/). Grape berries are produced commercially for fresh fruit, juice, raisins but the most part of harvested berries are fermented into wine, producing almost 287 million hectolitres in 2013. The economic impact of this beverage is far greater than the value of the grapes. For example, world wine export generated total revenues of 25.7 billion euros in 2013 allowed by growing prices. For the last two decades, the strong economic competitiveness has driven the winemakers to new objectives: optimizing grape production and wine quality.

Winemaking techniques and equipment improved a lot in the last decades, however wine quality still remains predominantly correlated to berry quality, i.e. its maturity and composition, at harvest[1]. Berry quality can be managed in function of the cultivation site by different manners such as the selection of the cultivars (grafts and rootstocks) and viticulture practices. A considerable amount of viticulture research has identified strategies that can be used to optimize grape maturity at harvest including irrigation[2-4], canopy management[5-10], and cropping levels[11-13], among others[14-17]; without fully understanding how these strategies work at a molecular level. Moreover, winegrowers have to cope with other factors that could dramatically affect grape quality and yield, such as biotic stresses as downy and powdery mildew, berry rot, etc., which need more study to obtain a complete understanding.

Consequently, the economic relevance of grapevine and the need to resolve field problems have encouraged the scientific community to study the processes that influence grape yield and quality, and more particularly to unravel the physiological, biochemical and molecular basis of berry development. According to the current knowledge, these profound changes reflect the expression and modulation of ~23,000 genes; nevertheless much remains unknown about these transcriptional changes and their regulation. Thanks to technological advances of the transcriptomic analysis tools, significant progresses have been done for the last two decades.

Transcriptomics background of grape berry development

Transcription is the first step of gene expression, in which a segment of deoxyribonucleic acid (DNA) is copied into ribonucleic acid (RNA) by the enzyme RNA polymerase. The transcriptome is the complete set of RNA molecules produced by a cell, a tissue or an organism including messenger RNAs (mRNAs), ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) and other non-coding RNAs such as micro RNAs. Generally, the mRNA profile is the most studied because it corresponds to the expression of protein-encoding genes. The transcriptome depends on the genes that are expressed and their corresponding expression level in function of the cell type and the developmental stage, but also in response to external conditions. The dynamic nature of the transcriptome is highly informative because it indicates the modulation patterns of the expressed genes throughout tissue development.

The first large-scale studies of gene expression in *Vitis vinifera* (2n = 38) were performed using low-throughput Sanger sequencing approaches, which were expensive and often did not provide quantitative data. This census sequencing of cDNA libraries generated collections of expressed sequence tags (ESTs) or short markers that were joined together in techniques such as serial analysis of gene expression (SAGE). For example, Davies and Robinson[18] carried out the differential screening of a Shiraz post-véraison berry cDNA library by Northen analysis using probes produced from pre-véraison and post-véraison RNA samples. Terrier and colleagues[19] analysed another Shiraz library comprising 275 ESTs produced from three developmental stages (green, softening and ripening berries). Ablett and co-workers[20] analysed 2,479 ESTs from a pré-veraison Chardonnay cDNA library and a collection of 1,743 ESTs Pinot Noir berries from the véraison phase was screened by Moser *et al.*[21]. During the same period, a gel-based amplified fragment length polymorphism (cDNA-AFLP) analysis method was developed to characterize gene expression in the absence of prior sequence information. This method was used to profile transcripts in Chardonnay berries at six developmental stages[22] and to find differences between immature and mature Clairette Blanche and Cabernet Sauvignon berries[23]. Moreover, the same technique was used to look at gene expression during Corvina berry development and withering either on or off the vine[24].

As the quantity and availability of grapevine sequence information increased, it became possible to carry out gene expression analyses *in silico*[25,26] allowing the

production of cDNA-based microarrays and oligonucleotide arrays based on these sequences, leading to hybridization-based gene expression profiling during grape berry development.

The first large-scale hybridization experiments used microarrays prepared from the sequences in cDNA libraries. First, Terrier and colleagues[27] analysed changes in gene expression in Shiraz berries during nine developmental stages using an array of 50-mers representing 3,175 grapevine unigenes. The same array was later used by Glissant *et al.*[28] to analyse changes in gene expression in Chardonnay berries as they softened during ripening. Furthermore, a cDNA array comprising 4,608 PCR-amplified berry clones, derived from a ripening Shiraz berry cDNA library, was used to monitor changes in gene expression in the berry skin during ripening[29]. Finally, a cDNA array containing 4,224 clones from a subtraction library was used to investigate differences in gene expression between two distinct pools of berries belonging to the high and low resveratrol producers, respectively eleven and ten varieties, at three berry developmental stages (véraison, ripening and post-ripening)[30].

In 2004, the first commercial high-density oligonucleotide array for grapevine was released in the form of the Affymerix GeneChip® Vitis vinifera Genome Array. This array was developed by selecting sequences from GenBank®, dbEST and RefSeq, and using sequence clusters that were created from the UniGene database (build October 7th 2003). With this platform, it was possible to monitor the expression of 14,000 *Vitis vinifera* transcripts and 1,700 transcripts from other Vitis species. Each transcript was represented by a probe set of 16 overlapping 25-mers including perfect match and mismatch sequences. Several reports were published with transcriptome data obtained using the Gene Chip® Vitis vinifera Genome Array. These included studies of developmental profiles of Cabernet Sauvignon[31] and Pinot Noir berries[32] and tissue-specific expression in mature Cabernet Sauvignon berries[33]. The array was also used to study the effect of environmental stresses, as heat[34] and water stress[35], and to investigate transcriptional changes in ripening Cabernet Sauvignon berry skin in response to abscisic acid treatment[36].

In 2007, the deciphering and availability of the complete grapevine genome sequence[37,38] enabled for the first time the profiling of grape berry developmental gene expression at the whole-genome level with initially the creation of many microarray platforms. First, the GrapeArray 1.2, consisting of 35–40-mer probes representing 25,471 transcripts based on TIGR Grape Gene Index (Release 5.0) and the 8.4-fold

draft grapevine genome sequence[37] was created using the CombiMatrix electrochemical oligonucleotide array assembly platform. GrapeArray 1.2 creation permitted to combine gene expression profiling during Corvina berry development and postharvest withering[39] with proteome and metabolome data sets from the same berry samples. From this combination, Zamboni and co-workers identified putative stage-specific biomarkers and described development and withering of Corvina berries using both hypothesis-free and hypothesis-driven systems biology approaches. Then, the NimbleGen 12x135K array was developed using the 12.0-fold assembly gene prediction V1[40] based on Combimatrix (GrapeArray3.0) and NimbleGen (090918_Vitus_exp_HX12) technologies. The NimbleGen 12x135K array enables the hybridisation of up to twelve independent samples on a single slide, sensibly reducing the costs of the experiment. Each of the twelve sub-arrays contains 135,000 60-mer oligonucleotide probes and allows detecting the expression of 29,549 grapevine transcripts. This array was used to develop the transcriptome atlas of the grapevine cv. Corvina from the gene expression profiling of 54 samples representing green and woody tissues and organs at different developmental stages as well as specialized tissues such as pollen and senescent leaves[41]. This study revealed the existence of a deep transcriptome shift driving the entire plant into a maturation program. Furthermore, Dal Santo and colleagues[42] used also the NimbleGen microarray platform to map the Corvina berry transcriptome and determine which genes are plastic (modulated in response to different environments) and which are non-plastic (regulated in the same manner regardless of the environment).

The recent technological advances in next-generation sequencing methods[43,44] have been applied to transcriptomics creating the novel high-throughput technology named RNA sequencing[45] (RNA-Seq). For the first time, a transcriptomic approach has provided absolute quantitative expression analysis results by mapping short cDNA sequences, called reads, onto the reference genome and counting the number of reads each gene is represented in a sample. In addition to the advantage to obtain quantitative expression, RNA-Seq shows many other benefits in comparison to microarray technology. Firstly, due to the static nature of microarray-based expression data, expression can only be determined for gene models included on the array. For genome sequences such as grapevine that are evolving in terms of both gene content and gene model structural annotation, the use of microarrays could result in missing expression information for a substantial number for genes not available or annotated

at the time of array design. Secondly, reliance of microarray technology on DNA-DNA hybridization can potentially lead to inaccurate expression estimates for genes sharing high sequence homology. Thirdly, due to background noise and signal saturation, the dynamic range of expression obtained from microarrays is limited and, therefore, may not be suitable for comparing very highly or lowly expressed genes. Finally, RNA-Seq has the potential to detect novel transcribed loci on annotated genomes[46], splicing variants[47], expressed single nucleotide polymorphisms[48] and allele-specific events[49], massively increasing the investigative capability over these molecular phenomena.

In the case of grapevine, the first use of RNA-Seq assay was performed to gain insights into the wide range of transcriptional responses that are associated with berry development in *Vitis vinifera* cv Corvina[50]. More than 59 million sequence reads, 36 to 44 bp in length, were generated from three developmental stages: post setting, véraison, and ripening using an Illumina genome analyzer II. With the same aim, another transcriptomic study was performed from pooled Shiraz berries sampled at four developmental stages: 3-week post-anthesis, 10- and 11-week post-anthesis, corresponding to early and late véraison, and at harvest, using paired-end reads of 100 bp generated by Illumina HiSeq 2000 sequencer[51]. In 2014, Degu and co-workers[52] carried out the skin transcriptome comparison of Cabernet sauvignon and Shiraz berries sampled at véraison and 20 days after véraison using the Illumina HiSeq 1000 sequencer. Finally, RNA-Seq approach was also used to reconstruct *Vitis vinifera* cv Corvina[53] and Tannat[54] transcriptomes by *de novo* assembly. This method permitted to identify varietal genes, i.e. genes that were not shared with the reference genome, mainly involved in the synthesis of phenolic and polyphenolic compounds in the case of Tannat green berries.

In conclusion, many transcriptomic studies have been carried out to shed light on the dynamic and complex development of grape berry for the last decade. The most part of these studies were realized using microarray platforms that only allow an indirect measurement of gene expression[31-33,41,42,55]. On the other hand, since 2010, the application of high-throughput sequencing technology on gene expression studies – namely, RNA sequencing – has arisen a new method to obtain quantitative transcriptomic results[50-52,54]. Nevertheless, most of these transcriptomic studies focused on a single variety at different developmental stages and determining any comparison between varieties very difficult to make. Furthermore, two of the few RNA-Seq studies[50,51] were performed with no biological replication; although considered

methodological breakthrough[56], those works could not significantly contribute to berry development knowledge due to the lack of statistical support for new biological findings.

Aim and outline of the thesis

This thesis describes the first berry transcriptome comparison carried out on ten cultivars: five red-skinned – Sangiovese, Barbera, Negro amaro, Refosco, Primitivo – and five white-skinned grape varieties – Vermentino, Garganega, Glera, Moscato bianco and Passerina – whose main features are briefly described in **chapter 2**. The novelty of this research project lies in performing the first RNA sequencing of 120 berry samples, corresponding to the 10-variety berries collected at four phenological growth stages in biological triplication, allowing obtaining more than 400 Gb of sequencing information (**chapter 3**). The aim of this project was to unravel the transcriptomic traits that were either common in all grape varieties – to increase understanding of gene functions in the general grape berry biological processes – or specific to each skin colour group – in order to understand the physiological and biochemical differences between red and white grape varieties at transcriptomic level – or specific to each variety – to define the transcriptional diversity between Italian grape varieties.

Chapter 4 reports the first steps of transcriptomic analysis, which consisted in the analysis of the gene expression of the 10-variety berries throughout development. From the alignment of the reads, generated by RNA-Seq assay, onto the reference genome PN40024 (12x)[37], 26,807 genes were detected as expressed, representing ~89.4% of the V1 predicted grapevine genes[40] (http://genomes.cribi.unipd.it/DATA). In order to handle this large data set and to avoid artificially high fold-change value discovery between samples, a cut-off of 1 FPKM (Fragments Per Kilobase of Mapped reads) was established. The 5,061 genes, exhibiting an expression magnitude below 1 FPKM among all berry samples, were removed from the data set and named as very low-expressed genes, described at a later stage in **chapter 10**. Then, differentially expressed genes (DEGs) between each subsequent pair of developmental stages were identified, allowing the study of the transcriptional modulation of each grape variety throughout berry development.

Different statistical methods were used to analyse the transcriptome data. Pearson's correlation distance approach and Principal Component Analysis (PCA)

allowed the identification of transcriptional relationships among samples – these findings are reported in **chapter 4** and **5** respectively.

In order to discover the transcriptomic traits that are common in all grape varieties, a comparison of the DEGs was performed in **chapter 6**, unravelling the commonly up- and down-regulated genes between each subsequent pair of growth stage. The analysis of those genes permitted to highlight the main biological processes occurring during berry development and to identify the genes involved in those biological functions. Putative biomarkers specific to either the developmental phase or the growth stage, independently of the skin colour, were discovered using an O2PLS-DA approach[39,57], as described in **chapter 5**. In **chapter 7**, a coupled clustering analysis was performed in order to establish the first grape berry development transcriptomic route, corresponding to the genes having similar patterns of expression during whole berry development independently of the variety. This route is composed of 993 genes and interestingly the most of the biomarker transcripts discovered in **chapter 5** were comprised in this analysis, confirming their involvement in the general program of berry development. This clustering analysis allowed the identification of sets of genes that exhibit the same expression profile in all the ten varieties and represent biological processes occurring at different stages or specifically during pre- or post-véraison phase.

At harvest time, the most evident phenotypic difference between berries is the skin colour. Red skin colour is caused by the biosynthesis of anthocyanins in red grapes, which is regulated by the additive activity of two MybA-type transcription factors: VvMybA1 and VvMybA2. In white cultivars, the insertion of the *Gret1* Gypsy-type retrotransposon into the promoter of *VvMybA1* prevents *VvMybA1* transcription[58], while a mutation of *VvMybA2* results in a double STOP codon responsible for the truncation of MybA2 protein, rendering it inactive[59]. In order to find out transcriptomic traits specific to each skin colour group – other than the well-known up-regulation of phenylpropanoid biosynthesis pathway-related genes after véraison – few genes were identified as only expressed in either red or white grapes throughout development in **chapter 4**. In **chapter 6**, the comparison of the DEGs allowed also the discovery of genes that were statistically up- and down-regulated only in red- or white-skinned berries. **Chapter 8** describes all the steps climbed up to understand differences of transcriptome patterns and behaviours between red and white grapes throughout maturation phase. To do that, several PCAs were performed and t-test analyses were

done in order to identify differentially expressed genes between red- and white-skinned berries at the two ripening stages. In addition, another approach, consisting in the extraction of transcript loadings from PCA outcome and the detection of the loading threshold-breaking point, was used to select genes involved in berry transcriptome differences at maturity. A deep analysis of these genes showed that phenylpropanoid biosynthesis pathway-related gene expression is not sufficient to explain the differences between red- and white-grape transcriptomes, and that many other biological processes are implicated in this phenomenon. Hence, the hypothesis that anthocyanin content could influence berry transcriptome over maturation and determine specific transcriptional trends that distinguish red and white grapes.

Chapter 9 reports the first exhaustive RNA-profiling of the phenylpropanoid/flavonoid biosynthesis pathway-related genes in ten varieties, allowing correlating each gene related to phenolic compound biosynthesis with the already-known biosynthesis timing, and identifying phenylpropanoid and flavonoid biosynthesis pathway-related genes participating in the downstream pathways.

As a part of this thesis, the red-skinned berry transcriptomes – namely a half of the whole dataset – were employed by Palumbo and colleagues[60] to develop a novel approach that integrates different network-based methods to analyse the correlation network arising from large-scale gene expression data (**chapter 11**). By studying grapevine (*Vitis vinifera*) and tomato (*Solanum lycopersicum*) gene expression atlases and the red-grape transcriptomic data set during the transition from immature to mature growth, a category named “fight-club hubs” was characterized by a marked negative correlation with the expression profiles of neighbouring genes in the network. A special subset named “switch genes” was identified, with the additional property of many significant negative correlations outside their own group in the network. Switch genes were found involved in multiple processes and include transcription factors that may be considered master regulators of the previously reported transcriptome remodelling that marks the developmental shift from immature to mature growth. All switch genes, expressed at low levels in vegetative/green tissues, showed a significant increase in mature/woody organs, suggesting a potential regulatory role during the developmental transition.

Concluding remarks of the entire thesis are comprised in **chapter 12**, where a deeper discussion on the novelty of this work is reported. The data provided by the

RNA-Seq assay of 120 berry samples provides a useful, data-rich resource for the scientific community.

REFERENCES

1. Bindon K., Holt H., Williamson P.O., Varela C., Herderich M., Francis I.L. Relationships between harvest time and wine composition in *Vitis vinifera* L. cv. Cabernet Sauvignon 2. Wine sensory properties and consumer preference. Food Chem. 154:90-101 (2014)
2. Bravdo B. & Hepner Y. "Water management and the effect on fruit quality in grapevines", Proceedings of the 6th Australian Wine Industry Technical Conference pp. 150-158 (1986)
3. Hardie W.J., Considine J.A. Response of grapes to water-deficit stress in particular stages of development. Am. J. Enol Viticul. 27:55-61 (1976)
4. Matthews M.A. & Anderson M.M. Reproductive development in grape (*Vitis vinifera* L.), responses to seasonal water deficits. Am. J. Enol Viticul. 40:52-60 (1987)
5. Crippen D.D. & Morrison J.C. The effects of sun exposure on the phenolic content of Cabernet Sauvignon berries during development. Am. J. Enol Viticul. 37:243-246 (1986)
6. Dokoozlian N.K. & Kliewer W.M. Influence of light on grape berry growth and development varies during fruit development. Journal of the American Society of Horticultural Science 121:869-874 (1996)
7. Hunter J.J. & Visser J.H. The effect of partial defoliation on growth characteristics of *Vitis vinifera* L. cv. Cabernet Sauvignon II. Reproductive growth. S. Afr. J. Enol. Vitic. 11:26-32 (1990)
8. Hunter J.J., Ruffner H.P., Volschenk C.G., Le Roux D.J. Partial defoliation of *Vitis vinifera* L. cv. Cabernet Sauvignon/99 Richter: effect on root growth, canopy efficiency, grape composition, and wine quality. Am. J. Enol Viticul. 46:306-314 (1995)
9. Rojas-Lara B.A. & Morrison J.C. Differential effects of shading fruit or foliage on the development and composition of grape berries. Vitis 28:199-208. (1989)
10. Smart R.E. "Vine manipulation to improve wine grape quality", University of California, Davis Grape and Wine Centenial Symposium Proceedings pp. 362-375 (1980)
11. Eschenbruch R., Smart R.E., Fisher B.M., Whittles J.G. "Influence of yield manipulations on the terpene content of juices and wines of Müller Thurgau", Proceedings of the 7th Australian Wine Industry Technical Conference pp. 89-93 (1986)
12. Kliewer W.M. & Dokoozlian N.K. "Leaf area/crop weight ratios of grapevines: influence on fruit composition and wine quality", Proceedings of the American Society for Enology & Viticulture 50th Anniversary Annual Meeting pp. 285-295 (2000)
13. Intrieri C. & Filippetti I. "Planting density and physiological balance: comparing approaches to European viticulture in the 21st century," Proceedings of the American Society for Enology & Viticulture 50th Anniversary Annual Meeting pp. 296-308 (2000)
14. Jackson D.I. & Lombard P.B. Environmental and management practices affecting grape composition and wine quality—a review quality. Am. J. Enol Viticul. 44:409-430 (1993)
15. Marais J. & van Wyk C.J. Effect of grape maturity and juice treatments on terpene concentrations and wine quality of *Vitis vinfera* L. cv. Weisser Riesling and Bukettraube. S. Afr. J. Enol. Vitic. 7:26-35 (1986)
16. Marais J., van Wyk C.J., Rapp A. Carotenoid levels in maturing grapes as affected by climatic regions, sunlight and shade. Am. J. Enol Viticul. 12:64-69 (1991)

17. Reynolds A.G. & Wardle D.A. Flavour development in the vineyard: impact of viticultural practices on grape monoterpenes and their relationship to wine sensory response. S. Afr. J. Enol. Vitic. 18:3-18 (1997)
18. Davies C. & Robinson S.P. Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterization of cDNAs encoding putative cell wall and stress response proteins. Plant Physiol. 122(3):803-812 (2000)
19. Terrier N., Ageorges A., Abbal P., Romieu C. Generation of ESTs from grape berry at various developmental stages. J Plant Physiol. 158(12):1575-1583 (2001)
20. Ablett E., Seaton G., Scott K., et al. Analysis of grape ESTs: global gene expression patterns in leaf and berry. Plant Sci. 159(1):87-95 (2000)
21. Moser C., Segala C., Fontana P., et al. Comparative analysis of expressed sequence tags from different organs of *Vitis vinifera* L. Funct Integr Genomics 5(4):208-217 (2005)
22. Venter M., Burger A.L., Botha F.C. Molecular analysis of fruit ripening: the identification of differentially expressed sequences in *Vitis vinifera* using cDNA-AFLP technology. Vitis 40(4):191-196 (2001)
23. Burger A.L. & Botha F.C. Ripening-related gene expression during fruit ripening in *Vitis vinifera* L. cv. Cabernet Sauvignon and Clairette blanche. Vitis 43(2):59-63 (2004)
24. Zamboni A., Minoia L., Ferrarini A., et al. Molecular analysis of post-harvest withering in grape by AFLP transcriptional profiling. J Exp Bot. 59(15):4145-4159 (2008)
25. Fei Z., Tang X., Alba R.M., White J.A., et al. Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. Plant J. 40(1):49-57 (2004)
26. da Silva F.G., Iandolino A., Al-Kayal F., et al. Characterizing the grape transcriptome. Analysis of expressed sequence tags from multiple Vitis species and development of a compendium of gene expression during berry development. Plant Physiol. 139(2):574-597 (2005)
27. Terrier N., Glissant D., Grimplet J., et al. Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (*Vitis vinifera* L.) development. Planta 222(5):832-847 (2005)
28. Glissant D., Dedaldechamp F., Delrot S. Transcriptomic analysis of grape berry softening during ripening. J Int Sci Vigne Vin 42(1):1-13 (2008)
29. Waters D.L.E., Holton T.A., Ablet E.M., Slade Lee L., Henry R.J. cDNA microarray analysis of developing grape (*Vitis vinifera* cv. Shiraz) berry skin. Funct Integr Genomics 5(1):40-58 (2005)
30. Gatto P., Vrhovsek U., Muth J., et al. Ripening and genotype control stilbene accumulation in healthy grapes. J Agric Food Chem. 56(24):11773-11785 (2008)
31. Deluc L.G., Grimplet J., Wheatley M.D., et al. Transcriptomic and metabololite analyses of Cabernet Sauvignon grape berry development. BMC Genomics 8:429 (2007)
32. Pilati S., Perazzolli M., Malossini A., et al. Genome-wide transcriptional analysis of grapevine berry ripening reveals a set of genes similarly modulated during three seasons and the occurrence of an oxidative burst at véraison. BMC Genomics 8:428 (2007)
33. Grimplet J., Deluc L.G., Tillet L.R., et al. Tissue-specific mRNA expression profiling in grape berry tissues. BMC Genomics 8:187 (2007)

34. Mori K., Goto-Yamamoto N., Kitayama M., Hashizume K. Loss of anthocyanins in red-wine grape under high temperature. J Exp Bot. 58(8):1935-1945 (2007)
35. Deluc L.G., Quilici D.R., Decendit A., et al. Water deficit alters differentially metabolic pathways important flavor and quality traits in grape berries of Cabernet Sauvignon and Chardonnay. BMC Genomics 10:212 (2009)
36. Koyama K., Sadamatsu K., Goto-Yamamoto N. Abscisic acid stimulated ripening and gene expression in berry skin of Cabernet Sauvignon grape. Funct Integr Genomics 10(3):367-381 (2010)
37. Jaillon O., Aury J.M., Noel B., et al. French-Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463-467 (2007)
38. Velasco R., Zharkikh A., Troggio M., et al. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. Plos One 2(12):e136 (2007)
39. Zamboni A., Di Carli M., Guzzo F., et al. Identification of putative stage-specific berry biomarkers and omics data integration into networks. Plant Physiol .154(3):1439-1459 (2010)
40. Forcato C. Gene prediction and functional annotation in the *Vitis vinifera* genome. PhD Thesis, Universita' Degli Studi Di Padova (2010)
41. Fasoli M., Dal Santo S., Zenoni S., Tornielli G.B., Farina L., Zamboni A., Porceddu A., Venturini L., Bicego M., Murino V., Ferrarini A., Delledonne M., Pezzotti M. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. Plant Cell 24:3489-3505 (2012)
42. Dal Santo S., Tornielli G.B., Zenoni S., Fasoli M., Farina L., Anesi A., Guzzo F., Delledonne M. and Pezzotti M. The plasticity of the grapevine berry transcriptome. Genome Biol. 14:r54 (2013)
43. Margulies M., Egholm M., Altman W.E., et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376-380 (2005)
44. Bennet S. Solexa Ldt. Pharmacogenomics 5(4):433-438 (2004)
45. Mortazavi A., Williams B.A., McCue K., Schaeffer L., Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods 5(7):621-628 (2008)
46. Roberts A., et al. Identification of novel transcripts in annotated genomes using rna-seq. Bioinformatics 27:2325-2329 (2011)
47. Richard H., et al. Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. Nucleic Acids Res. 38:e112 (2010)
48. Wang Z., Gerstein M., Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10:57-63 (2009)
49. Zhang K., et al. Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. Nat. Methods 6:613-618 (2009)
50. Zenoni S., Ferrarini A., Giacomelli E., et al. Characterization of transcriptional complexity during berry development of *Vitis vinifera* using RNA-seq. Plant Physiol. 152(4):1787-1795 (2010)
51. Sweetman C., Wong D.C.J., Ford C.M., Drew D.P. Transcriptome analysis at four developmental stages of grape berry (*Vitis vinifera* cv. Shiraz) provides insights into regulated and coordinated gene expression. BMC Genomics 13:691 (2012)

52. Degu A., Hochberg U., Sikron N., et al. Metabolite and transcript profiling of berry skin during fruit development elucidates differential regulation between Cabernet Sauvignon and Shiraz cultivars at branching points in the polyphenol pathway. BMC Plant Biology 14:188 (2014)
53. Venturini L., et al. De novo transcriptome characterization of *Vitis vinifera* cv. Corvina unveils varietal diversity. BMC Genomics 14:41 (2013)
54. Da Silva C., Zamperin G., Ferrarini A., et al. 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(12):4777-4788 (2013)
55. Pastore C., et al. Increasing the source/sink ratio in *Vitis vinifera* (cv Sangiovese) induces extensive transcriptome reprogramming and modifies berry ripening. BMC genomics 12:631 (2011)
56. Vaux D.L., Fidler F., Cumming G. Replicates and repeats--what is the difference and is it significant? A brief discussion of statistics and experimental design. EMBO Rep. 13(4):291-296 (2012)
57. Wiklund S., Johansson E., Sjöström L., Mellerowicz E.J., Edlund U., Shockcor J.P., Gottfries J., Moritz T., and Trygg, J. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Anal. Chem. 80:115-122 (2008)
58. Kobayashi S., Goto-Yamamoto N., Hirochika H. Retrotransposon-induced mutations in grape skin color. Science 304(5673):982 (2004)
59. Walker A.R., et al., White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant J. 49:772 (2007)
60. Palumbo M.C., Zenoni S., Fasoli M., Massonnet M., Farina L., Castiglione F., Pezzotti M., Paci P. Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development. Plant Cell 26(12):4617-4635 (2014)

Chapter 2

Overview of the ten Italian wine grape varieties

INTRODUCTION

Historical origins of wine grapes

The grape is a unique not only for its worldwide cultivation but because its history has been tightly connected with the development of human culture since ancient time. In fact, the main product, wine, has always had a major role in the life style of Mediterranean people. In particular, ancient Greeks and Romans considered wine as divine, and dedicated the gods Dionysus and Bacchus to this beverage[1].

Among the *Vitaceae* family, *Vitis vinifera* species is the only one expensively used in wine industry. Its first apparition is suggested to occur about 65 million years ago in Eurasia. The domestication of grape seems linked to the discovery of wine, even if it is unclear which process came first. The earliest evidence of wine production was found in Iran at the Hajji Firuz Tepe site in the northern Zagros Mountains during the Neolithic period around 7400-7000 years ago[2]. Nevertheless, the discovery of remained seeds suggests also a primo-domestication in Western Europe.

From the primo-domestication sites, a gradual spread occurred to adjacent regions such as Egypt and Lower Mesopotamia. Then, there was further dispersal around the Mediterranean following the main civilizations[2]. Under the influence of Romans, *Vitis vinifera* expanded inland following the main trade river routes. By the end of Roman Empire, grape growing was common in most of the European locations where they are grown today. The Romans were the first to give names to cultivars and they probably acted in the differentiation between table and wine grapes in addition to the different colour types[3].

In the middle ages, the Catholic Church replaced the Romans in spreading grape cultivation to new regions through the crusades. Following the Renaissance (16th century), *Vitis vinifera* colonized the New World countries, where it was not indigenous, first as seeds in the case of America then by cuttings from their places of origin. Cuttings were also introduced to America, South Africa, Australia and New Zealand in the 19th century and later to North Africa. At the End of the 19th century, the introduction of disease-causing agents (mildews and Phylloxera) from America resulted in devastation and destruction of the major part of the European vineyards. This accident drastically reduced the diversity for both cultivated and wild grapes. European viticulture was saved from extinction by using American non-*vinifera* species as rootstocks in order to obtain disease-resistant inter-specific hybrids.

In conclusion, during *Vitis vinifera* domestication, the biology of grapes underwent several changes to ensure greater sugar content for better fermentation, greater yield and more regular production. Today, several thousand cultivars exist but only few of them are still cultivated. In fact, over the last 50 years, a drastic reduction of diversity has occurred due to the globalization of wine companies and marketing pressure.

Presentation of the ten selected grapevine varieties

For this research project, ten wine cultivars were selected: five red-skinned – Sangiovese, Barbera, Negro amaro, Refosco, Primitivo – and five white-skinned grape varieties – Vermentino, Garganega, Glera, Moscato bianco and Passerina. These cultivars were chosen among several hundreds of Italian varieties for their diversity in agronomical and oenological traits, and for adaptation to different growing locations (http://catalogoviti.politicheagricole.it) (**Fig. 1**).

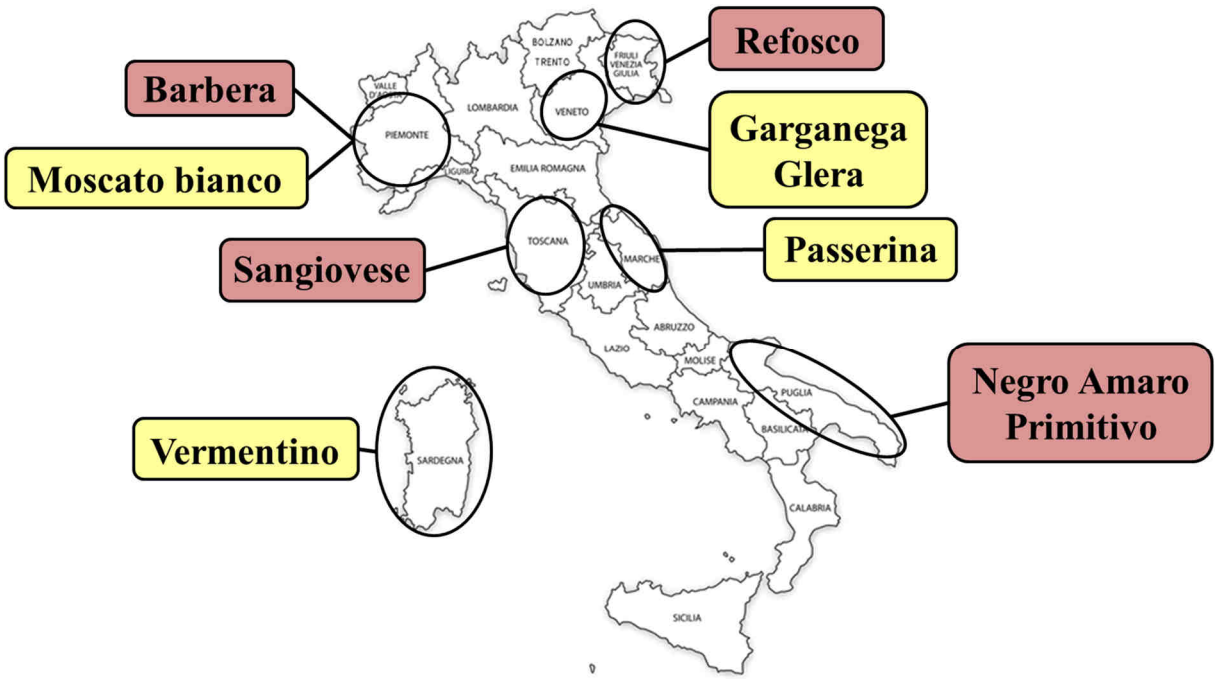


Figure 1: Map of Italian regions showing the growing areas of the selected grapevine varieties.

In this chapter, a brief description of each grapevine variety is provided in order to help the reader to have an overview of their main features.

FIVE RED GRAPE VARIETIES

Sangiovese

Synonym: Sangioveto.

Growth locations: In Italy, Sangiovese is the most widely planted red grape variety. It is an officially recommended variety in 53 provinces and an authorized planting in an additional 13 ones[4]. It is the main grape used in the popular red wines of Tuscany, where it is the solitary grape of Brunello di Montalcino and the primary component of the wines of Chianti, Vino Nobile di Montepulciano and many "Super Tuscans". Outside of Tuscany, Sangiovese is found throughout central Italy where it places an important role in the Denominazione di Origine Controllata e Garantita (DOCG) wines of Montefalco Sagrantino secco and Torgiano Rosso Riserva in Umbria, Conero in Marche and the Denominazione di Origine Controllata (DOC) wines of Lazio and Rosso Piceno in Marche. Significant Sangiovese plantings can also be found outside of central Italy in Lombardia, Emilia-Romagna, Valpolicella and as far south as Campania and Sicily[5].
Total growth surface in Italy: 71,558 ha in 2010 (from ISTAT data).

Bunch characteristics at harvest: Bunch is medium-sized (17-25 cm long), more or less compact, conical-shaped with one or two wings. Peduncle is thick, visible and semi-lignified.
Berry is medium-sized (12-15 mm diameter), spherical- or slightly elliptical-shaped. Skin is very waxy, dark-purple-coloured and medium-thick. Flesh is soft and pinkish-coloured.

Wine: Sangiovese is the fundamental grapevine variety of Chianti wine. In a typical Chianti blend, the portion of Sangiovese grapes can vary in function of the zone of production and can reach until 7/10 of the mix. Usually, Sangiovese is blend with "Canaiolo nero" (5-15%), "Malvasia" (5-15%) and "Trebbiano" (5-15%). Sangiovese

can be also used for a single-variety vinification but the obtained wine is often unpleasant in mouth, which ageing changes into a too-much-orange-coloured wine.

Barbera

Synonym: -

Growth locations: The highest quality Barbera wines come from the Piedmont region, where fifteen times more acreage is devoted to it than to Nebbiolo. Barbera d'Asti and Barbera del Monferrato each produce about three times as much wine as Barbera d'Alba. Colli Toronesi is produced in such small quantities that it is rarely found outside its own region. Barbera is also produced in Lombardy, Emilia-Romagna and Sardinia[6].
Total growth surface in Italy: 20,524 ha in 2010 (from ISTAT data).

Bunch characteristics at harvest: Bunch is medium-sized, pyramidal-shaped, more or less compact depending of the environmental and cultural conditions. Peduncle is medium-longed, semi-lignified and red-brown-coloured.
Berry is medium-sized and ovoid-shaped. Skin is very waxy, intense-blue-coloured and medium-thick. Flesh is very juicy, uncoloured, simple-flavoured, sweet but acidulous.

Wine: The resulting wines are deep, purplish black in their youth, but tend to early browning and lightening as they age. Tannin from oak aging can help somewhat to stabilize colour. Although normally indistinct in aroma, when cultivated in temperate areas and cropped for quality, Barbera can exhibit an attractive ripe aroma of red fruit, currants or blackberries that can be enhanced by vanilla, smoky or toasty notes added by barrel aging. On the other hand, neutral aroma, high colour and acidity are all good characteristics for blending with other grapes and this is how Barbera is most frequently used[6].

Negro amaro

Synonyms: Negroamaro, Nero amaro.

Growth locations: The cultivar grows up almost exclusively in Puglia and particularly in Salento, the peninsula which can be visualised as the “heel” of Italy.
Total growth surface in Italy: 11,460 ha in 2010 (from ISTAT data).

Bunch characteristics at harvest: Bunch is medium-sized (14-20 cm long), conical-shaped and compact. Peduncle is visible, medium-sized and woody at the two first centimetres.
Berry is medium-to-large-sized with a regular oval shape. Skin is thick, waxy and purple-coloured. Flesh is juicy, neutral-flavoured, sweet and coloured.

Wine: Negro amaro grapes can produce wines very deep in colour. Wines made from Negro amaro tend to be very rustic in character, combining perfume with an earthy bitterness. The grape produces some of the best red wines of Puglia, particularly when blended with the highly scented Malvasia Nera, as in the case of Salice Salentino.

Refosco (dal peduncolo rosso)

Synonyms: Refosco, Malvoise.

Growth locations: Refosco dal peduncolo rosso is grown predominantly in the Friuli-Venezia Giulia region of northeast Italy. It is found in the Denominazione di origine controllata (DOC) of Colli Orientali del Friuli, Friuli Aquileia, Friuli Grave and Friuli Latisana. Refosco is also found in the Veneto portion of the Lison Pramaggiore.
Total growth surface in Italy: 1,076 ha in 2010 (from ISTAT data).

Bunch characteristics at harvest: Bunch is medium-sized, pyramidal-shaped, winged and sparse. At maturity, peduncle is wine-red-coloured (hence the name) and lignified. Berry is medium-sized and spheroidal. Skin is thick, dark-coloured, opaque, and slightly waxy. Flesh is very sweet and soft.

Wine: The wines from this grape can be quite powerful and tannic, with a deep violet colour and a slight bitterness. On the palate, there are strong currant, wild berry and plum flavours. In the past, simple table wines were produced from Refosco but changing attitudes to the grape potential to produce finer, more age-worthy wines, has resulted in a general improvement in quality. Refosco wines have particularly high acidity and are mostly dark and densely coloured with violet and grassy aromas. Flavours of dark peppery spices and plums abound on the palate, and the wine often displays a slightly astringent, almond-skin like finish on the palate. The wine can stand some aging (depending on variety), and after a period of four-to-ten years, it achieves a floral quality as well.

Primitivo

Synonym: Zinfandel.

Growth locations: Widespread throughout northern Puglia.
Total growth surface in Italy: 12,234 ha in 2010 (from ISTAT data).

Bunch characteristics at harvest: Bunch is medium-sized (14-17 cm long), medium-compact, conical-cylindrical-shaped and winged. Peduncle is visible, short, thick and semi-lignified.
Berry is medium-sized (13-17 mm) and spheroidal-shaped. Skin is bluish, medium-thick and waxy. Flesh is pinkish, sweet with variety-typical flavours.

Wine: Tannic red wine, with a beautiful colour that can be very dark. High alcoholic if its production is reasonable and sometimes lacking acidity. Aromas of cherry, spice, floral for young wines, raspberry, blackberry, liquorice and blond tobacco.

FIVE WHITE GRAPE VARIETIES

Vermentino

Synonyms: Favorita, Pigato.

Growth locations: Vermentino is widely planted in Sardinia, in Liguria primarily under the name Pigato, in Piedmont under the name Favorita[7].
Total growth surface in Italy: 4,562 ha in 2010 (from ISTAT data).

Bunch characteristics at harvest: Bunch is medium-to-large-sized (15-20 cm long), conical-shaped and medium-compact. Peduncle is visible, herbaceous and medium-thick.
Berry is medium-to-large-sized and spherical-shaped. Skin is waxy, medium-thick, greenish-yellow- or amber-yellow-coloured depending of the environment. Flesh is uncoloured, sweet and neutral-flavoured.

Wine: The Gallura's Vermentino wine (Sardinia) shows an intense straw yellow colour with bright gold reflections, intense and refined aromas of ripe white-pulp fruit, broom, herbs. In mouth, it offers a feeling of softness and fresh acidity with a final of warm mineral notes[8].

Garganega

Synonym: Garganego.

Growth locations: Garganega is a variety of white Italian wine grape widely grown in the Veneto region of North East Italy, particularly in the provinces of Verona and Vicenza.
Total growth surface in Italy: 11,292 ha in 2010 (from ISTAT data).

Bunch characteristics at harvest: Bunch is large-sized (20-25 cm long), conical-shaped, winged (with a wing more developed than the other one) and loosely knit. Peduncle is thick, visible, green-yellow-coloured and herbaceous.
Berry is medium-sized (15 mm) and spherical-shaped. Skin is yellow-gold-coloured (even yellow-amber-colour if the berry is well sun exposed), waxy, thin and tough. Flesh is soft, simply flavoured and uncoloured.

Wine: Garganega is a delicate white wine with notable flavours of almond, lemon and a light spices. It has a pale colour with soft lemon hues. The Garganega grape variety is one of Italy's most planted grapes and comes from the Veneto region, which is a province of Verona. Garganega is the primary component of the Venetian wine Soave. Soave is a full-bodied wine with good acidity and has the aromas of honeydew melon, pear and yellow or white flowers. Garganega grapes are also the key grape used in "recioto" or "Straw Wine" which is known as the warm climate ice wine style. It also provides a basis to Gambellara wine – a light, crisp, dry white wine that is similar to the Soave wines but generally not outstanding)[9].

Glera

Synonym: Serprino.

Growth locations: Glera is grown mainly in the Veneto region of Italy, traditionally in an area near Conegliano and Valdobbiadene, in the hills in the north of Treviso.
Total growth surface in Italy: 18,255 ha in 2010 (from ISTAT data).

Bunch characteristics at harvest: Bunch is large-sized (20-25 cm long), elongated, pyramidal-shaped, winged. Peduncle is long, thin and herbaceous.
Berry is medium-sized and spherical-shaped. Skin is waxy, yellow-gold-coloured and thin.
Wine: Italian wine produced from Glera is usually either slightly fizzy or sparkling (in Italian, *frizzante* and *spumante* respectively). A few still wines are also made from

Glera, but on nowhere near the same scale as the sparkling wines that are so exported around the globe. The worldwide thirst for Prosecco has resulted in many imitations of the style. Glera has high acidity and a fairly neutral palate, making it ideal for sparkling wine production. Glera's aromatic profile is characterized by white peaches, with an occasional soapy note. The wine is light-bodied and low in alcohol (8.5% is the minimum permitted alcohol level for Prosecco), well suited to be drinking in the summer months or as an aperitif[10].

Moscato bianco

Synonyms: Muscat blanc à petit grain, Muscat de Chambave, Moscato, Moscatello, Moscatellone, Muscar, Muskatteller, Moscato reale, Gelbert Muskateller.

Growth locations: In Piedmont region, it is really the most intensely white variety cultivated and one of the leading in many cities of the provinces of Cuneo, Asti and Alessandria. The cultivar is present, also occasionally, in many other wine-growing areas of Piedmont, including mountain and foothill areas. In the rest of Italy, the variety is known and used for the production of aromatic wines (Valle d'Aosta, Oltrepò Pavese, the Euganean Hills, Tuscany, Apulia, Sicily and Sardinia)[11].

Total growth surface in Italy: 11,506 ha in 2010 (from ISTAT data).

Bunch characteristics at harvest: Bunch is medium-sized, compact, conical-cylindrical-shaped and winged. Peduncle is visible, medium-thick and semi-lignified.

Berry is medium-sized and spherical-shaped. Skin is slightly waxy, yellow-gold-coloured to amber-coloured on the sun-exposed part and medium-thick. Flesh is firm, sweet with typical Muscat flavours.

Wine: Sweet and intensely aromatic sparkling wines are produced with Moscato bianco grapes. After withering, they also get dessert wines of excellent quality. Dry wines can

also be produced from these berries but they are rare. In this case, they are aromatic wines to be consumed as aperitifs[12].

Passerina

Synonym: -

Growth locations: The majority of Passerina vines are grown in Marche's Piceno province (of Rosso Piceno fame), but there are also cultivated all across central Italy such as in Abruzzo, Emilia-Romagna and Lazio[13].
Total growth surface in Italy: 894 ha in 2010 (from ISTAT data).

Bunch characteristics at harvest: Bunch is large-sized, conical-shaped, winged and medium-compact. Peduncle is short, thick and lignified.
Berry is large-sized (17-24 mm long) and spherical-shaped. Skin is yellow-coloured, medium-waxy and thick. Flesh is uncoloured, sweet and soft.

Wine: Passerina grapes give a wine straw-yellow-coloured with golden reflections. At the nose, the wine is fruity with hints of tropical fruit, and floral with notes of broom, full (round) and dry. With good acidity, it can be used both for making sparkling and for withering wines[14].

CONCLUSION

The transcriptomic study of these 10-variety berries was a part of an important nationwide research project. Indeed, “Vigneto” project – Valorizzazione dei Principali Vitigni Autoctoni Italiani e dei loro Terroir – was launched by the Italian Ministry of Agricultural and Forestry Policies in 2010 to defend and enhance the extraordinary Italian wine heritage. The main goal of this research project was to draw up a genetic passport of the main Italian grape varieties, which could describe from a molecular perspective the diversity of Italian grape varieties that leads to even greater variability within wine production.

REFERENCES

1. This P., Lacombe T., Thomas MR. Historical origins and genetic diversity of wine grapes. Trends in genetics 22:511-519 (2006)
2. McGovern P.E. “Ancient wine: the search for the origins of viniculture”. Princeton University Press (2004)
3. Roxas C. “Essai sur les variétés de la vigne qui végètent en Andalousie”. Imprimerie Poulet (1814)
4. Oz Clarke. Encyclopedia of Grapes Harcourt Books pp. 209-216 (2001)
5. Robinson J. The Oxford Companion to Wine (Third Edition). Oxford University Press pp. 606–607 (2006)
6. Robinson J. The Oxford Companion to Wine (Third Edition) Oxford University Press pp. 62-63 (2006)
7. Robinson J. The Oxford Companion to Wine (Third Edition) Oxford University Press pp. 732 (2006)
8. http://www.sardegnaagricoltura.it/index.php?c=4402&s=45552&v=2&xsl=443
9. http://www.grapeheaven.com/learn/Garganega-i72.aspx
10. http://www.wine-searcher.com/grape-392-glera-prosecco
11. http://www.regione.piemonte.it/agri/politiche_agricole/viticoltura/dwd/vitigni/varieta_cloni/moscato_descr.pdf
12. http://www.regione.piemonte.it/agri/politiche_agricole/viticoltura/dwd/vitigni/varieta_cloni/moscato_descr.pdf
13. http://www.wine-searcher.com/grape-1988-passerina
14. http://www.quattrocalici.it/vitigni/passerina

Chapter 3

RNA-sequencing assay of 120 berry samples

INTRODUCTION

Grape berry structure

According to standard botanical use, the ovary wall develops into the fruit wall or pericarp. The pericarp is in turn subdivided into an outer exocarp, the middle mesocarp, and the inner endocarp[1]. However, from a winemaking perspective, the grape berry has three major types of tissue: skin and flesh representing the berry pericarp, and seed[2] (**Fig. 1**). Pericarp tissue contributes about 65% of the berry volume.

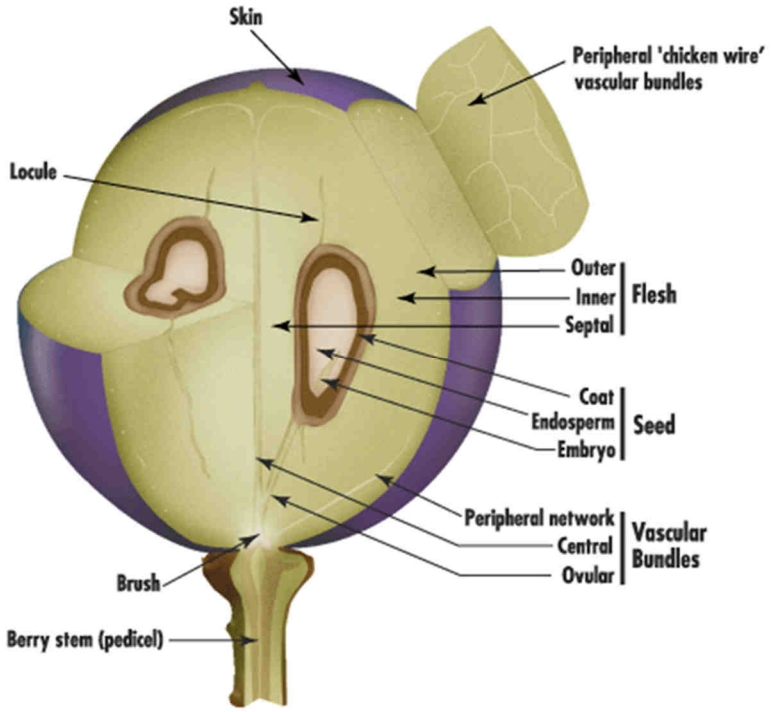


Figure 1: Structure of a ripe grape berry partially sectioned on the long and central axis to show internal parts. Illustration by Jordan Koutroumanidis, Winetitles[2].

Skin

The epidermis consists of a single layer of flattened, disk-shaped cells, with irregularly undulating edges[3]. These nonphotosynthetic cells may develop thickened or lignified cell walls, depending of the cultivar. Unlike some other plant surfaces, the skin of a grape berry does not contain a significant number of functional stomata[4] and is covered by a relatively thick layer of cuticle and epicuticular wax[5]. During ripening, preharvest water loss may occur due to a decreased vascular flow of water into the berry combined with continued transpiration and cuticle disruption, a relatively slow process[6].

Just under the outer layer of skin cells lies the hypodermis – a tightly packed layer of flattened cells with thickened corner walls. The hypodermis commonly contains ten layers of cells, but this can vary depending on the cultivar. When, the cells are young, they are photosynthetic[7]. After véraison, the plastids lose their chlorophyll and starch contents[8].

The hypodermal cells tend to accumulate phenolic compounds in relatively high concentrations as the grape berry matures. Among their other effects, these compounds strongly influence aroma, flavour, and colour attributes. Therefore, they play a key role in determining final juice and wine quality and sensory characteristics. The compounds concentrated in these cells include many types of tannins, which strongly influence elements of mouth feel such as astringency, and flavour such as bitterness, and many specific flavour notes. Pigments, mainly anthocyanins in red grapes, accumulate in these cells beginning at véraison. Many aromatic compounds such as the terpenoids, responsible for pleasant, fruity aromas, begin accumulating in these cells during this time as well.

Flesh

The cells that make up grape berry flesh are typically larger and rounder than the cells that make up the skin. These cells contain large vacuoles, which are the primary sites for the accumulation of sugars. Acids and phenolic compounds are also concentrated in these vacuoles, with the primary acids being tartaric and malic acid[9]. Although the outer mesocarp cells contain plastids, they rarely contribute significantly to photosynthesis.

The vascular tissue of the fruit consists primarily of a series of peripheral bundles that ramify throughout the outer circumference of the berry, and axial bundles that extend directly up through the septum. During its development, the grape berry obtains water and other nutrients through this vascular system. Xylem serves to transport nutrients primarily from the roots. These nutrients include minerals, hormones, and other compounds. Phloem serves to transport nutrients primarily from the canopy; these are mainly sugars, mostly sucrose.

Seeds

Grape seeds are contained in locules, which are indistinct in mature berries. The seeds are composed of an outer seed coat, the endosperm, and the embryo. As with most

seeds, the endosperm comprises the bulk of the grape seed and serves to nourish the embryo during early growth. The seed coat also contains relatively high concentrations of tannins. Similar to the tannins and phenols found in the flesh, these tannins are reduced in concentration on a per-berry basis as the berries mature. In particular, phenols responsible for bitterness are altered or rendered less soluble/extractable.

These tissues vary considerably in size and thus contribute differently to overall wine composition. As a general rule, wines made from smaller berries will have a higher proportion of skin and seed derived compounds. In addition to the effect of berry size on the proportion of skin and seed-derived components in wine, the actual number of seeds in the berry can influence the proportion of seed-derived components in wine. The normal or perfect number of seeds in the grape is four but the actual number is much less[10].

Grape berry development

Development of the grape berry, a non-climateric fruit, follows a double-sigmoidal growth curve composed of two main phases called green phase and maturation phase separated by a lag phase[11] (**Fig. 2**).

Green phase: berry formation

From flowering to approximatively 60 days, the berry is formed and the seeds are produced. Berry size increases rapidly by cell division then by cell expansion. During this phase, the role of the pericarp is to protect the seeds in development against animal attacks. Therefore, the berry is small, hard, and green as the rest of the plant due to the presence of chlorophyll, highly acidic with the accumulation of tartaric acid principally at the periphery of the berry, vegetal-odorous with the synthesis of methoxypyrazines and bitter due to the high concentration of tannins in the skin[12]. Hydroxycinnamic acids, that are precursors of volatile phenols[13], are also accumulated in the flesh and skin of the developing berry during this period.

In most cultivars, the first growth phase is followed by a lag phase. During this period, berry growth slows markedly while the organic acid concentration of the berry reaches its highest level with the accumulation of malic acid in the flesh cells. These organic acids, i.e. malate and tartrate, act as chemical stores of energy and may also be

involved in defence of the berry before seeds are fully developed. In fact, it is also during this period that the embryo within the seed matures and the seed coat lignifies. The duration of this phase depends both of the variety and the temperature. Its end corresponds to the end of the herbaceous phase of the fruit.

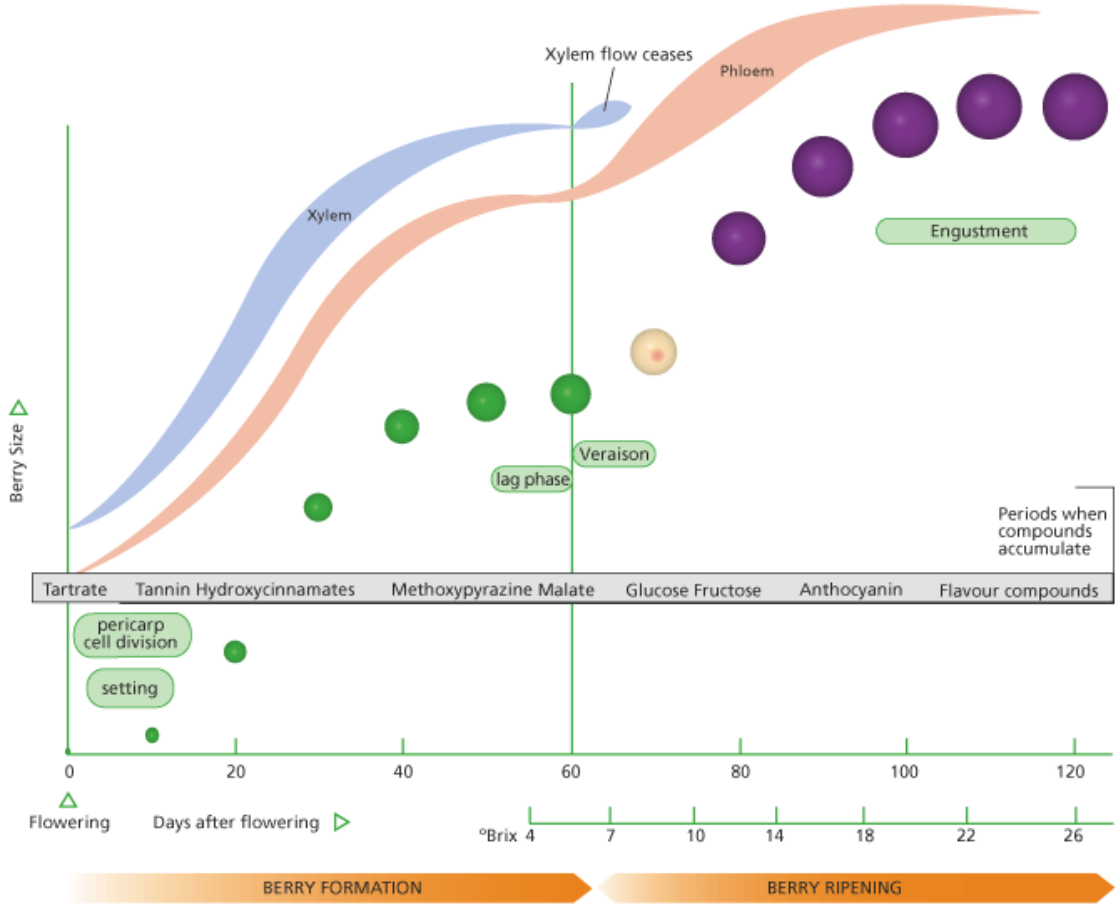


Figure 2: Diagram showing relative size and colour of berries at 10-day intervals after flowering, passing through major developmental events (round boxes). Also shown are the periods when compounds accumulate, the levels of juice °Brix, and an indication of the rate of inflow of xylem and phloem vascular saps into the berry. Illustration by Jordan Koutroumanidis, Winetitles[2].

Maturation phase: berry ripening

Once seeds are mature, the berry needs to become attractive to guarantee the dispersal of the seeds by herbivores, and the onset of the ripening phase, called véraison, occurs. The French word véraison was first used to describe the change in berry skin colour and then adopted to describe the onset of ripening. Consequently, the process includes the change of colour with the synthesis of anthocyanins in red grape skins, but also the cell expansion coupled with a water influx, the softening of

the berry, the increase of sugars in the pulp, the decline of organic acids and tannins, and the synthesis of volatile aromas[12]. During this phase, water, sugars, and nitrogen compounds are transported to the berry via the phloem. Sucrose is hydrolysed to glucose and fructose in the berry flesh. Tartaric acid concentration decreases as the berry increases in volume whereas malic acid concentration declines because it is used as an energy source for berry metabolic processes. Berry flavour and aroma compounds are synthesized within the berry while methoxypyrazine concentration generally decreases as the berry matures.

At harvest time, the final product of this complex development, i.e. the fully ripe berry, is composed of many components that will have a great impact on berry quality. In fact, mature berry quality is usually evaluated using several or many of this following criteria: aroma/flavour and intensity of aroma/flavour, grape skin tannins and tannin extractability, stem lignification or 'ripeness', seed numbers per berry, seed 'ripeness' or tannin extractability, sugar composition per berry, red fruit colour in case of red grape varieties, °Brix, acidity, pH, berry softness and size/weight[14].

Reference genome and gene annotation versions

For the last decade, the substantial economic relevance of grapevine and the need to resolve field issues have led to develop physiology and pathology studies coupled with the expansion of genetic resources and tools. In 2007, a milestone for grapevine biology occurs with the publication of the first draft of the grapevine genome by the French-Italian Consortium[15]. The French-Italian sequencing was performed on the Pinot noir 40024 line selected for its high degree of homozygosis (approximatively 93%). This particular line was obtained through multiple self-pollination cycles in order to by-pass the high heterozygosis that characterizes grapevine. From the consortium sequencing, a total of 6.2 million end-reads were generated, representing an 8.4-fold coverage of the genome. From this 8X coverage assembly, a set of 30,434 protein-coding genes was predicted from the 475 Mb genome using the GAZE software[16]. Then, the sequencing and the assembly of the Pinot noir 40024 have been rapidly updated from a 8X to a 12X coverage of the genome sequence and a 12X assembly[17]. The 8X and 12X assemblies are accompanied by the respective gene structure predictions, which contain different types of subsequence predictions. These include genes, mRNAs, UTRs, introns, exons, and inter-genic spaces. Three versions of the

12X prediction are available. Version 0 (V0) was performed with the GAZE software by the Genoscope in France and contains 26,346 annotated genes (http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/). Version 1 (V1) is the result of the union of V0 and a gene prediction performed with JIGSAW software[18] at the CRIBI in Padova (Italy) thus gaining 29,971 predicted genes and increased reliability[19]. The last version (V2) was published by Vitulo et al.[20] in 2014. This new gene prediction improves the previous one with 2,258 new coding genes and 3,336 putative long non-coding RNAs by integrating the information derived from the considerable amount of newly available RNA-Seq data and setting up rigorous bioinformatic procedures based on several filtering steps. The public release of grapevine reference genome enabled the profiling of grape berry developmental gene expression at the whole-genome level with initially the creation of microarray platforms and then the use of next-generation-sequencing technologies as the RNA-Sequencing.

Basic workflow of a RNA-sequencing assay using the Illumina HiSeq 1000 instrument

The workflow of high-throughput sequencing with the HiSeq 1000 can be divided into three major steps: the library preparation using the Illumina TruSeq RNA Sample preparation protocol, the cluster generation using the Illumina cBot and the sequencing by the Illumina HiqSeq 1000 sequencer (Illumina Inc, San Diego, CA, USA).

Library preparation

The Illumina TruSeq RNA Sample preparation protocol follows a specific workflow (**Fig. 3**).

Firstly, mRNA molecules, containing poly-A tails, are selected using poly-T oligo-attached magnetic beads through two rounds of purification. During the second elution of the poly-A RNA, the RNA is also fragmented (165 bp average) and primed with random hexamers for cDNA synthesis.

Secondly, the RNA fragments are reverse-transcribed into a first strand cDNA using the reverse transcriptase Superscript II (Invitrogen, Carlsbad CA, USA). Then, the synthesis of the second strand is performed and RNA templates are removed. Agentcourt Ampure XP beads (Beckman Coulter, Fullerton CA, USA) are used to separate the double-stranded cDNAs from the reaction mix.

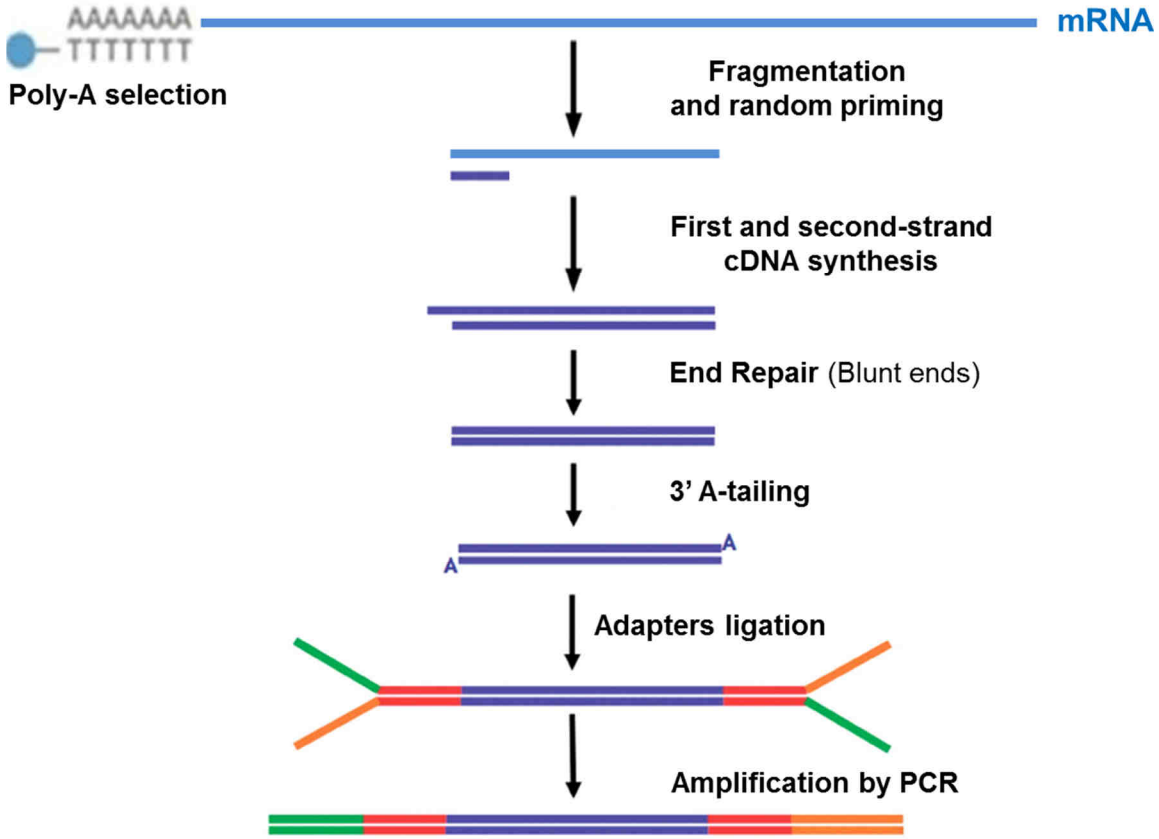


Figure 3: Workflow of Truseq cDNA library preparation protocol.

Thirdly, the End Repair process is performed in order to convert the overhangs resulting from fragmentation into blunt ends. In fact, the reaction mix is composed of a 3' to 5' exonuclease activity of this mix removes the 3' overhangs and the polymerase activity fills in the 5' overhangs.

Fourthly, a single Adenine nucleotide is added to the 3' ends of the blunt fragments to prevent them from ligating together during the adapter ligation reaction. A corresponding single Thymine nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragment. This strategy ensures a low rate of chimera (concatenated template) formation.

Fifthly, universal and indexed adapters are ligated to the ends of the double-stranded cDNA that will allow the hybridization onto a flow cell.

Finally, a PCR (Polymerase Chain Reaction) is performed with adapter-specific primers in order to selectively enrich the cDNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the library.

Cluster Generation

The Cluster Generation is performed on the Illumina cBot. Single cDNA-fragments are attached to the flow cell by hybridizing to oligos on its surface that are complementary to the ligated adapters. The cDNA-molecules are then amplified by a so called bridge amplification which results in a hundred of millions of unique clusters. Finally, the reverse strands are cleaved and washed away and the sequencing primer is hybridized to the DNA-templates (**Fig. 4 a-i**).

Sequencing

During sequencing the huge amount of generated clusters are sequenced simultaneously. The cDNA-templates are copied base-by-base using the four nucleotides (ACGT) which are fluorescently labelled and reversibly terminated (**Fig. 4 j-m**). After each synthesis step, the clusters are excited by a laser, which causes fluorescence of the last incorporated base. After that, the fluorescence label and the blocking group are removed allowing the addition of the next base. The fluorescence signal after each incorporation step is captured by a built-in camera and converted into a single base using a base-calling software.

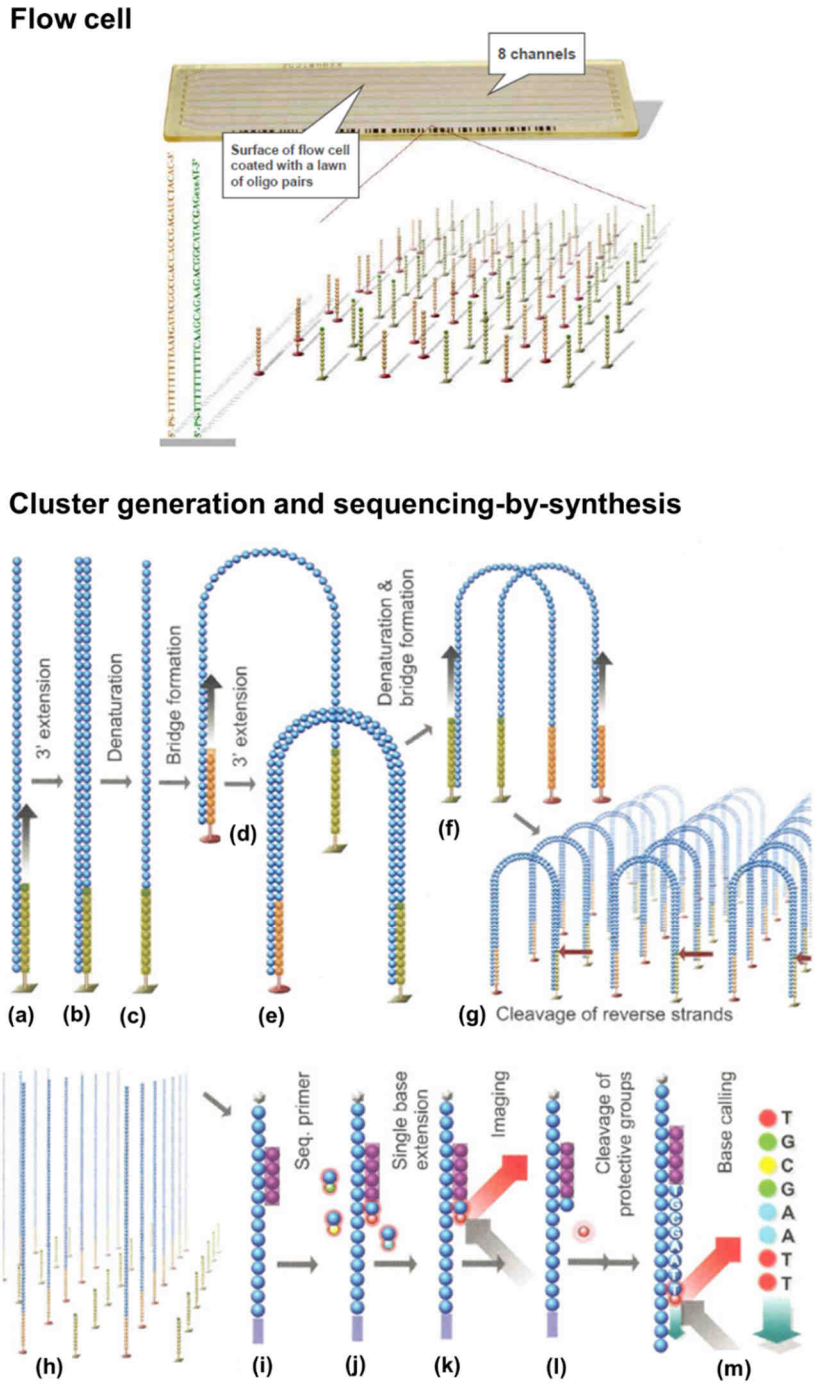


Figure 4: Flow cell overview and sequencing-by-synthesis process

Cluster formation: Denatured DNA is bound to primers on the flow cell surface (a), extended (b), and then denatured (c). The single-stranded DNA forms a bridge to adjoining lawn primers (d), is extended again to create a double-stranded bridge (e), denatured and hybridized to form a new singe-stranded DNA bridges (f). The process is repeated 35 times to give discrete clusters with ~2000 molecules (g). Reverse strands are cleaved (arrow, g), 3' ends are blocked and the strands are hybridized to sequencing primers (i).

Sequencing: DNA polymerase incorporates a reversible terminator (j), the fluorescent base is imaged (k), the fluorescent tag is then cleaved off, and the terminator is unblocked (l). The process is repeated for 100 cycles to determine the sequence (m).

GRAPE BERRY MATERIALS

Vineyard features

Grape berries were collected from ten *Vitis vinifera* cultivars, five red-skinned ones: Sangiovese, Barbera, Negro amaro, Refosco, Primitivo, and five white-skinned ones: Vermentino, Garganega, Glera, Moscato bianco and Passerina all grafted onto S.O.4 rootstock and cultivated in the same experimental vineyard (Conegliano, Veneto region, Italy) and with similar culture conditions. The grapevine varieties were placed by block of five plants, in an area of 50 meters radius range composed of a homogeneous soil. Berries were collected during the 2011 season at the same time of the day (between 9 and 10 am).

Sample collection

In order to cover the whole berry development, berries were collected at four phenological growth stages, two pre- and two post-véraison, by freezing the whole berry in liquid nitrogen (**Fig. 5**).

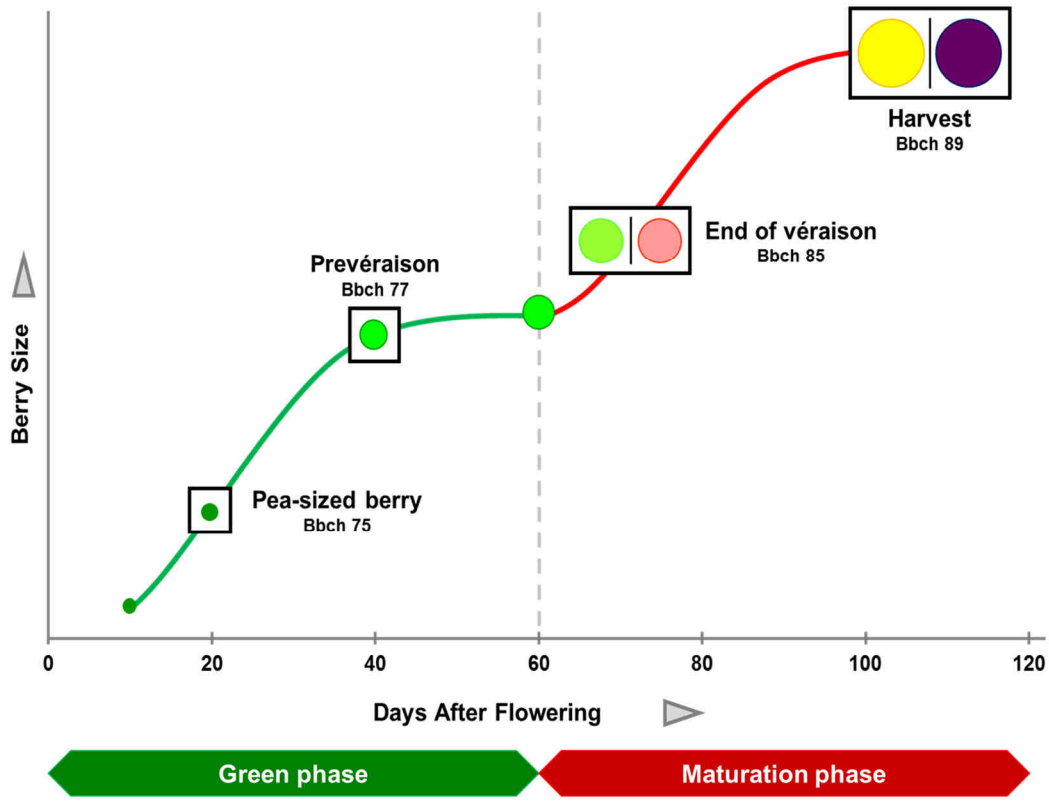


Figure 5: Four sampling phenological growth stages covering whole berry development. Each time point is represented according to the Bbch-scale for grapevine phenological growth stages defined by Lorenz et al.[21].

The two stages of the green phase corresponded to pea-sized berries (Bbch 75) at almost 20 days after flowering and berries beginning to touch (Bbch77) just prior véraison. The two stages representing the maturation phase are the softening of berries (Bbch 85) at the end of véraison and berries fully ripe for harvest (Bbch 89) (**Supplementary table 1**). Berries were collected from the entire five-vine block, taking care to pick up clusters from both sides of the canopy. The sampling of full-ripe berries was done considering the commercial Brix degree value of each variety but also the sanitary status of the berries – clusters were harvested at the first sign of rotting. At each stage, three berries were randomly selected from each cluster representing almost one hundred berries and immediately frozen in liquid nitrogen. From each berry sample, three biological replicates were created in order to obtain representative results. Brix degree of the must was determined using the digital DBR35 refractometer (Giorgio Bormac, Italy) for post-véraison samples (**Supplementary table 2**).

GRAPE BERRY SAMPLE PREPARATION

RNA extraction from berry pericarps

High quality RNA is required for optimal cDNA library elaboration. Consequently, in order to overcome the problem of high water content and contaminants, Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, Saint-Louis, MO, USA) was used following the company instruction manual with some specific modifications[22] to extract total RNA from berry pericarp.

1. Solution preparation and basic instruction

Prepare Lysis Solution/2-Mercaptoethanol Mixture: calculate the total amount of Lysis Solution Mixture needed for processing samples and add 10 µL of 2-Mercaptoethanol to each mL of Lysis Solution Mixture needed.

All centrifugation steps are performed at room temperature and at maximum speed (14,000-16,000 rpm) in a standard microcentrifuge.

2. Lyse Tissue Sample

Weight 400 mg of berry pericarp powder split in two tubes. Pipet 900 µL of the Lysis Solution/2-Mercaptoethanol Mixture to 200 mg of powder and vortex immediately and

vigorously for at least 30 seconds. Incubate the sample at 56°C for 3-5 minutes with shaking (800rpm). Do not vortex the sample during or after the heat incubation.

3. Pellet Cellular Debris

Centrifuge the sample at maximum speed for 10 minutes to pellet cellular debris.
Note: Mark the tube orientation before centrifugation in order to avoid the cellular debris pellet in the next step.

4. Filter Lysate

Pipet the 700-µL lysate supernatant into a Filtration column (blue retainer ring) seated in a 2-mL Collection Tube by positioning the pipette tip at the bottom of the tube but away from the pellet. If there is a layer of floating particulates, position the pipette tip below the floating layer and away from the pellet before pipetting the supernatant. Do not be concerned with carry-over of the floating particulates to the Filtration Column, but avoid taking the pellet. Close the cap and centrifuge at maximum speed for 1 minute to remove residual debris. Repeat this step until the lysate supernatant is finished using the same Filtration column, but collecting every 700 µL of the clarified flow-through lysate in separate tubes (three tubes with almost 700 µL of lysate supernatant each).

5. Bind RNA to Column

Pipet 750 µL of Binding Solution into each of the clarified lysates and mix immediately and thoroughly by pipetting at least 5 times or vortex briefly. Do not centrifuge.
Pipet 700 µL of the mixture into a Binding Column (red retainer ring) seated in a 2-mL Collection Tube. Close the cap and centrifuge at maximum speed for 1 minute to bind RNA. Decant the flow-through liquid and tap the Collection Tube (upside down) briefly on a clean absorbent paper to drain the residual liquid. Return the column to the Collection Tube and pipet the remaining mixture to the column and repeat the centrifugation and decanting steps. Go on this procedure until you have filtered everything onto the same red Binding Column.

6. First Column Wash

Pipet 500 µL of Wash Solution 1 into the column. Close the cap and centrifuge at maximum speed for 1 minute. Decant the flow-through liquid and tap the Collection

Tube (upside down) briefly on a clean absorbent paper to drain the residual liquid. Return the column to the Collection Tube.

7. Second Column Wash

Ensure that Wash Solution 2 Concentrate has been diluted with ethanol as described in the Preparation Instructions. Pipet 500 µL of the diluted Wash Solution 2 into the column. Close the cap and centrifuge at maximum speed for 30 seconds. Discard the flow-through liquid and tap the Collection Tube (upside down) briefly on a clean absorbent paper to drain the residual liquid. Return the column to the Collection Tube.

8. Third Column Wash

Pipet another 500 µL of the diluted Wash Solution 2 into the column, close the cap and centrifuge at maximum speed for 30 seconds. Discard the flow-through liquid and tap the Collection Tube (upside down) briefly on a clean absorbent paper to drain the residual liquid. Return the column to the Collection Tube.

9. Dry Column

Centrifuge the column at maximum speed for 1 minute to dry. Carefully remove the column-tube assembly from the centrifuge to avoid splashing the residue flow-through liquid to the dried column. If the residual flow-through liquid does accidentally contact the dried column, re-centrifuge the column for 30 seconds before proceeding to the elution step.

10. Elution

Transfer the column to a new 2-mL Collection Tube. Pipet 55 µL of Elution Solution directly onto the centre of the filter inside the column. Close the cap and let the tube sit for 1 minute. Centrifuge at maximum speed for 1 minute to elute. Purified RNA is now in the flow-through elute and ready for immediate use or storage at -20 °C (short-term) or -80 °C (long-term).

RNA Quality Control

RNA quality and quantity were determined using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Illumina recommends starting with 2.5 µg

of total RNA for each cDNA-library preparation with a sufficient purity, namely ratios of absorbance at 260 to 280 nm and 260 to 230 nm must be both higher than 1.8. When ratio of absorbance at 260 to 230 nm was below 1.8, total RNAs were precipitated with LiCl to remove contaminants that absorbed at 230 nm. LiCl was mixed with total RNA to a final concentration of 2.5 M, incubated overnight at 4°C, and centrifuged at 13,000g, and the pellet was washed with 70% ethanol before resuspending in water.

RNA Integrity

The state of RNA intactness of each RNA sample was characterized using a Bioanalyzer Chip RNA 7500 series II (Agilent, Santa Clara, CA, USA) (**Fig. 6**), as reported in the manufacture instruction.

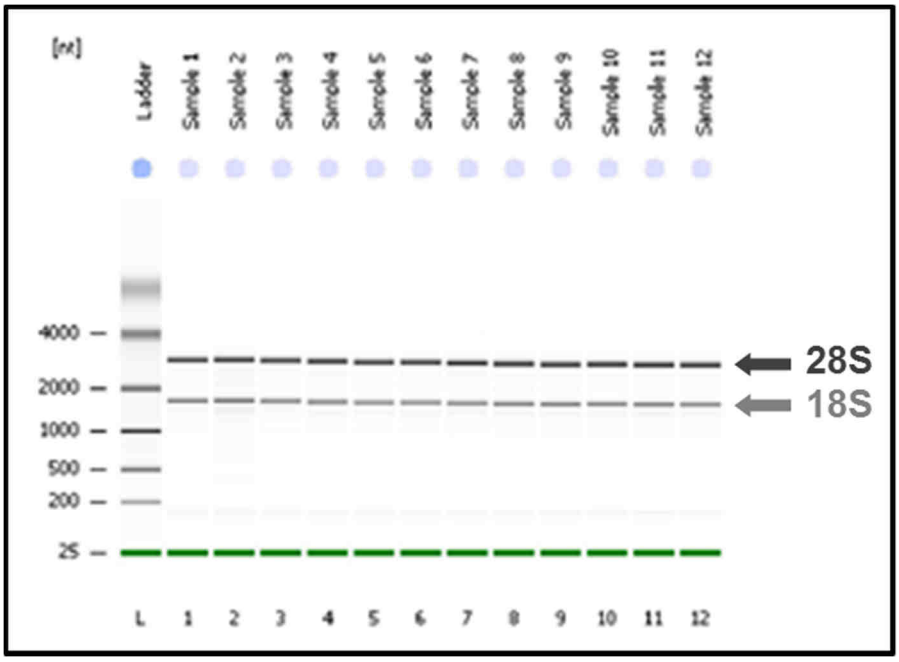


Figure 6: Example of a Bioanalyzer-RNA-analysis result. Ladder and 12 RNA samples are visualized as in a virtual electrophoresis. Arrows indicate the signals corresponding to ribosomal 28S and 18S RNAs.

GRAPE BERRY SAMPLE RNA-SEQUENCING

Library Preparation and Sequencing

All 120 non-directional cDNA libraries were prepared from 2.5 µg of total RNA using the Illumina TruSeq RNA Sample preparation protocol (Illumina Inc, San Diego, CA, USA). The Illumina protocol was two-days-long and permitted the elaboration of

sixteen c-DNA libraries per week. After each library elaboration, a Bioanalyzer Chip DNA 1000 series II (Agilent, Santa Clara, CA, USA) was realized in order to check the size distribution and the purity of the sample (**Fig. 7**). Each library had an average size of 286 bp corresponding to the addition of the median insert length (165 bp) and the universal and indexed adapter lengths (58 and 63 bp respectively).

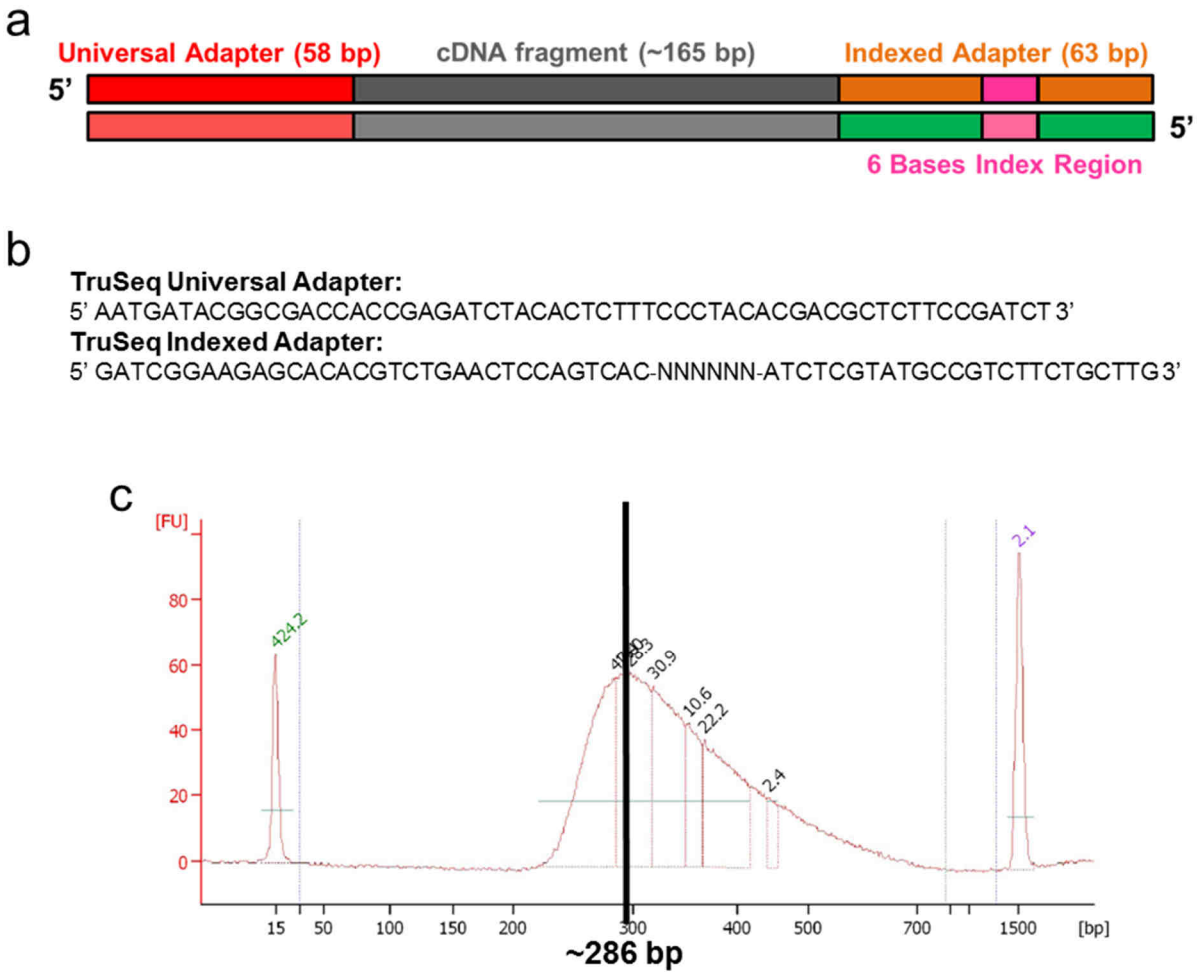


Figure 7: (**a**) TruSeq c-DNA library structure, (**b**) TruSeq Universal and Indexed adapter sequences and (**c**) example of a Bioanalyzer-DNA output showing the length of the library (bp) in function of the fluorescence measure (FU:Fluorescence Unit).

Library quantification was done by quantitative PCR using KAPA SYBR® FAST qPCR kit (KapaBiosystems, Wilmington, MA, USA) in order to pool all the libraries in equimolar concentrations. Primers used for the real-time qPCR were specific to the sequence of the universal and indexed adapters, and furnished by Illumina. The PCR involved a 95°C hold for 3 minutes followed by 40 cycles at 95°C for 10 seconds and

60°C for 30 seconds, then a 95°C hold for 1 minute, a 55°C hold for 30 seconds and finally 95°C hold for 30 seconds. Each real-time PCR was performed three times.

Dilutions and pooling of the Tru-Seq libraries were done by set of six samples in order to obtain a pool concentration of 10 nM. Then, the pools quality was determined with a Bioanalyzer Chip DNA 1000 series II (Agilent, Santa Clara, CA, USA) by Genny Buson. She also performed the quantification of the pool-libraries by real-time qPCR as described above. Finally, she diluted again all the pools to a final concentration of 2 nM before the flow-cell loading procedure.

Single-ended reads of 100 nucleotides were obtained using an Illumina Hiseq 1000 sequencer in collaboration with the Functional Genomic Lab directed by Massimo Delledonne at the University of Verona. Illumina Casava 1.8 software was used for the base calling in order to generate the data (average value of reads per sample ± standard deviation: 33,438,747 ± 5,610,087).

SUPPLEMENTARY MATERIAL

Supplementary table 1: Date of berry collection of each variety at each growth stage.

Variety	Pea-sized berry	Prevéraison	End of véraison	Harvest
Sangiovese	21/06/11	12/07/11	09/08/11	22/09/11
Barbera	21/06/11	12/07/11	05/08/11	06/09/11
Negro amaro	21/06/11	12/07/11	09/08/11	22/09/11
Refosco	21/06/11	12/07/11	10/08/11	20/09/11
Primitivo	21/06/11	12/07/11	01/08/11	20/09/11
Vermentino	23/06/11	15/07/11	29/07/11	06/09/11
Garganega	23/06/11	15/07/11	05/08/11	13/09/11
Glera	21/06/11	12/07/11	09/08/11	06/09/11
Moscato bianco	21/06/11	12/07/11	17/08/11	01/09/11
Passerina	23/06/11	15/07/11	17/08/11	12/09/11

Supplementary table 2: Brix degree value of each variety at the two ripening stages. Each Brix degree value represents the mean ± standard deviation of three biological replicates.

Variety	End of véraison	Harvest
Sangiovese	13.3 ± 0.2	19.3 ± 0.7
Barbera	14.3 ± 0.4	22.7 ± 0.2
Negro amaro	14.2 ± 0.3	21.8 ± 0.6
Refosco	13.8 ± 0.3	23.3 ± 0.5
Primitivo	14.1 ± 0.1	18.1 ± 0.2
Vermentino	9.6 ± 0.3	18.7 ± 1.2
Garganega	11.4 ± 0.5	17.7 ± 0.5
Glera	11.5 ± 0.5	18.4 ± 0.7
Moscato bianco	11.9 ± 1.4	20.9 ± 0.8
Passerina	13.1 ± 0.9	15.9 ± 0.4

REFERENCES

1. Jackson R.S. Grapevine Structure and Function, Chapter 3 in Wine Science, 2nd Edition. San Diego: Academic Press. pp. 66-71 (2000)
2. Kennedy J. Understanding grape berry development. Practical Winery & Vineyard (2002)
3. Pucheu-Plante B. & Mercier M. Etude ultrastucturale de l'interrelation hôte-parasite entre le raisin et le champignon Botrytis cinerea: exemple de la pourriture noble en Sauternais. Can. J. Bot. 61:1785-1797 (1983)
4. Blanke M.M. & Leyhe A. Stomatal and cuticular transpiration of the cap and berry of grape. J. Plant Physiol. 132:250-253 (1988)
5. Blaich R., Stein U., Wind R. Perforation in der Cuticula von Weinbeeren als morphologischer Faktor der Botrytisresistenz. Vitis 23:242-256 (1984)
6. Rogiers S.Y, Hatfield J.M, Jaudzems V.G, G. White R.G, Keller M. Grape berry cv. Shiraz epicuticular wax and transpiration during ripening and preharvest weight loss. Am. J. Enol. Vitic. 55(2):121-127 (2004)
7. Sweetman C., Deluc L.G., Cramer G.R., Ford C.M., Soole K.L. Regulation of malate metabolism in grape berry and others developing fruits. Phytochemistry 70(11-12):1329-1344 (2009)
8. Hardie W.J., O'Brien T.P., Jaudzems V.G. Morphology, anatomy and development of the pericarp after anthesis in grape, *Vitis vinifera* L. Aust. J. Grape Wine Res. 2:97:142 (1996)
9. Fontes N., Côrte-Real M., Gerós H. New observations on the integrity, structure, and physiology of fresh cells of fully ripened grape berry. Am. J. Enol. Vitic. 62:279:284 (2011)
10. Bioletti F. "Outline of ampelography for the Vinifera grapes in California," Hilgardia 11:227-293 (1938)
11. Coombe B.G. Research on the development and ripening of the grape berry. Am. J. Enol. Vitic. 43:101-110 (1992)
12. Conde C., Silva P., Fontes N., et al. Biochemical changes throughout grape berry development and fruit and wine quality. Food 1:1-22 (2007)
13. Romeyer F.M., Macheix J.J., Goiffon J.P., Reminiac C.C., Sapis J.C. The browning capacity of grapes. 3. Changes and importance of hydroxycinnamic acid-tartaric acid esters during development and maturation of the fruit. J. Agric. Food Chem. 31:346-349 (1983)
14. Zoecklein B.W., Fugelsang K.C., Gump B.H., Nury F.S. "Wine analysis and production", New York, Kluwer Academic/Plenum (1999)
15. Jaillon O., Aury J.M., Noel B., Policriti A., Clepet C., et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-467 (2007)
16. Howe K.L., Chothia T., Durbin R. GAZE: a generic framework for the integration of gene-prediction data by dynamic programming. Genome Res. 12(9):1418-1427 (2002)
17. Grimplet J., Van Hemert J., Carbonell-Bejerano P., Díaz-Riquelme J., Dickerson J., Fennell A., Pezzotti M., Martínez-Zapater J.M. Comparative analysis of grapevine whole-genome gene predictions, functional annotation, categorization and integration of the predicted gene sequences. BMC Res Notes. 5:213 (2012)

18. Allen J.E., Salzberg S.L. JIGSAW: integration of multiple sources of evidence for gene prediction. Bioinformatics 21(18):3596-3603 (2005)
19. Forcato C. Gene prediction and functional annotation in the *Vitis vinifera* genome. PhD Thesis, Universita' Degli Studi Di Padova (2010)
20. Vitulo N., et al. A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. BMC Plant Biology 14:99 (2014)
21. Lorenz D.H., Eichhorn K.W., Bleiholder H., Klose R., Meier U., Weber E.. Phenological growth stages and BBCH-identification keys of grapevine (*Vitis vinifera* L. ssp. vinifera). Vitic. Enol. Sci. 49:66-70 (1994)
22. Fasoli M., Dal Santo S., Zenoni S., Tornielli G.B., Farina L., Zamboni A., Porceddu A., Venturini L., Bicego M., Murino V., Ferrarini A., Delledonne M., Pezzotti M. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. Plant Cell 24:3489-3505 (2012)

Chapter 4

Gene expression analysis of the ten varieties during berry development

INTRODUCTION

Fruit ripening depends of global transcriptional programs. Therefore, defining the subsets of expressed genes and their expression patterns is crucial to understand the transcriptional dynamic occurring during fruit development. The recent technological advances in next-generation sequencing methods[1,2] have been applied to transcriptomics creating the novel high-throughput technology named RNA-sequencing (RNA-Seq)[3]. RNA-Seq technology has already been used to study fruit development of many crops as *Solanum lycopersicum* (tomato)[4], *Capsicum annuum* L (chili pepper)[5], *Myrica rubra* (Chinese bayberry)[6], *Pyrus ussuriensis* (Chinese pear)[7], *Citrus reticulate* (Ponkan fruit)[8], *Mangifera indica* Linn (mango fruit)[9] and watermelon[10]. With regard to *Vitis vinifera* L., four previous studies were carried out on cultivars Corvina[11], Shiraz[12,14], Tannat[13] and Cabernet Sauvignon[14] berries. These transcriptomic studies were done only for one[11-13] or two varieties[14], and the developmental stages chosen differed from one study to the others making any comparison quite imprecise. Furthermore, two of them[11,12] were performed using only one biological replicate that does not embrace the necessary statistical representation of the experimental biological variability thus they have been mainly considered methodological breakthrough.

In this work, the first transcriptomic study of berry development on ten different grapevine varieties, five red and five white, was carried out by performing a RNA-Seq assay on berry samples collected at the same four phenological stages in biological triplication. The gene expression analyses included in this chapter focused on studying the global expression profile in different varieties and conditions. Both the total number of genes expressed in each variety, and skin colour- and variety-specific transcripts were identified. Moreover transcriptional modulation patterns during development were defined by performing differential gene expression analysis. In order to understand the relationship between the ten varieties in different growth stages in terms of gene expression profiles, the transcriptome correlation was valuated.

RESULTS AND DISCUSSION

Gene expression analysis from RNA-Seq reads

From the alignment of the reads, obtained by RNA-sequencing of 120 total RNA samples, onto the reference genome PN40024 (12x)[15], a normalized averaged expression value per transcript (FPKM – Fragments Per Kilobase of Mapped reads) was calculated for each triplicate using the geometric normalization method (Cuffmerge program from Cufflinks package[16]). In total, 26,807 genes were detected as expressed – namely genes having an averaged FPKM value greater than 0 in at least one sample – representing ~89.4 % of the V1 predicted grapevine genes[17] (http://genomes.cribi.unipd.it/DATA) (**Fig. 1**).

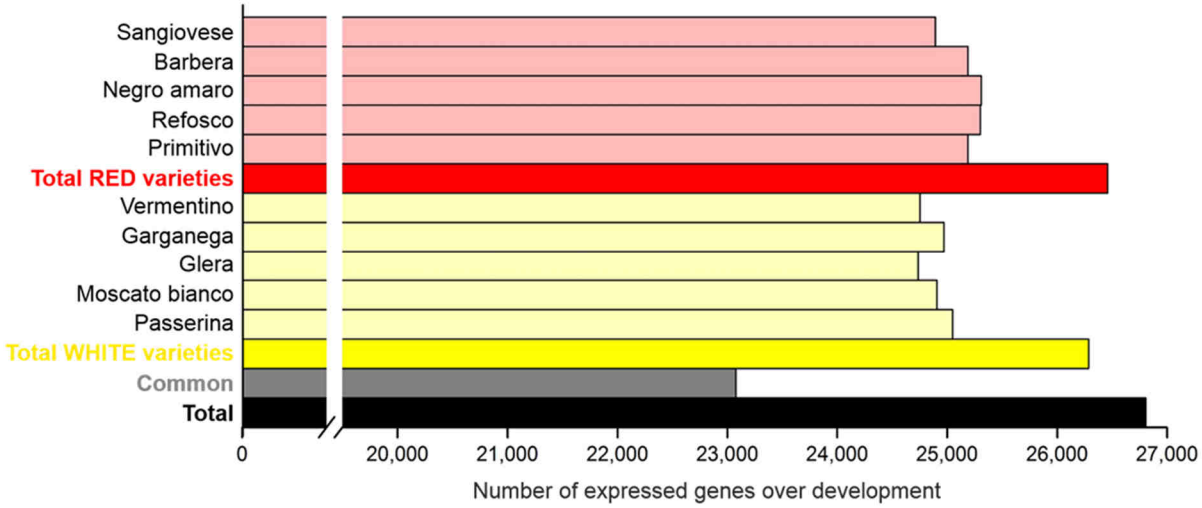


Figure 1: Number of transcripts over development in each variety. **Total RED varieties**: Number of expressed genes in at least one red-coloured variety (26,459). **Total WHITE varieties**: Number of expressed genes in at least one white-coloured variety (26,285). **Total**: Number of expressed genes in at least one variety (26,807). **Common**: genes expressed in each variety at least at one growth stage (23,079).

These numbers were calculated on the basis that the average value of the biological triplicate normalized by Cuffmerge of a gene had to be greater than 0.

Comparing to other RNA-Seq assays, this result corresponds to almost 9,500 genes more than the total number of expressed genes in Corvina variety[11] and ~6,000 ones more than in Shiraz cultivar[12]. Hence, carrying out RNA-Seq assay with more varieties and more biological replicates allowed covering much more transcriptome and variety-specific patterns. Additionally, it is worth to notice that the number of genes expressed over berry development did not change much among the ten varieties. Nevertheless,

red-skinned berries, except for Sangiovese berry development, expressed slightly more genes than the white ones (~1%). In total, 23,079 genes were found to be expressed during berry development of all ten varieties.

This RNA-Seq assay produced a wide range of gene expression values (from 2.65E-11 to 174,049 FPKM). However, Warden and colleagues[18] showed that genes with a low coverage of reads could produce artificially high fold-change values between samples, although the true expression levels might not differ greatly. In order to avoid this problem, they proposed to define a minimum reliable threshold for FPKM values comprised between 0.01 and 1. In order to avoid using an arbitrary-chosen threshold, the effect of different cutoffs on data distribution was exploited[19]. The density plots of the 10-variety transcriptomes at each stage after applying four FPKM cutoffs (0.01, 0.1 and 1) showed that a threshold of 1 FPKM allowed the data to be the closest to normal-distributed data (**Fig. 2**).

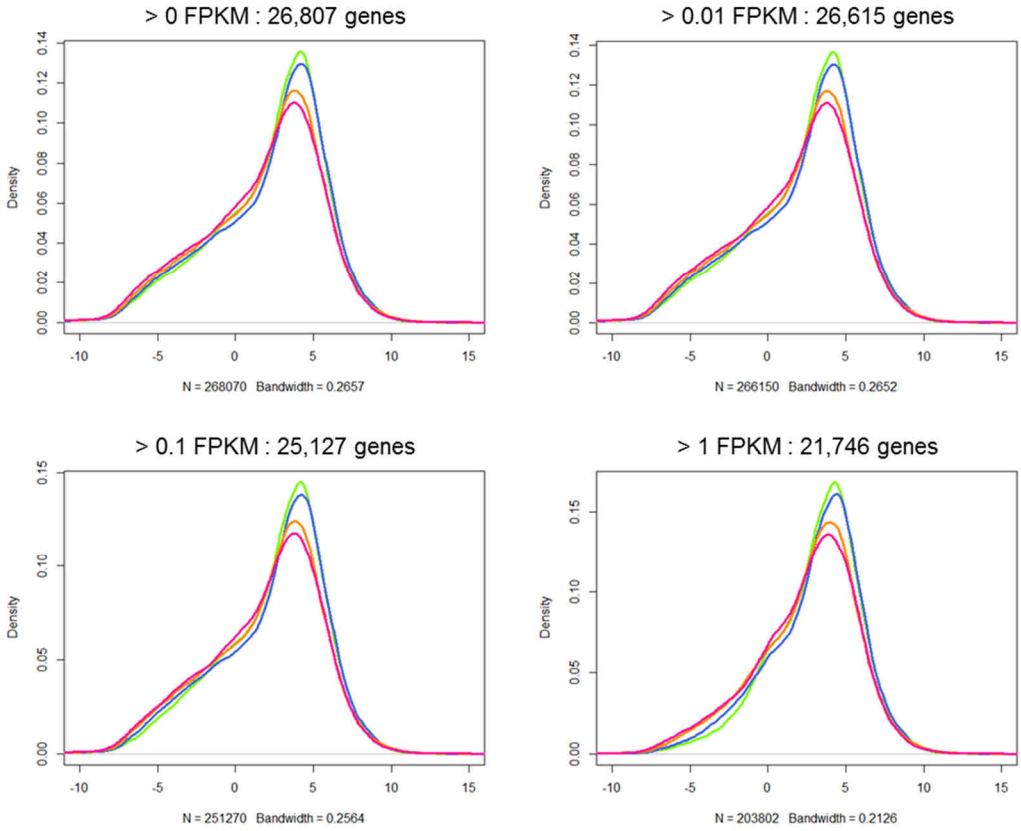


Figure 2: Density plots showing the distribution of log_2(FPKM) values obtained after applying a FPKM cutoff of 0.01; 0.1 and 1 for the 10 variety transcriptomes at each stage: Pea-sized berry stage (green), Prevéraison (blue), End of véraison (orange), Harvest (pink). The density plots were created using R software.

The choice of the optimal cutoff was not obvious, however this seemed like a reasonable criterion to use. Consequently, genes showing an expression intensity smaller than 1 FPKM in all 40 samples were classified as very low-expressed and then removed from the gene expression dataset. The resulting dataset was composed of 21,746 transcripts and used in the following analyses. The 5,061 very low-expressed genes are described in **chapter 10**.

Based on the 21,746-transcript dataset, the analysis of gene expression activity of each variety at each growth stage showed that the number of expressed genes slightly decreased after véraison for all varieties (**Fig. 3**).

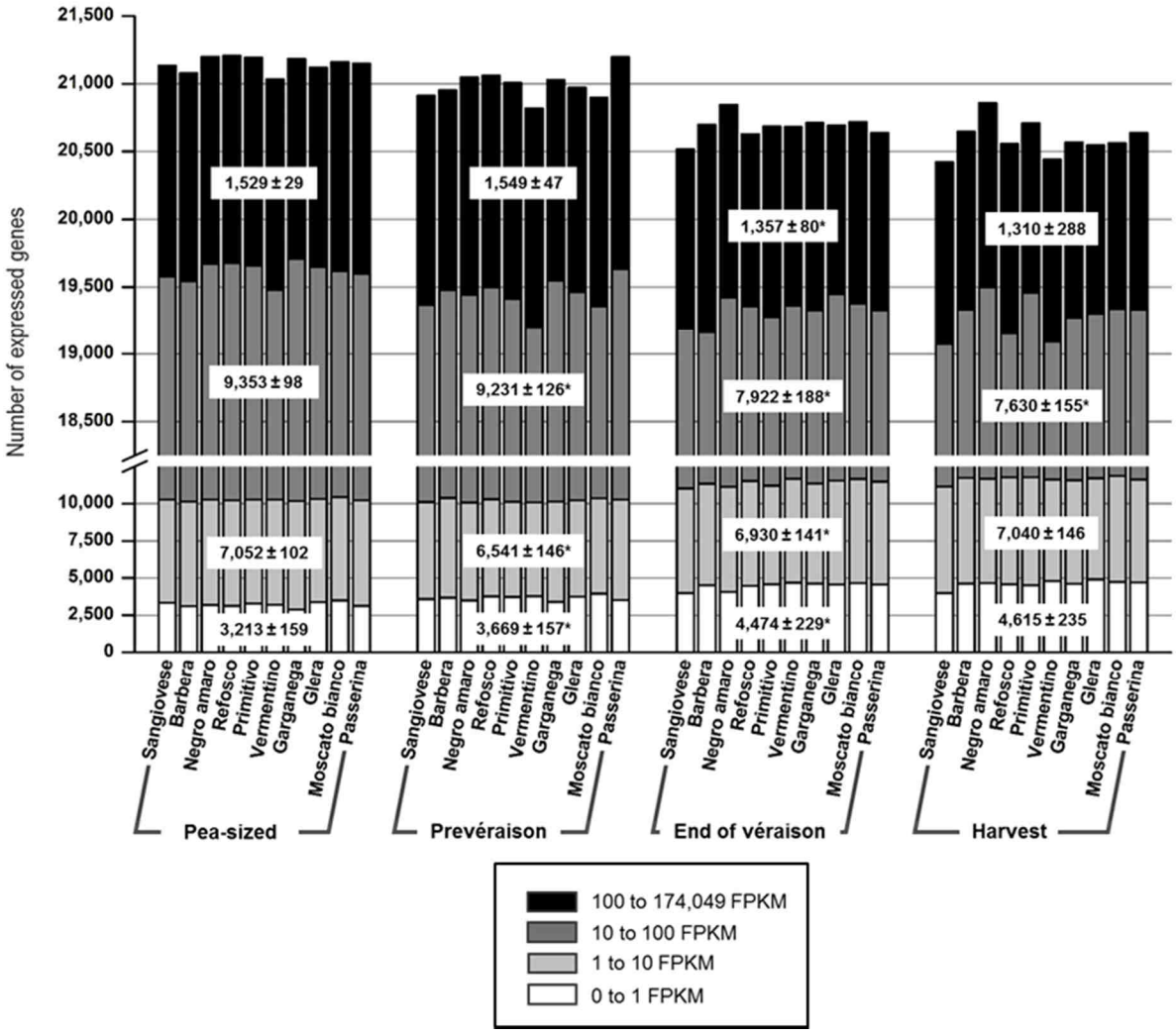


Figure 3: Number of expressed genes at each phenological stage in each variety. Transcripts were attributed to each selected FPKM intervals on the basis of the average value of the biological triplicate normalized by Cuffmerge. The average number of expressed genes of each RPKM interval and the corresponding standard deviation are represented in the white squares; * shows significant differences at 1% between each subsequent pair of phenological stages.

This decline seemed to affect more particularly genes with a high and medium expression value (> 100 and 10 FPKM respectively) combined to an increase of the low- and very low-expressed genes (< 10 and 1 FPKM respectively). The same trend was observed for the density plots of the ten varieties performed at each stage (**Supplementary figure 1**). However, the decrease of expressed genes with a medium expression intensity seemed to be gradual over the maturation phase in Negro amaro, Refosco and Glera berries in comparison to the others. These observations suggest that green development engages the expression of more genes but also with a higher expression level than the maturation process. Moreover, in **chapter 6**, a high transcriptional modulation was observed after véraison with more genes down- rather than up-regulated during berry development, which could explain this decline of transcripts after véraison. Finally, this phenomenon of down-regulation/suppression of gene expression seemed to be the trend of any maturation processes in *Vitis vinifera* (**chapter 11**), involving the suppression of vegetative pathways rather than the activation of mature pathways.

Skin colour- and variety-specific transcripts

In order to find out genes expressed only in either red- or white-skinned berries, any gene expressed even only in one white grape variety were removed from the transcript list of the red ones and vice versa. As result, few transcripts were identified as red- (53) or white-skin (36) specific. In addition, most of these genes did not have a functional annotation (37 and 29 for red- and white-skin specific transcripts respectively), making difficult the understanding of the function of these genes. A comparison of these transcripts was done using Venn diagrams (http://bioinformatics.psb.ugent.be/webtools/Venn/) (**Fig. 4**), which allowed also understanding the varietal specificity within the same colour group.

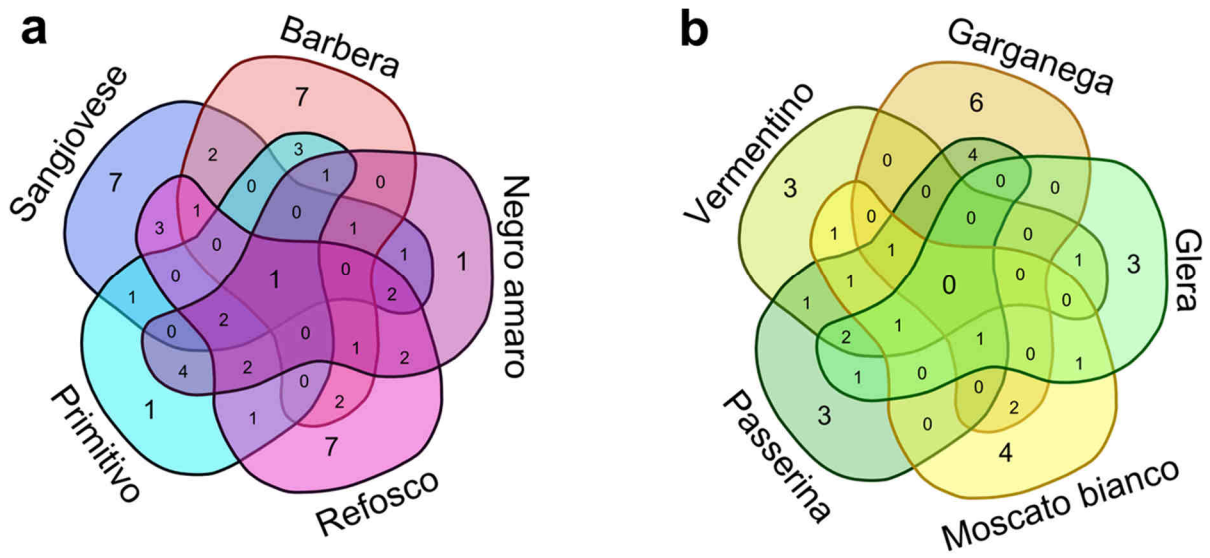


Figure 4: Shared and specific transcripts present only in red-skinned berries (**a**) or white-skinned berries (**b**) during berry development.

Concerning red-skinned berries, only one red-grape-specific transcript was shared by all the five red grape varieties (VIT_02s0025g01930) (**Fig. 4 a**). This gene coding for a cellulose synthase-like protein G3 was expressed in Sangiovese (3.81 FPKM), in Refosco and Primitivo full-ripe berries, and in Barbera and Negro amaro pea-sized berries at a very low-expression intensity (below 1 FPKM). Cellulose synthase-like proteins belongs to the glycosyltransferase family 2. It was proposed the cellulose synthase-like families might be involved in the synthesis of the backbones of hemicelluloses such as xyloglucan, xylan, mannan and other β-D-glycans in the cell wall[20,21]. Nevertheless, as the expression level is below the set threshold, it may not be considered a commonly expressed gene in the five red-skinned varieties. The other red-grape-specific transcripts were either shared by more varieties or identified as variety-specific. Among the shared ones, we found transcripts coding for a patatin (VIT_00s0567g00040) in Barbera, Negro amaro and Refosco full-ripe berries, a translation initiation factor eIF-4A and a Emp24/gp25L/p24 family protein in Negro amaro and Refosco green berries (VIT_05s0029g00800 and VIT_05s0029g00790 respectively), a glycine hydroxymethyltransferase (VIT_07s0005g06570) in Sangiovese and Refosco ripe berries. Patatin is a glycoprotein found in potato (*Solanum tuberosum*), which mainly functions as a storage protein but it also has lipase activity that could be involved in the cleavage of fatty acids from membrane lipids[22]. Nevertheless, its role in grape berry development is still unknown. Translation initiation factor eIF-4A is required for translation initiation to bind mRNA to 40S ribosomal subunits[23] and Emp24/gp25L/p24 family protein is

involved in intracellular transport[24]. The expression of these genes in Negro amaro and Refosco green berries may be linked to a specific translation activity and protein transport occurring during berry formation. Furthermore, 23 transcripts were identified as variety-specific (**Supplementary table 1**) as an enolase gene (VIT_11s0016g01570) and a receptor kinase 2 gene (VIT_19s0014g00490) expressed only in Sangiovese berries, the nitric-oxide synthase 1 gene (VIT_16s0022g00140) expressed only in Barbera grapes and a lipase GDSL gene (VIT_14s0066g01260) only expressed in Refosco berries. However, all these function-predicted genes were low-expressed, so we can wonder if their gene expression specificity is relevant at a physiological level.

Among the 29 white-skin-specific transcripts, none was shared by the five white-grape varieties and only seven were functionally annotated. Concerning the shared transcripts, we identified an ankyrin repeat protein gene (VIT_07s0151g00040) and an argonaute gene (VIT_01s0010g01240) orthologous to *AtAGO1* expressed in Garganega and Passerina berries, a double-stranded DNA-binding protein gene (VIT_00s0163g00070) expressed in Vermentino and Glera berries and the 12-oxophytodienoate reductase 1 gene (VIT_18s0041g02090) involved in alpha-linolenic acid metabolism expressed in Vermentino, Glera and Passerina berries. In *Arabidopsis thaliana*, AtAGO1[25] is associated with numerous microRNAs (miRNAs) and short interfering RNAs (siRNAs) to mediate target repression via mRNA cleavage and inhibition of translation[26-28]. This gene is involved in many processes as leaf proximal-distal, venation and stomatal patterns[29], auxin and light control of adventitious rooting[30], and plant defence[31] but its role in *Vitis vinifera* is not well understood. However, this gene showed a similar pattern of expression in Garganega and Passerina berries (**Table 1**). Therefore, we can suppose that this encoded protein could have a similar and specific role in these two grape varieties.

Table 1: Expression intensity (FPKM average value) of VIT_01s0010g01240 Argonaute gene in Garganega and Passerina varieties during berry development.

Variety	Pea-sized	Prevéraison	End of véraison	Harvest
Garganega	0.177813	1.45875	7.19521	4.30352
Passerina	0.165152	1.49112	4.25119	2.76961

In addition, 19 variety-specific transcripts were identified, as a polyprotein gene (VIT_15s0024g01360) expressed only in Moscato bianco pea-sized berries, the Heat

shock 70kDa protein 5 gene (VIT_15s0046g02740) and an acyl-[acyl-carrier-protein] desaturase gene (VIT_00s0327g00070) expressed only in Passerina berries.

Transcriptional modulation analysis

Differentially expressed genes (DEGs) between each subsequent pair of developmental stages (i.e. Pea-sized berry-Prevéraison, Prevéraison-End of véraison, End of véraison-Harvest) were identified using Cuffdiff v2.0.2[32,33] with a False-Discovery Rate of the Benjamini-Hochberg multiple tests of 5%. The analysis was performed for all the 10 varieties. The results indicated that berry development engages the expression modulation of a large number of genes – from 12,137 DEGs in Passerina to 16,999 ones in Barbera – which is in line with the number of DEGs observed during development of Corvina berries[11]. In order to visualize the transcriptional modulation occurring in each variety during berry development, the number of genes significantly modulated in each subsequent comparison – pairwise DE analysis of developmental stages – was plotted for each grape variety (**Fig. 5**). **Figure 5** shows that grape varieties followed different dynamics of gene expression modulation.

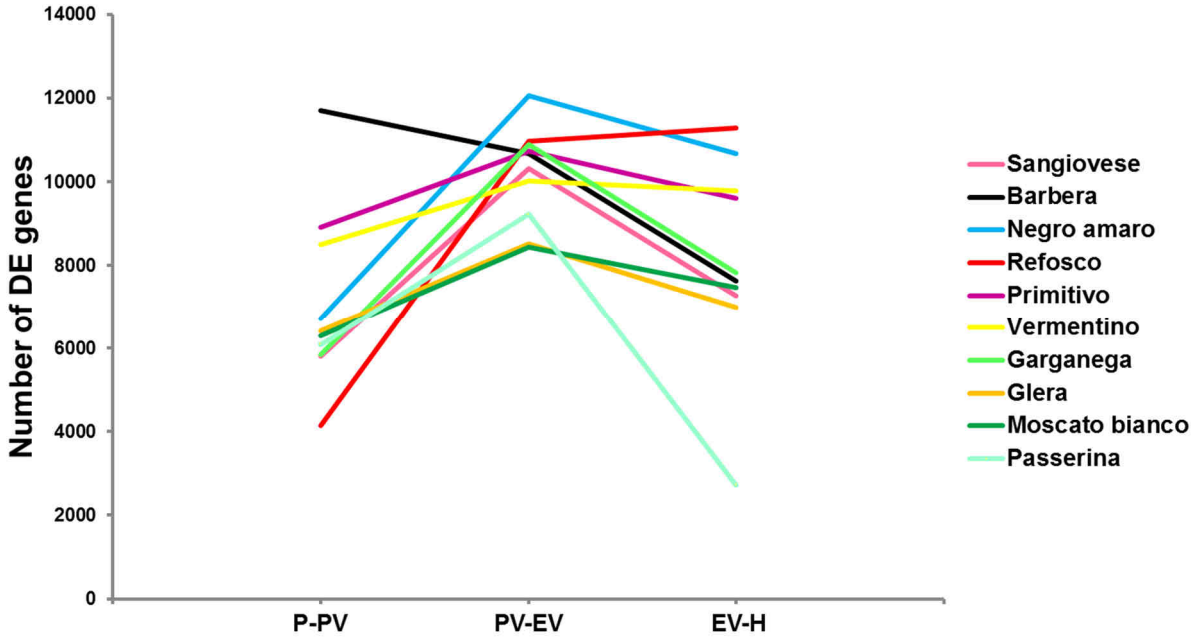


Figure 5: Number of DEGs of each grape variety between each subsequent pair of developmental stages: Pea-sized berry-Prevéraison (P-PV), Prevéraison-End of véraison (PV-EV), End of véraison-Harvest (EV-H).

The profiles could be grouped by similarity and then common trends of gene modulation could be recognized for some varieties. For Refosco, the gene modulation profile consisted of a marked increase of the number of DEGs from prevéraison to the end of véraison – from 4,258 to 11,045 DEGs – followed by slight rise towards harvest – 11,384 DEGs. A second dynamic also concerned a high increase at the end of véraison but in this case followed by an important decline from the end of véraison to harvest for Negroamaro, Garganega, Sangiovese and Passerina. The third dynamic was similar to the second one but with a smaller amplitude in the case of Primitivo, Vermentino, Glera and Moscato bianco. Finally, Barbera exhibited a completely different profile with a decreasing number of DEGs over development, from 11,903 at prevéraison stage to 7,687 at harvest. This last profile differed from the others for the high number of DEGs between pea-sized berries and prevéraison stage. Indeed, 3,030 Barbera berry-specific DEGs (1,434 up- and 1,596 down-regulated genes) between the two first growth stages were identified. This result suggests either that Barbera berries exhibited a higher transcriptional activity during green development or that pea-sized or prevéraison berries of Barbera cultivar were not collected at the exactly same growth stages than the other varieties, which could explain the higher gene modulation in the first comparison.

Variety and growth stage transcriptome relationships

In order to compare forty sample transcriptomes, a correlation matrix was generated using Pearson's coefficient as distance metric (**Fig. 6 a**). The matrix showed a strong correlation between samples of the first two stages, and between end of véraison and harvest, showing an evident distinction of the transcriptional traits before and after véraison, independently on skin colour and variety. A dendrogram (**Fig. 6 b**) – generated converting correlation coefficients in distance values – pointed out that in pre-véraison phase varieties strongly clustered together independently of the skin colour. After véraison, skin colour seemed to influence Refosco, Barbera and Negro amaro berry transcriptomes that clustered separately from the other varieties at harvest, suggesting that these three red grape varieties might have specific transcriptional traits at maturity. Moreover, we noticed that Pearson correlation value between Barbera and the other varieties during the green phase (**Fig. 6 a**) was very high and similar to the other ones despite a higher DEG number at prevéraison stage.

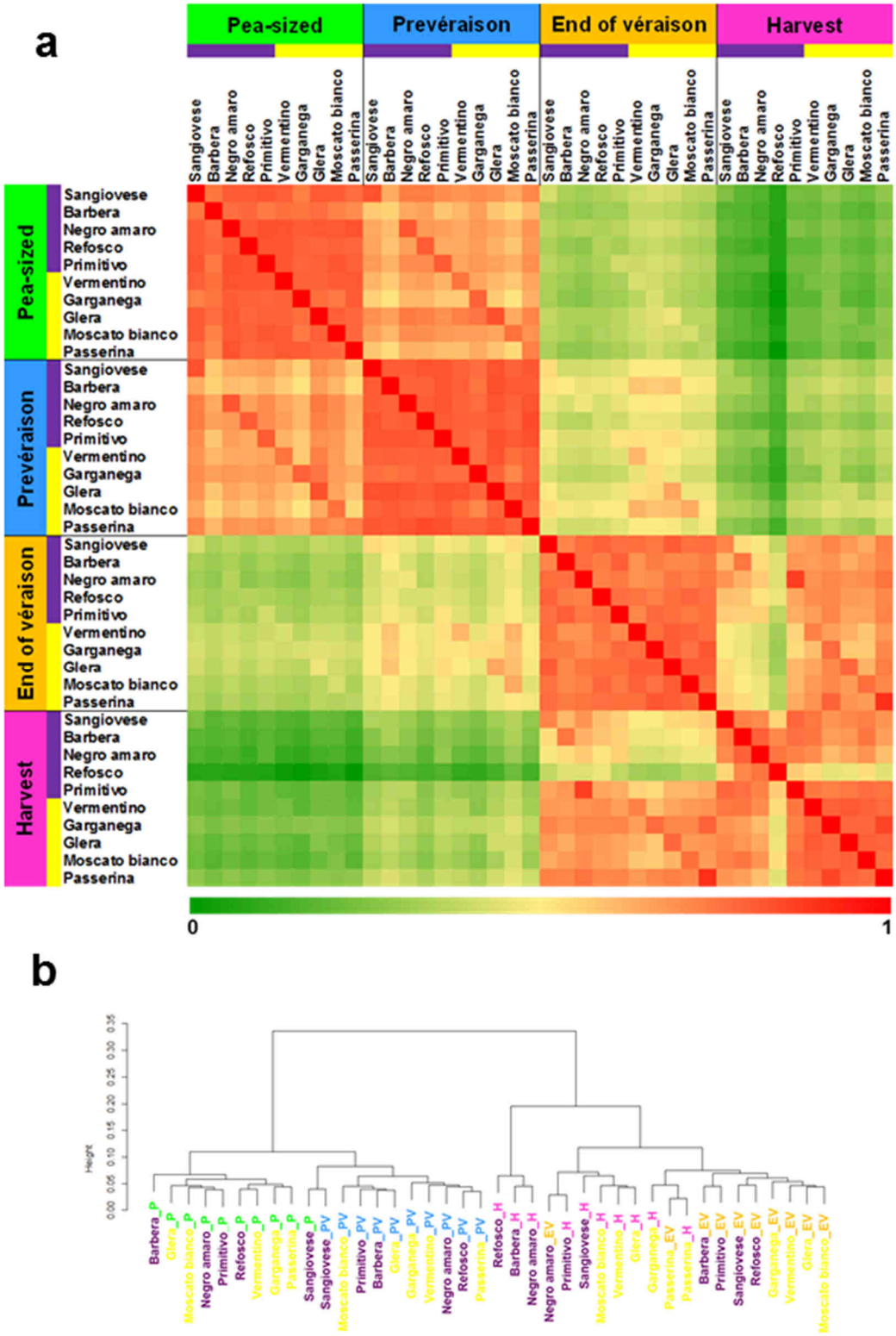


Figure 6: (**a**) Correlation matrix of the 40-sample data set. The analysis was performed by comparing the values of the whole transcriptome in all 40 samples, using the $\log_2$-transformed FPKM values and Pearson's distance as the metric. Correlation analysis was performed using R software.
(**b**) Cluster dendrogram of the whole data set. The Pearson's correlation values were converted into distance coefficients to define the height of the dendrogram.

Nonetheless, we also noticed that Barbera was separated from the other varieties in the pea-sized berry stage-cluster of the dendrogram (**Fig. 6 b**). This result suggests either that Barbera pea-sized berries exhibited a specific transcriptome at this stage or that Barbera pea-sized berries were not collected exactly at the same growth stage than the other varieties.

CONCLUSION

Overall, gene expression of 10-variety berries at four growth stages was analysed, allowing detecting the expression of about 89 % of the V1 predicted grapevine genes[17]. In order to better handle the wide range of gene expression level, a cutoff of 1 FPKM was applied to the 40-sample data set resulting in the creation of a novel data set composed of 21,746 transcripts. The analysis of this data set showed that transcript number slightly decreased after véraison for all the ten varieties affecting particularly genes with a high and medium expression level (> 100 and 10 FPKM respectively).

We observed that skin colour- and variety-specific transcripts were only a few and devoid of a predictive functional annotation for the most of them. No hit and unknown-function transcripts could represent good candidates for further functional characterization analysis in order to discover if their expression is responsible for varietal physiological and morphological characteristics as berry shape and size or flavour production and composition.

Differentially gene expression analysis between each subsequent pair of growth stages showed that eight of the ten varieties presented a similar transcriptional modulation profile – namely an increase of DEG numbers at the end of véraison followed by a decrease of them at harvest – whereas Barbera exhibited a decreasing profile and Refosco an increasing profile. However, the higher DEG number of Barbera berries at prevéraison stage seemed not affect Pearson correlation value between Barbera and the other varieties during the same period (**Fig. 6 a**) but it could explain the separation between Barbera and the other varieties in the pea-sized berry stage-cluster of the dendogram (**Fig. 6 b**). This result either could be due to a specific transcriptome of Barbera pea-sized berries or might be caused by a sampling issue of Barbera pea-sized berries.

Regarding transcriptome relationship, Pearson correlation matrix (**Fig. 6 a**) showed a strong correlation between berry samples of the first two stages, and

between the two last stages, indicating a deep transcriptome shift occurring at véraison independently on skin colour and variety. Indeed, in **chapter 11**, this green-to-maturation transcriptome shift was defined as a trend of any maturation processes in *Vitis vinifera*, which involves the suppression of diverse metabolic processes related to vegetative growth, and the induction of only a few pathways as secondary metabolic processes and responses to biotic stimuli. Nevertheless, the dendogram revealed a split of Refosco, Negro amaro and Barbera full-ripe berry transcriptomes from the other ones, suggesting that these three red grape varieties might have specific transcriptional traits at full-maturity.

METHODS

Sequencing Data Analysis

Single-ended reads of 100 nucleotides generated by the RNA-Seq assay of 120 RNA samples (**chapter 3**) were aligned onto the *Vitis vinifera* reference genome PN40024 (12x)[15] using the program TopHat v2.0.6[34], with default parameters. An average of 85.55% of reads was mapped for each sample (**Supplementary table 2**). Mapped reads were used to reconstruct the transcripts using the program Cufflinks[16] v2.0.2 and the reference genome annotation V1[17] (http://genomes.cribi.unipd.it/DATA). Then, all reconstructed transcripts were merged into a single non-redundant list of 29,287 transcripts, using the program Cuffmerge, from the Cufflinks package. Basically, Cuffmerge program merges transcripts whose reads are overlapping and share a similar exon structure (or splicing structure) and generates a longer chain of connected exons. Using this novel list of transcripts, the normalized averaged expression value of each transcript was calculated as FPKM (Fragments Per Kilobase of Mapped reads) for each triplicate using the geometric normalization method.

Differential gene expression analysis

Differential gene expression analysis was performed using the Cuffdiff program within Cufflinks version v2.0.2[32,33] and the mapping results described above. To do gene expression comparisons between each subsequent pair of developmental stages (i.e. Pea-sized berry-Prevéraison, Prevéraison-End of véraison, End of véraison-Harvest) for each variety, Cuffdiff first calculated the variance of the transcript abundance in each sample accounting both for the variation across replicates using the geometric

normalization method. The variance for a gene abundance was then determined using the variance in abundance for each isoform of that gene. The statistic T-test was calculated as the log of the ratio of FPKM values for the two samples divided by the variance of the log of the ratio. A modified Student's t-test was then used to determine if genes were differentially expressed. A False-Discovery Rate of 5% after Benjamini-Hochberg correction for multiple testing was applied in order to obtain a q-value for each test[35]. Dr. Mario Altieri realized the above-described bioinformatics analyses.

Density plots

Density plots of **Figure 2** were created from FPKM average values of each developmental stage 10-sample transcriptome after applying the subsequent cutoffs using R software. Density plots of each variety (**Supplementary Fig. 1**) were created from 21,746 FPKM values for each developmental stage by using the following R script (Example for Sangiovese).

```
>rawdata = read.csv(file="RPKM_Sangiovese.txt",sep="\t", header=T, row.names=1)
>data= log2(rawdata)

>plot(density(data[,1]), col="lawngreen", xlim=c(-10,15), lwd=3)
>lines(density(data[,2]), col="royalblue", lwd=3)
>lines(density(data[,3]), col="darkorange", lwd=3)
>lines(density(data[,4]), col="deeppink", lwd=3)
```

Correlation Analysis

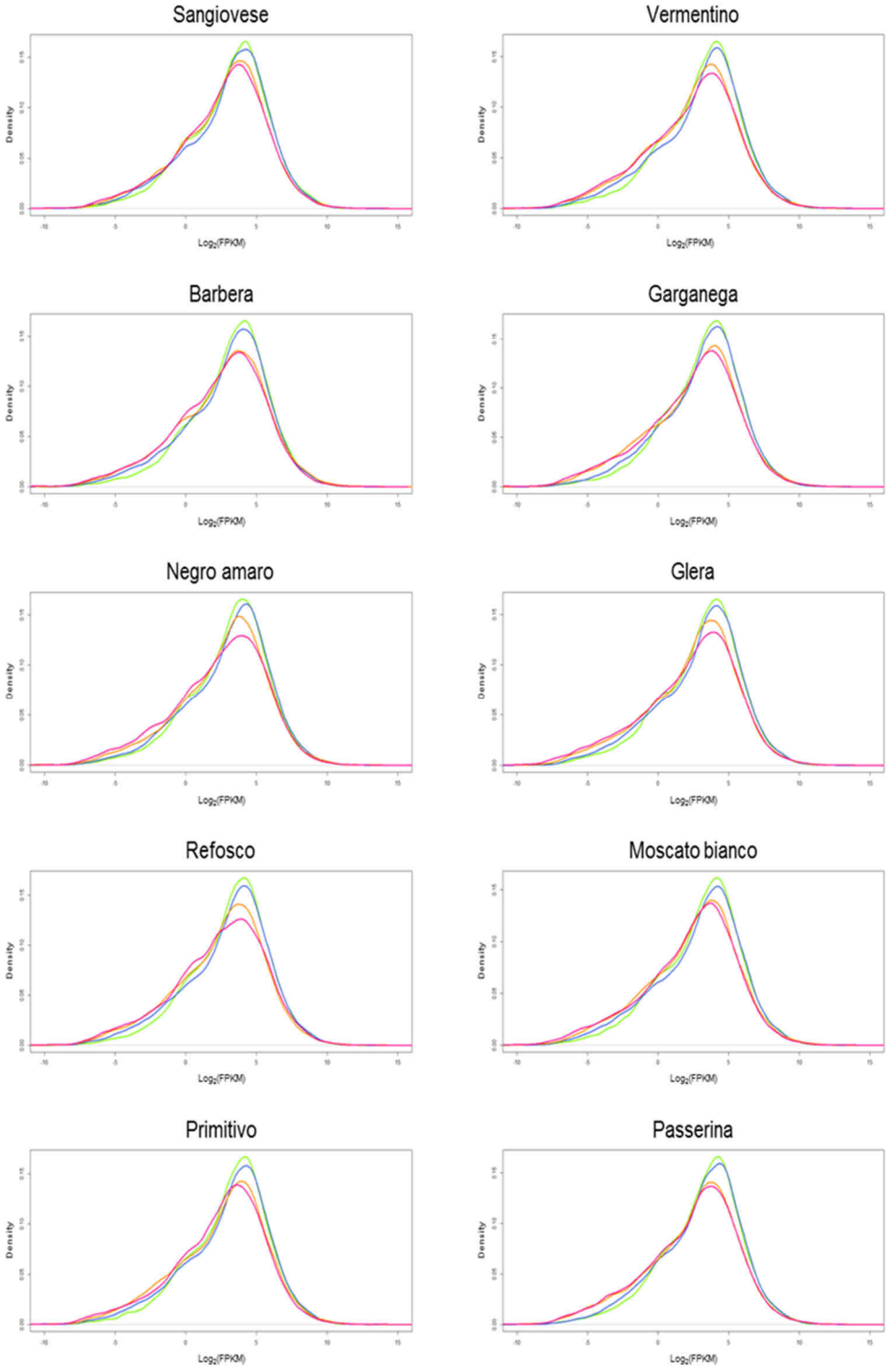
A correlation matrix was prepared using R software and Pearson's correlation coefficient as the statistical metric to compare the values of the whole transcriptome in all 40 samples (21,746-transcript data set), using the $\log_2$(FPKM mean + 1). Correlation values were converted into distance coefficients to define the height scale of the dendrogram. Correlation matrix and dendogram were created by using the following R script:

```
>rawdata = read.csv(file="file_name",sep="\t")
>data=rawdata[,2:x]
>rownames(data)=rawdata[,1]
```

```
>cormatrix = cor(data)
>write.csv(cormatrix,file="cormatrix")

>plot(hclust(as.dist(1-cor(data))),cex=1)
>savePlot ("dendrogram", type="tiff")
```

SUPPLEMENTARY MATERIAL

Supplementary Figure 1: Density plots showing the distribution of log_2(FPKM) values obtained for each variety at each stage: pea-sized berry (green), prevéraison (blue), end of véraison (orange), harvest (pink). Density plots were created using R software.



Supplementary Table 1: Variety-specific transcripts.

Variety	Gene_ID	Functional annotation	P	PV	EV	H
Sangiovese	VIT_10s0003g01950	No hit	0	0	0	7.811
	VIT_12s0059g01840	No hit	0	0	0	3.905
	VIT_19s0015g01240	No hit	0	0	0	2.752
	VIT_11s0016g01570	Enolase	2.041	0.759	0	0.0744
	VIT_10s0003g04520	No hit	1.712	0	0	0
	VIT_06s0004g02150	No hit	1.687	0	0	0
	VIT_19s0014g00490	Receptor kinase 2	1.668	0	0	0
Barbera	VIT_10s0003g04140	No hit	0	6.749	1.541	3.843
	VIT_16s0148g00270	No hit	6.201	0	0	0
	VIT_07s0005g06370	No hit	0	0	3.143	3.923
	VIT_18s0001g15180	No hit	3.885	0	0	0
	VIT_19s0140g00220	No hit	0	0	2.352	0
	VIT_16s0022g00140	Nitric-oxide synthase 1 (AtNOS1)	0	1.228	0	0
	VIT_19s0015g02850	No hit	0	1.122	0	0
Negro amaro	VIT_15s0048g02790	No hit	0	0	1.321	0
Refosco	VIT_02s0025g02380	No hit	0	0	76	0
	VIT_13s0064g01670	No hit	40.304	0	0	0
	VIT_09s0054g01320	No hit	1.582	6.781	0	2
	VIT_12s0034g00700	No hit	0	3.344	0	0
	VIT_03s0038g00520	No hit	0.303	0.391	0	2.531
	VIT_06s0080g01030	No hit	0	1.246	0	0.742
	VIT_14s0066g01260	Lipase GDSL	1.065	0	0	0
Primitivo	VIT_19s0014g00820	No hit	0	2.428	0	0
Vermentino	VIT_14s0060g00660	No hit	200.294	0	0	0
	VIT_04s0008g04680	No hit	28.944	0	0	0
	VIT_18s0089g00890	No hit	0	0	0	1.202
Garganega	VIT_17s0000g02380	No hit	10.091	0	0	30.073
	VIT_18s0001g01150	No hit	10.484	0	0	0
	VIT_18s0001g08400	No hit	9.391	0	0	0
	VIT_14s0036g01300	No hit	0	4.414	0	0
	VIT_02s0154g00270	No hit	0.463	2.56	1.072	0
	VIT_16s0022g01370	No hit	0	0	0	1.127
Glera	VIT_15s0024g01550	No hit	0.596	0	0	1.902
	VIT_01s0127g00110	No hit	0	1.258	0	0
	VIT_04s0008g07200	No hit	1.061	0	0	0
Moscato bianco	VIT_05s0051g00430	No hit	3.184	0	7.149	9.468
	VIT_15s0024g01360	Polyprotein	7.831	0	0	0
	VIT_04s0023g02910	No hit	0	0	0	7.085
	VIT_02s0012g02440	No hit	0	1.931	0	0
Passerina	VIT_19s0014g03860	No hit	0	0	16.242	0
	VIT_15s0046g02740	Heat shock 70kDa protein 5	4.619	1.781	0.22	0.109
	VIT_00s0327g00070	Acyl-[acyl-carrier-protein] desaturase	0	1.599	0	0.706

Supplementary table 2: Read features of the 120 sequenced samples.

Variety	Developmental Stage	Replicate	Total reads	Mapped Reads	Average reads	Average mapped reads
Sangiovese	Pea-sized berry	1	30,424,284	24,903,276	35,361,610	28,761,685
		2	33,364,279	27,096,778		
		3	42,296,267	34,285,000		
	Prevéraison	1	28,322,782	23,050,616	29,290,926	23,904,188
		2	30,196,178	24,643,257		
		3	29,353,817	24,018,692		
	End of véraison	1	18,946,189	12,909,984	22,842,624	15,378,135
		2	26,455,515	17,835,227		
		3	23,126,169	15,389,194		
	Harvest	1	31,887,707	21,065,690	30,886,738	20,635,076
		2	32,903,151	21,904,339		
		3	27,869,357	18,935,198		
Barbera	Pea-sized berry	1	30,540,721	26,774,602	32,096,887	27,837,539
		2	31,155,671	26,862,428		
		3	34,594,268	29,875,588		
	Prevéraison	1	37,407,969	32,320,057	37,126,294	32,196,851
		2	33,889,396	29,197,255		
		3	40,081,517	35,073,242		
	End of véraison	1	26,258,398	23,054,109	33,971,705	29,587,204
		2	46,299,820	40,169,568		
		3	29,356,896	25,537,935		
	Harvest	1	33,013,166	28,560,867	34,961,301	30,219,550
		2	29,816,915	25,690,191		
		3	42,053,822	36,407,593		
Negro amaro	Pea-sized berry	1	31,246,381	27,219,497	36,615,736	31,936,444
		2	33,211,052	29,034,628		
		3	45,389,774	39,555,207		
	Prevéraison	1	34,112,134	30,471,759	33,204,203	29,638,379
		2	31,666,051	28,221,968		
		3	33,834,423	30,221,411		
	End of véraison	1	28,935,792	24,887,081	32,065,530	27,565,001
		2	32,120,870	27,343,169		
		3	35,139,927	30,464,752		
	Harvest	1	36,191,903	31,458,395	38,740,774	33,822,747
		2	45,433,034	39,708,169		
		3	34,597,384	30,301,677		
Refosco	Pea-sized berry	1	27,016,935	23,431,033	35,632,158	30,842,795
		2	37,663,524	32,483,077		
		3	42,216,016	36,614,274		
	Prevéraison	1	35,647,744	31,110,575	36,322,254	31,721,225
		2	33,532,597	29,349,922		
		3	39,786,420	34,703,179		
	End of véraison	1	27,335,956	22,663,223	29,915,903	24,830,476
		2	33,996,348	28,221,195		
		3	28,415,405	23,607,011		
	Harvest	1	31,357,178	25,490,831	31,326,102	25,019,893
		2	32,426,013	25,693,231		
		3	30,195,115	23,875,617		
Primitivo	Pea-sized berry	1	27,023,286	23,649,830	35,334,678	31,237,590
		2	28,198,631	25,076,693		
		3	50,782,117	44,986,246		
	Prevéraison	1	35,866,946	31,616,019	31,750,847	28,052,520
		2	30,504,750	27,164,591		
		3	28,880,846	25,376,949		
	End of véraison	1	28,833,723	25,038,141	30,775,656	26,674,054
		2	29,279,416	25,408,967		
		3	34,213,829	29,575,054		
	Harvest	1	32,564,412	27,757,804	38,666,916	32,931,704
		2	43,611,853	36,989,000		
		3	39,824,483	34,048,308		

(Table continues on following page)

Supplementary table 2 (Continued from previous page)

Variety	Developmental Stage	Replicate	Total reads	Mapped Reads	Average reads	Average mapped reads
Vermentino	Pea-sized berry	1	32,859,906	28,401,555	30,202,921	25,961,822
		2	29,845,815	25,424,997		
		3	27,903,042	24,058,913		
	Prevéraison	1	29,281,688	25,497,993	29,509,863	25,583,833
		2	29,515,472	25,974,373		
		3	29,732,428	25,279,132		
	End of véraison	1	29,963,367	25,832,014	32,685,540	28,235,060
		2	38,314,116	33,138,656		
		3	29,779,137	25,734,509		
	Harvest	1	32,541,338	26,550,836	33,750,340	28,029,079
		2	32,275,151	27,094,774		
		3	36,434,531	30,441,627		
Garganega	Pea-sized berry	1	26,745,783	22,196,172	31,039,955	26,370,030
		2	34,930,815	29,727,322		
		3	31,443,266	27,186,595		
	Prevéraison	1	31,428,846	27,172,548	32,962,599	28,482,564
		2	29,777,091	25,722,268		
		3	37,681,860	32,552,875		
	End of véraison	1	34,360,728	30,157,657	29,957,574	25,975,566
		2	27,435,887	23,481,549		
		3	28,076,107	24,287,493		
	Harvest	1	30,689,527	26,270,935	31,504,911	26,902,300
		2	30,122,784	25,519,317		
		3	33,702,422	28,916,647		
Glera	Pea-sized berry	1	29,792,740	25,614,875	36,458,410	31,465,055
		2	31,409,419	27,327,341		
		3	48,173,071	41,452,949		
	Prevéraison	1	30,339,314	26,267,278	30,311,090	26,396,630
		2	28,387,580	24,815,843		
		3	32,206,375	28,106,768		
	End of véraison	1	29,946,975	25,936,764	34,615,330	29,933,415
		2	32,698,172	28,209,311		
		3	41,200,842	35,654,169		
	Harvest	1	39,857,627	34,238,653	42,235,446	36,731,711
		2	49,662,036	43,161,417		
		3	37,186,674	32,795,064		
Moscato bianco	Pea-sized berry	1	32,911,860	28,940,051	39,080,721	34,304,437
		2	38,256,129	33,480,427		
		3	46,074,173	40,492,834		
	Prevéraison	1	31,875,074	27,979,819	34,429,091	30,214,921
		2	35,859,739	31,413,791		
		3	35,552,461	31,251,152		
	End of véraison	1	32,473,567	28,268,808	30,276,053	26,395,199
		2	29,385,127	25,818,254		
		3	28,969,465	25,098,535		
	Harvest	1	26,566,972	22,772,078	35,528,577	30,727,212
		2	47,177,431	40,839,579		
		3	32,841,327	28,569,980		
Passerina	Pea-sized berry	1	30,601,249	27,130,022	33,343,213	29,614,243
		2	36,798,419	32,745,407		
		3	32,629,971	28,967,300		
	Prevéraison	1	34,268,249	30,500,643	32,871,018	29,318,297
		2	36,498,569	32,587,692		
		3	27,846,236	24,866,555		
	End of véraison	1	42,715,900	37,323,963	35,254,834	30,882,388
		2	31,913,129	28,007,618		
		3	31,135,472	27,315,582		
	Harvest	1	31,198,147	26,977,226	34,643,566	29,984,566
		2	35,791,471	31,044,469		
		3	36,941,079	31,932,003		

REFERENCES

1. Margulies M., Egholm M., Altman W.E., et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376-80 (2005)
2. Bennet S. Solexa Ldt. Pharmacogenomics 5(4):433-8 (2004)
3. Mortazavi A., Williams B.A., McCue K., Schaeffer L., Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods 5(7):621-8 (2008)
4. Koenig D., et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc. Natl. Acad. Sci. USA 110, E2655–E2662 (2013)
5. Martínez-López et al. Dynamics of the chili pepper transcriptome during fruit development. BMC Genomics 15:143 (2014)
6. Feng C., Chen M., Xu C.J., Bai L., Yin X.R, Li X., Allan A.C., Ferguson I.B., Chen K.S. Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. BMC Genomics 13:19 (2012)
7. Huang G. et al. Comparative transcriptome analysis of climacteric fruit of chinese pear (*Pyrus ussuriensis*) reveals new insights into fruit ripening. Plos One vol.9 (2014)
8. Lin Q., Wang C., Dong W., et al. Transcriptome and metabolome analyses of sugar and organic acid metabolism in Ponkan (*Citrus reticulata*) fruit during fruit maturation. Gene 554(1):64-74 (2014)
9. Wu H.X., Jia H.M., Ma X.W., et al. Transcriptome and proteomic analysis of mango (*Mangifera indica* Linn) fruits. J Proteomics 105:19-30 (2014)
10. Grassi S., Piro G., Lee J.M., Zheng Y., Fei Z., et al. Comparative genomics reveals candidate carotenoid pathway regulators of ripening watermelon fruit. BMC Genomics 14:781 (2013)
11. Zenoni S., Ferrarini A., Giacomelli E., et al. Characterization of transcriptional complexity during berry development of *Vitis vinifera* using RNA-seq. Plant Physiol 152(4):1787-1795 (2010)
12. Sweetman C., Wong D.C.J., Ford C.M., Drew D.P. Transcriptome analysis at four developmental stages of grape berry (*Vitis vinifera* cv. Shiraz) provides insights into regulated and coordinated gene expression. BMC Genomics 13:691 (2012)
13. Da Silva C., Zamperin G., Ferrarini A., et al. 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(12):4777-4788 (2013)
14. Degu A., Hochberg U., Sikron N., et al. Metabolite and transcript profiling of berry skin during fruit development elucidates differential regulation between Cabernet Sauvignon and Shiraz cultivars at branching points in the polyphenol pathway. BMC Plant Biology 14:188 (2014)
15. Jaillon O., Aury J.M., Noel B., Policriti A., Clepet C., et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-467 (2007)
16. Roberts A., Pimentel H., Trapnell C., Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics 27:2325-2329 (2011)
17. Forcato C. Gene prediction and functional annotation in the *Vitis vinifera* genome. PhD Thesis, Universita' Degli Studi Di Padova (2010)
18. Warden C.D., Yuan Y.C., Wu W. Optimal Calculation of RNA-Seq Fold-Change Values. International Journal of Computational Bioinformatics and In Silico Modeling 2(6):285-292 (2013)

19. Hebenstreit D., Fang M., Gu M., Charoensawan V., van Oudenaarden A., Teichmann S.A. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. Mol Syst Biol. 7:497 (2011)
20. Richmond T.A., Somerville C.R. The cellulose synthase superfamily. Plant Physiol. 124(2):495-498 (2000)
21. Lerouxel O., Cavalier D.M., Liepman A.H., Keegstra K. Biosynthesis of plant cell wall polysaccharides - a complex process. Curr Opin Plant Biol. 9(6):621-630 (2006)
22. Andrews D.L., Beames B., Summers M.D., Park W.D. Characterization of the lipid acyl hydrolase activity of the major potato (*Solanum tuberosum*) tuber protein, patatin, by cloning and abundant expression in a baculo-virus vector. Biocheml J. 252:199-206 (1988)
23. Browning K.S. The plant translational apparatus. Plant Mol. Biol. 32(1-2):107-144 (1996)
24. Wada I., et al. SSR alpha and associated calnexin are major calcium binding proteins of the endoplasmic reticulum membrane. J Biol Chem. 266:19599-19610 (1991)
25. Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., et al. *AGO1* defines a novel locus of Arabidopsis controlling leaf development. EMBO J. 17:170-180 (1998)
26. Baumberger N. & Baulcombe D.C. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci USA. 102:11928-11933 (2005)
27. Qi Y., Denli A.M., Hannon G.J. Biochemical specialization within Arabidopsis RNA silencing pathways. Mol Cell. 19:421-428 (2005)
28. Brodersen P., Sakvarelidze-Achard L., Bruun-Rasmussen M., Dunoyer P., Yamamoto Y.Y., et al. Widespread translational inhibition by plant miRNAs and siRNAs. Science 320:1185-1190 (2008)
29. Jover-Gil S., Candela H., Robles P., Aguilera V., Barrero J.M., Micol J.L., Ponce M.R. The microRNA pathway genes *AGO1*, *HEN1* and *HYL1* participate in leaf proximal-distal, venation and stomatal patterning in Arabidopsis. Plant Cell Physiol. 53:1322-1333 (2012)
30. Sorin C., Bussell J.D., Camus I., et al. Auxin and light control of adventitious rooting in Arabidopsis require ARGONAUTE1. Plant Cell 17:1343-1359 (2005)
31. Morel J., Godon C., Mourrain P., Beclin C., Boutet S., Feuerbach F., Proux F., Vaucheret H. Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. Plant Cell 14:629-639 (2002)
32. Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 28:511-515 (2010)
33. Trapnell C., Hendrickson D.G., Sauvageau M., Goff L., Rinn J.L., et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 31:46-53 (2012)
34. Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14:R36 (2013)
35. Benjamini Y., Hochberg Y. Controlling the False Discovery Rate: A Practicle and Powerfull Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 57(1):289-300 (1995)

Chapter 5

Putative stage- and phase-specific molecular biomarkers of berry development

INTRODUCTION

Transcriptomic analysis methods, as RNA-sequencing, generate large data sets. In this case, the handling of the transcriptional data set is quite complicated by the fact that the samples represented berry development of ten different grapevine varieties. In line with this, it is important to consider that the transcriptional modulation observed entailed trends that were likely involved both in common programs – berry development and maturation – and varietal-specific traits – i.e. white versus red varieties.

To dissect this complex biological system and to extract information describing relevant biological processes, diverse strategies were applied. In particular, multivariate statistics analysis (MVA) can be described as a group of tools for reducing the dimensionality of large dataset to render the visualization and interpretation more manageable[1,2]. MVA methods can be generally grouped into supervised and unsupervised approaches. Unsupervised approaches, as Principal Component Analysis (PCA), use any information on group identity to construct the model. The data are analysed as belonging to a single block (X-block) of observations and variables, allowing the identification of subgroupings in the data. Supervised approach means that group identity is defined, focusing the analysis on extracting the variables important for group separation. The data are divided into two separate blocks: the X-block containing the predictor variables (e.g. FPKM data) and the Y-block containing the response variables (e.g. stages, phases, skin colour, variety…)[3]. Bidirectional orthogonal projections to latent structures (O2PLS)[4,5] method is a MVA technic that allows the partitioning of the systematic variability in X and Y into three parts: the X/Y joint predictive variation; the Y-orthogonal variation in X; and the X-unrelated variation in Y[2]. O2PLS can be also used for discriminant analysis (DA) to differentiate between individual classes[6]. O2PLS-DA is a powerful method to identify discriminating variables as physiological and molecular features contributing to sample separation.

In this work, different patterns among sample distribution were identified by using unsupervised PCA, confirming the tendency previously observed in **chapter 4**. Then, O2PLS-DA was used firstly to analyse the FPKM data set and obtain the best model – namely the model explaining the best sample organization – and to extract putative molecular biomarkers. These marker transcripts provided new insights into the grape berry development features.

RESULTS AND DISCUSSION

Data set exploration

In order to reveal the general pattern contributing to sample separation, a Principal Component Analysis (PCA) was performed on the 40-sample FPKM data set (**Fig. 1**). The first two principal components (PC) explained 44.3% of the total transcriptional variance – cumulative variance. PC1 could explain the time course of berry development from pre-véraison transcriptomes – placed to the left – and post-véraison transcriptomes to the right. PC2 distinguished clearly pea-sized berry stage from prevéraison stage transcriptomes, whereas end of véraison and harvest stages slightly overlapped. Interestingly, PC2 seemed to describe a separation between white- and red-skinned berry samples starting from the end of véraison and more evidently at harvest. In particular, Barbera, Negro amaro and Refosco full-ripe berry transcriptomes were distinguished from the other harvest transcriptomes.

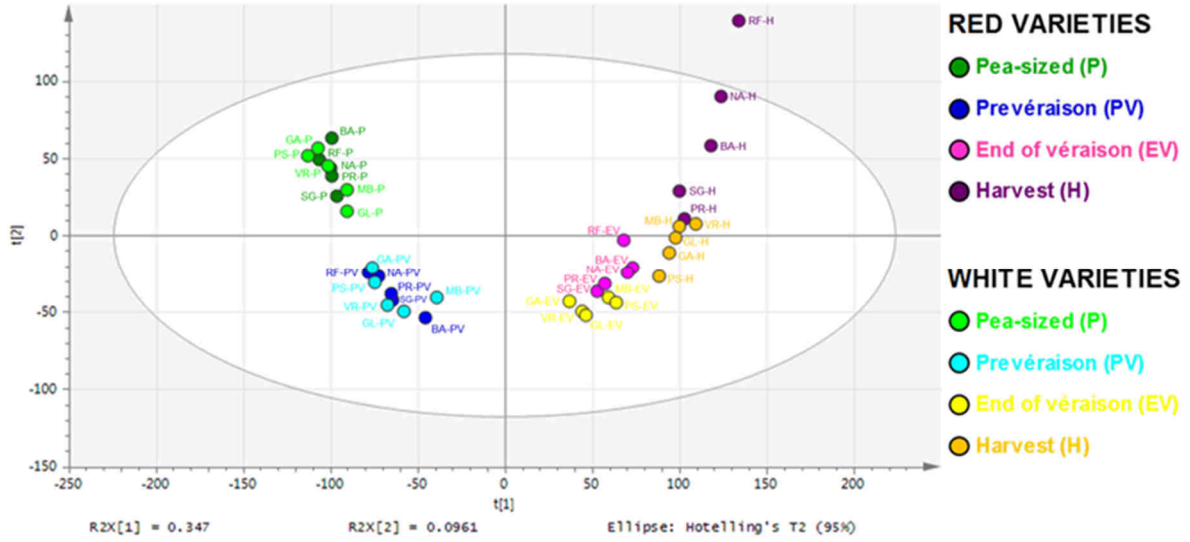


Figure 1: Plot of 40-sample data set PCA showing the first two principal components t[1]=34.7% and t[2]=9.61%.

Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest) and for each variety: SG (Sangiovese), BA (Barbera), NA (Negro amaro), RF (Refosco), PR (Primitivo) for the red grape varieties and VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina) for the white grape varieties.

This PCA was used to supervise the best Orthogonal bidirectional Projections to Latent Structures (O2PLS) model, namely the model with the highest and the most similar overall fit and cross-validated correlation values (R^2Y and Q^2 values

respectively)[3]. As result, a four-class O2PLS-DA model (3+6+0, UV, R^2Y=0.988, Q^2=0.901, p(CV-ANOVA) = 5.885 E-21), corresponding to the four growth stages, was obtained (**Fig. 2**). The three principal components (PC) explained 48.3% of the total variance of the model. PC1 could represent again the time course of berry development. PC2 and PC3 allowed distinguishing clearly the four growth stages. The 3-dimension plot underlined again the clear separation of transcriptomes between pre- and post-véraison phases, but also permitted to split the transcriptomes into four growth stages independently on berry skin colour.

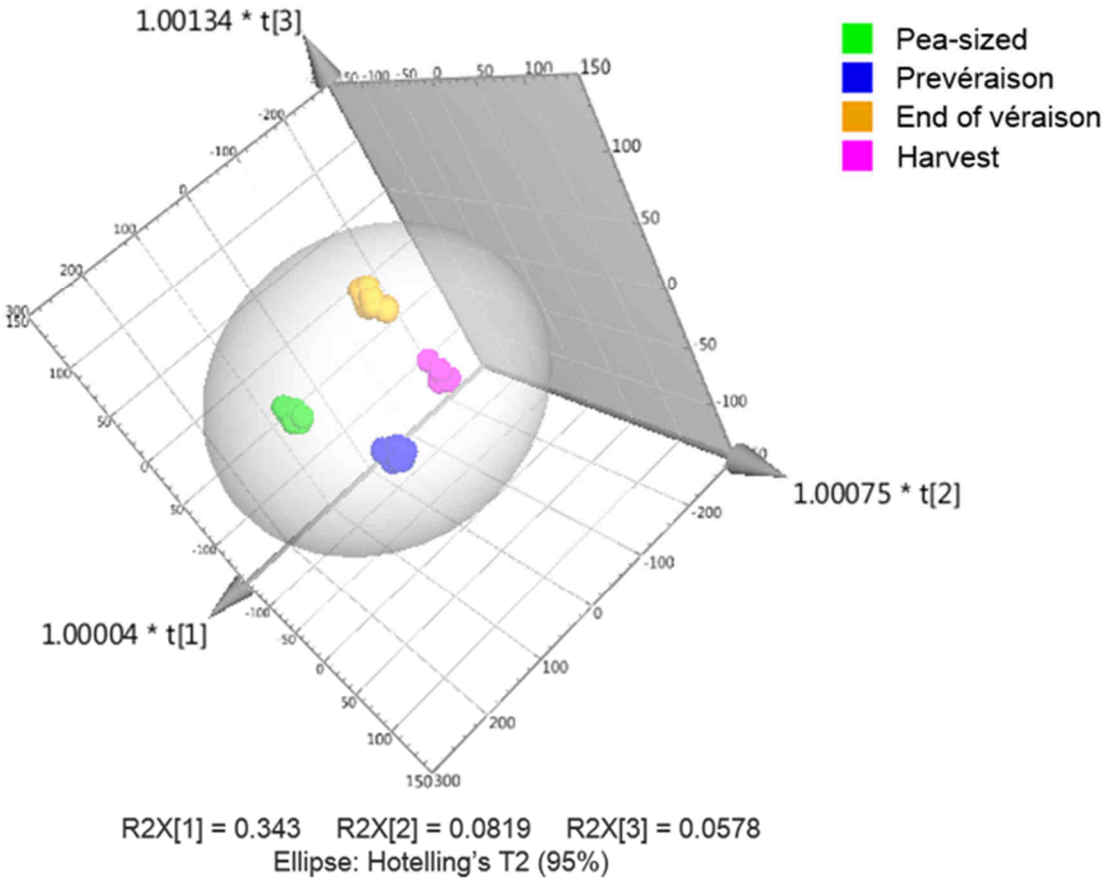


Figure 2: Variables and scores scatterplot of the O2PLS-DA model (3 + 6 + 0, UV, R2Y = 0.988, Q2 = 0.901, p(CV-ANOVA) = 5.885 E-21) applied to 40-sample data set. The plot shows the first three principal components t[1] = 34.3%; t[2] = 8.19%; t[3] = 5.78% (UV = unit variance scaling method).

Identification of putative stage-specific molecular biomarkers

The previous observation was exploited to identify putative stage-specific biomarkers – namely transcripts whose presence or absence defines all 10-variety samples at a given stage – applying four distinct two-class O2PLS-DA models. In each case, the transcriptomes of all varieties at a specific stage (10 samples) were used as a reference group/class against all the other sample transcriptomes at the remaining three stages (a group of 30 samples) in one unique class[7]. A S-plot[8] was used to select biomarkers with a loading correlation coefficient – called p(corr) – greater than 0.9 (positive biomarkers) and lower than -0.9 (negative biomarkers). Using these parameters, only putative positive biomarkers were obtained: 326, 17, 3 and 17 transcripts for pea-sized berry stage, prevéraison, end of véraison and harvest, respectively (**Fig. 3**).

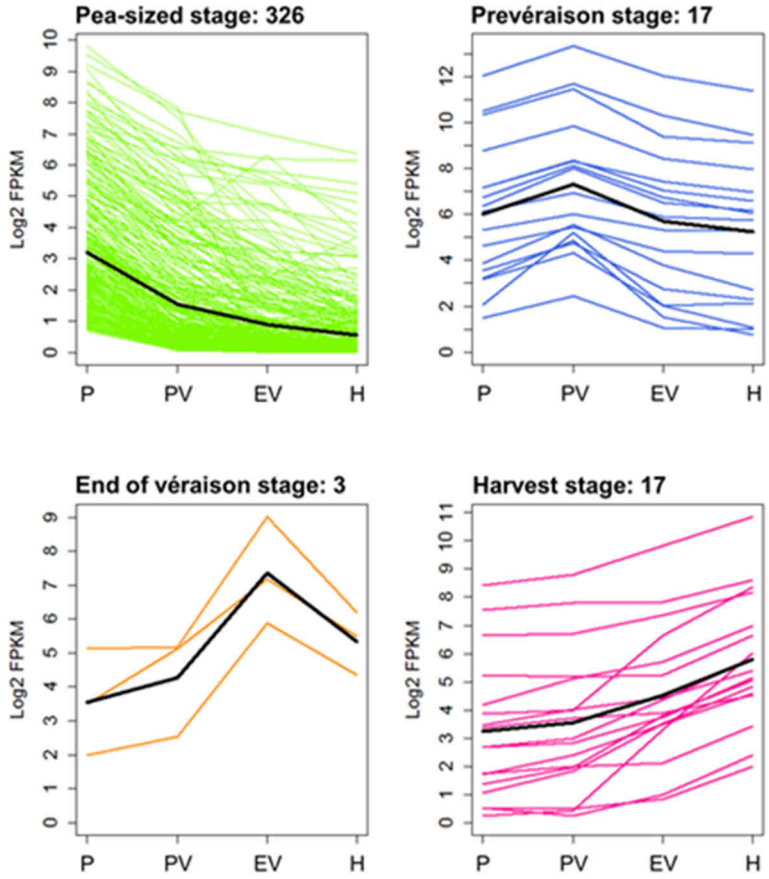


Figure 3: Expression profiles of the putative positive molecular biomarkers of each phenological stage. Putative biomarkers were selected using an S-plot[8] with a p(corr) > 0.9. Each single line represents the log_2-transformed average of the mean FPKM values for an individual transcript (R software).

In **Figure 3**, the plots show that the major expression level corresponds to the stage, which the genes are putative biomarkers of. However, their expression was not completely exclusive of the relative stage and then did not exhibit a unique peak of expression, but rather a smooth profile. For example, putative pea-sized stage-specific biomarkers did not exhibit a profile completely flat until harvest – the average expression remained slightly high in prevéraison and end of véraison berries. The same trend was also observed for the other groups.

Putative pea-sized berry stage-specific biomarker transcripts

The 326 genes selected as putative pea-sized berry stage-specific biomarker genes were analysed for overrepresented functions using the gene ontology (GO) enrichment tool BiNGO[9] in Cytoscape with the GoSlim Plants annotation, only referred to biological processes. The BiNGO-derived graph illustrates the most significantly overrepresented functions at pea-sized berry stage (**Fig. 4**).

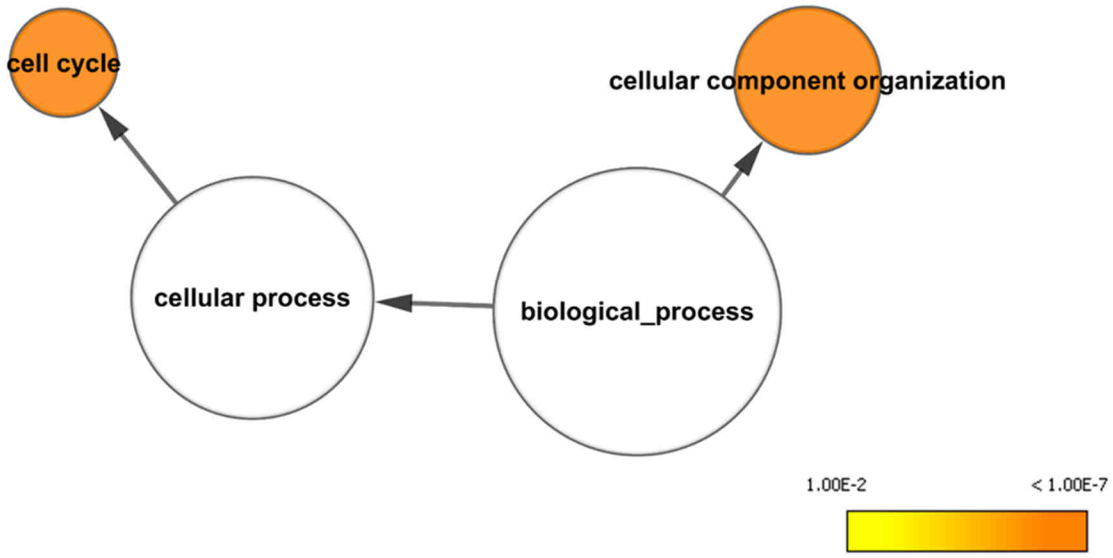


Figure 4: Enriched GO terms in putative pea-sized berry stage-specific biomarker transcripts.
The network graph shows BiNGO visualization of the overrepresented GO terms for the 326 putative pea-sized berry stage-specific biomarker transcripts. Node size is positively correlated with the number of genes belonging to the category. Non-coloured nodes are not overrepresented, but they may be the parents of overrepresented terms. Coloured nodes represent GO terms that are significantly overrepresented (Benjamini and Hochberg corrected p-value < 0.01), with the shade indicating significance as shown in the colour bar.

Putative pea-sized berry stage-specific biomarker genes were significantly enriched for genes involved in cell cycle (GO:0007049) and cellular component organization (GO:0016043). These two functions are representative of the berry size increase through rapid cell division occurring during this stage[10]. Indeed, a period of rapid berry growth – both through rapid cell division and cell enlargement – comes immediately after bloom until around 3-4 weeks later, establishing the number of cells in a grape berry. In fact, no further cell division occurs after this period[10].

Among the genes belonging to the cell cycle GO category, the three types of cyclin gene necessary to the plant cell-cycle machinery[11] were identified: two cyclin A genes (VIT_18s0001g02060; VIT_04s0008g04260), two cyclin B genes (VIT_13s0067g01420; VIT_08s0040g00930) and the corresponding cyclin-dependent kinase (VIT_18s0122g00550) and five cyclin D genes (VIT_00s0199g00210; VIT_18s0001g07220; VIT_07s0129g01100; VIT_18s0001g09920; VIT_00s0203g00160). Concerning cellular component organization function, three histone-core genes were found: H2A (VIT_00s0179g00340), H3 (VIT_06s0004g01710) and H4 (VIT_06s0004g04370) genes all necessary for chromatin and chromosome structure during DNA duplication[12]. In addition, genes involved in cell wall metabolism were identified such as six pectin methyl-esterase genes (VIT_07s0005g00730; VIT_03s0017g01950; VIT_07s0005g01940 VIT_11s0016g03020; VIT_14s0060g01960; VIT_11s0016g00330), a cellulose synthase-like gene (VIT_08s0007g03990), a cellulose synthase 3 gene (VIT_08s0007g08380), the expansin gene *VvEXPA16* (VIT_14s0108g01020) and an extensin gene (VIT_11s0016g02900).

Putative prevéraison stage-specific biomarker transcripts

Only 17 genes were identified as prevéraison stage-specific putative biomarker genes (**Table 1**). The encoded proteins were involved in various biological processes like signal transduction (VIT_02s0087g00720; VIT_08s0040g02360), regulation of transcription/transcription factor activity (VIT_19s0014g01250; VIT_14s0068g01690), transport (VIT_14s0060g00690; VIT_13s0073g00560; VIT_18s0117g00260; VIT_15s0107g00270), cytokinin biosynthesis with the isopentenyltransferase 5 gene[13] (VIT_07s0104g00270) and in translation process as an elongation factor 1-alpha gene[14] (VIT_08s0040g02330) for example.

Table 1: Putative prevéraison stage-specific biomarker transcripts. For each transcript, a mean FPKM value of the 10-variety samples was calculated for each growth stage. Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

O2PLS-DA Model, UV, 1+3+0, R2Y = 0.991, Q2 = 0.94

Gene_id	p(corr)	Functional annotation	P	PV	EV	H
VIT_05s0020g02490	0.9548	Unknown protein	38.78	62.76	38.53	37.84
VIT_05s0020g01170	0.9510	Translationally-controlled tumor protein	4,196.24	10,281.28	4,159.81	2,683.68
VIT_07s0191g00180	0.9478	Homeobox-leucine zipper protein ATHB-6	142.34	323.47	131.04	94.77
VIT_19s0014g01250	0.9317	TUBBY like protein 7 TLP7	69.54	120.96	58.64	52.10
VIT_02s0087g00720	0.9274	Receptor protein kinase	3.21	35.40	3.05	1.11
VIT_06s0004g01390	0.9256	S-receptor kinase	13.42	45.89	12.89	5.59
VIT_14s0060g00690	0.9209	ADP-ribosylation factor A1B	1,284.45	2,770.65	664.73	551.80
VIT_08s0040g02360	0.9196	Calmodulin binding protein	81.07	254.58	89.55	69.97
VIT_13s0073g00560	0.9165	DnaJ homolog, subfamily A, member 2	1,428.96	3,295.28	1,273.95	711.68
VIT_04s0023g03880	0.9133	Unknown	106.33	278.99	107.03	64.77
VIT_06s0004g06650	0.9096	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase	1.77	4.38	1.11	1.00
VIT_08s0040g02330	0.9068	Elongation factor 1-alpha 1	434.52	928.30	343.55	251.53
VIT_18s0117g00260	0.9058	Co-chaperone-curved DNA binding protein A	10.85	26.07	5.70	3.91
VIT_14s0068g01690	0.9035	TCP family transcription factor TCP20	23.71	42.40	20.10	18.58
VIT_15s0107g00270	0.9020	Mechanosensitive ion channel	8.24	28.04	1.88	0.70
VIT_18s0001g05800	0.9016	Dehydration-responsive protein	141.82	312.83	172.43	123.44
VIT_07s0104g00270	0.9015	Isopentenyltransferase 5	8.03	18.87	3.01	3.42

Putative end of véraison stage-specific biomarker transcripts

For the end of véraison stage, only three genes were classified as putative stage-specific biomarker genes (**Table 2**).

Table 2: Putative end of véraison stage-specific biomarker transcripts. For each transcript, a mean FPKM value of the 10-variety samples was calculated for each growth stage. Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

O2PLS-DA Model, UV, 1+4+0, R2Y = 0.988, Q2 = 0.843

Gene_id	p(corr)	Functional annotation	P	PV	EV	H
VIT_05s0051g00590	0.9436	Pectate lyase	34.01	34.95	516.52	72.27
VIT_12s0142g00160	0.9072	Unknown protein	10.23	34.12	142.80	44.02
VIT_12s0057g00700	0.9002	Glucan endo-1,3-beta-glucosidase 3 precursor	2.96	4.80	57.70	19.23

These genes coded for a pectate lyase (VIT_05s0051g00590), an uncharacterized protein (VIT_12s0142g00160) and a glucan endo-1,3-beta-glucosidase 3 precursor (VIT_12s0057g00700). Pectate lyases catalyse the eliminative cleavage of de-esterified pectin – prior de-methylation by pectin methyl-esterases –, which is a major component of the primary cell walls of many higher plants[15], contributing to pectin depolymerization and solubilization that takes part to fleshy fruit softening. Plant endo-1,3-β-glucanases are able to catalyse the endo-type hydrolytic cleavage of β-1,3-D-glucosidic linkages in β-1,3-glucans[16]. Although endo-1,3-β-glucanase activity has been generally related to antimicrobial activity, these enzymes seem also to participate to the polysaccharide modifications taking place in the cell wall of fruit during ripening. An example is a banana endo-1,3-β-glucanase isolated from pulp tissues found to be likely involved in

ripening and tissue softening[17]. The peak of expression of these two transcripts at the end of véraison confirmed their involvement in cell wall loosening and berry softening during fruit ripening[18].

Putative harvest stage-specific biomarker transcripts

Regarding harvest stage, 17 genes were selected as putative stage-specific biomarker transcripts (**Table 3**).

Table 3: Putative harvest stage-specific biomarker transcripts. For each transcript, a mean FPKM value of the 10-variety samples was calculated for each growth stage. Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

O2PLS-DA Model, UV, 1+4+0, R2Y = 0.991, Q2 = 0.892

Gene_id	p(corr)	Functional annotation	P	PV	EV	H
VIT_12s0057g00310	0.9257	Gamete expressed1 (GEX1)	0.17	0.35	8.75	63.48
VIT_04s0008g05770	0.9169	CBL-interacting protein kinase 25 (CIPK25)	13.49	14.75	99.78	328.50
VIT_02s0109g00250	0.9163	4-coumarate-CoA ligase	0.44	0.39	0.80	2.99
VIT_11s0118g00560	0.9155	Nodulin	5.33	6.92	19.59	56.32
VIT_06s0004g07930	0.9128	Nucleotidyltransferase family	10.00	15.29	20.48	40.55
VIT_05s0102g00250	0.9107	Unknown protein	2.34	2.94	3.32	9.80
VIT_01s0011g05810	0.9097	Adagio protein 1	2.27	4.25	10.18	22.71
VIT_16s0098g01230	0.9095	Cytochrome b5	184.59	220.37	226.43	385.88
VIT_00s1916g00010	0.9087	DNA binding	36.98	34.94	36.91	99.17
VIT_05s0020g03870	0.9068	MLK/Raf-related protein kinase 1	1.58	2.89	12.85	32.01
VIT_18s0001g11910	0.9067	1-acyl-sn-glycerol-3-phosphate acyltransferase 4	0.42	0.17	0.97	4.27
VIT_00s0225g00010	0.9062	Unknown protein	17.02	33.20	51.55	124.39
VIT_13s0074g00470	0.9062	OTU cysteine protease	5.33	6.15	12.28	34.34
VIT_05s0020g04080	0.9052	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 2	343.29	441.20	910.12	1848.21
VIT_18s0001g10480	0.9018	Unknown protein	1.08	2.59	10.34	27.35
VIT_08s0217g00030	0.9016	Ran-binding protein 1 RanBP1	98.96	103.00	162.11	286.39
VIT_00s0226g00040	0.9011	Myb family	9.20	11.92	13.71	21.89

For example, a 4-coumarate-CoA ligase gene (VIT_02s0109g00250) involved in phenylpropanoid biosynthesis pathway[19] and the 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 2 (VIT_05s0020g04080) also named acidoreductone oxygenase 2 (*ARD2*) were identified. ARD2 protein is involved in the 5-methylthioadenosine or Yang cycle, a set of reactions that recycle 5-methylthioadenosine to methionine. In plants, 5-methylthioadenosine is a byproduct of polyamine, ethylene, and nicotianamine biosynthesis[20-22]. However, recently Gonda and colleagues[23] showed that L-methionine could be incorporated into aroma volatiles through two distinct catabolic routes: the first route through L-methionine aminotransferase as previously noted for other amino acids[24], and the second route through L-methionine-ɣ-lyase[25]. They also showed that L-methionine may be metabolized into L-isoleucine in melon fruits, and may ultimately contribute to the formation of L-isoleucine-derived aroma volatiles. Even if the contribution of L-

methionine to the formation of aroma compounds in grape berry remains unexplored, it could be coherent with the expression peak of *ARD2* in full-ripe grape berries.

Surprisingly, the gamete expressed 1 (*GEX1*) (VIT_12s0057g00310) gene was classified as biormarker of full-ripe berries, maybe due to an erroneous gene prediction. Indeed, *GEX1* is known to have dual functions in gametophyte development and early embryogenesis in *Arabidopsis thaliana*[26] but not for its involvement in fruit ripening.

CBL-interacting protein kinase 25 (CIPK25) (VIT_04s0008g05770) encoding gene was also identified as putative harvest stage biomarker. Calcineurin B-like (CBL) proteins are calcium-binding proteins that forms a complex network with their target kinases being the CBL-interacting protein kinases (CIPKs)[27]. The expression peak of CIPK25 in full-ripe berries shows that calcium signals might be involved in the achievement of full maturity.

Identification of putative phase-specific molecular biomarkers

The above-described approach for the selection of putative stage-specific biomarkers was also used to determine genes involved into the green-to-maturation phase shift by grouping the transcriptomes of the two stages corresponding to pre- and post-véraison phases. In this case, 1,009 putative positive biomarker transcripts of the green phase and 44 of the maturation phase were selected (**Fig. 5 a**). Considering the huge number of putative green phase-specific biomarker genes, the first part of grape berry development, which consists in berry formation, seems to be a common process among the studied varieties. On the other hand, the small amount of putative maturation phase-specific biomarker transcripts suggests that more genotype-dependent traits may appear during this phase. Furthermore, by comparing the biomarker gene lists in Venn diagrams, we observed that the phase-specific biomarkers did not overlap the stage-specific biomarkers (**Fig. 5 b**). That was not surprising as genes expressed in a specific stage were clearly down-regulated in the other stage of the same developmental phase (**Fig. 3**).

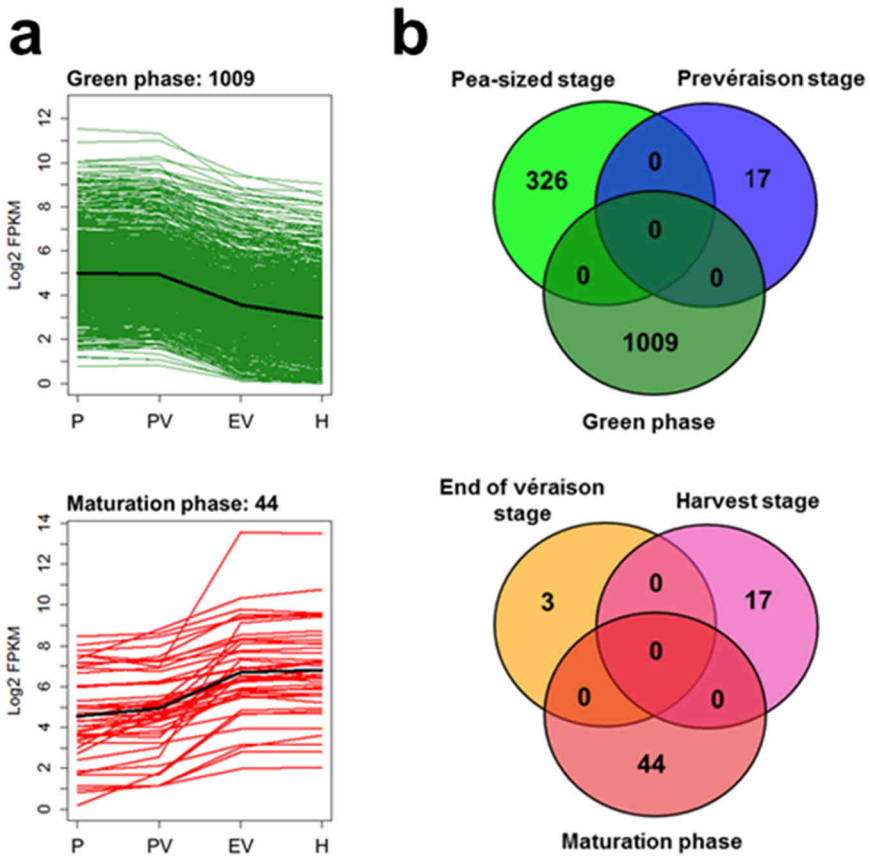


Figure 5: (**a**) Expression profiles of the putative positive molecular biomarkers of each phase. Biomarkers were selected using an S-plot[8] with a p(corr) > 0.9. Each single line represents the log_2-transformed average of the mean FPKM values for an individual transcript.
(**b**) Venn diagrams showing the specificity of the putative positive biomarkers of each pre-véraison stage and the green phase (above), and of each post-véraison stage and the maturation phase (below).

Putative green phase-specific biomarker transcripts

In order to gain insights into the biological processes taking place in the related developmental phase, the 1,009 putative green phase-specific biomarker genes were analysed for overrepresented functions using the GO enrichment tool BiNGO[9] and the GoSlim Plants annotation (**Fig. 6**). Five GO categories were found significantly enriched: the photosynthesis process (GO:0015979), generation of precursor metabolites and energy (GO:0006091), cellular process (GO:0009987), response to abiotic stimulus (GO:0009628) and translation process (GO:0006412). All these functions are essential for the first phase of rapid berry growth.

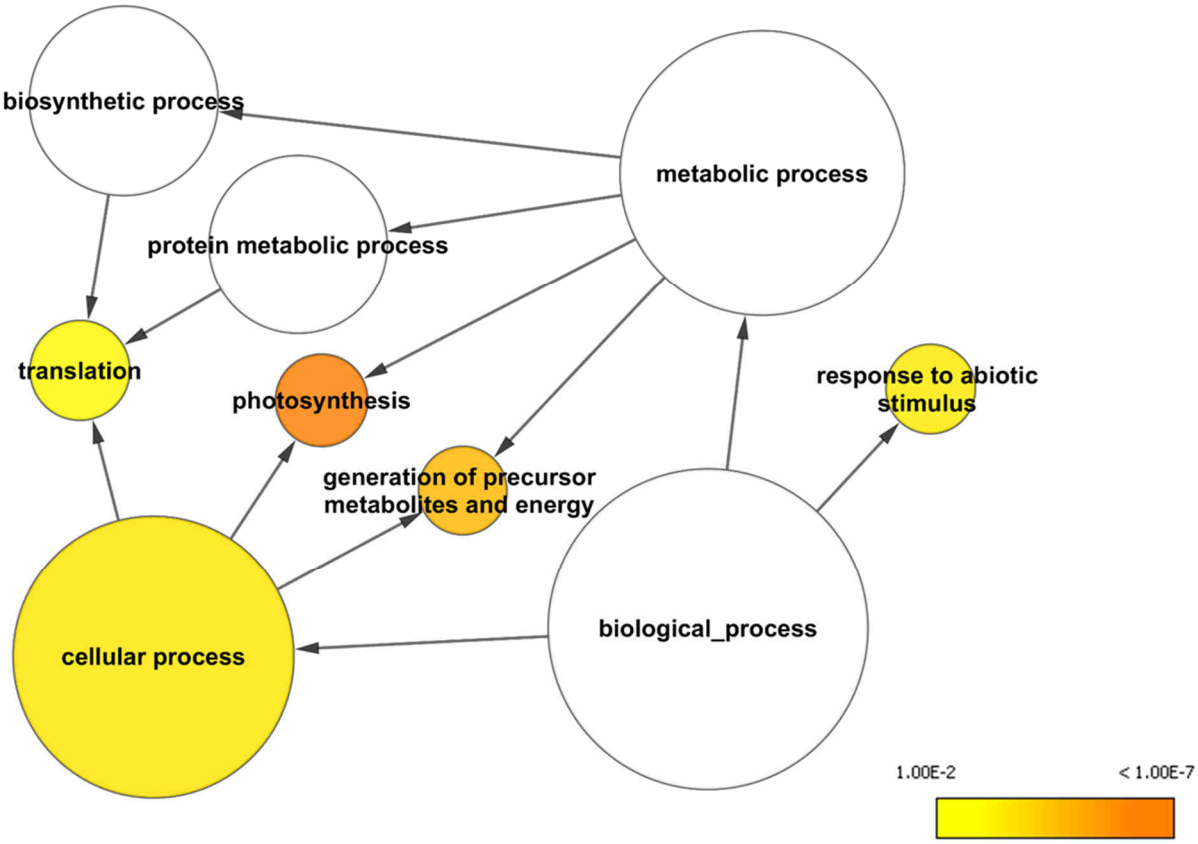


Figure 6: Enriched GO terms in putative green phase-specific biomarker transcripts.
The network graph shows BiNGO visualization of the overrepresented GO terms for the 1,009 putative green phase-specific biomarker transcripts. Node size is positively correlated with the number of genes belonging to the category. Non-coloured nodes are not overrepresented, but they may be the parents of overrepresented terms. Coloured nodes represent GO terms that are significantly overrepresented (Benjamini and Hochberg corrected p-value < 0.01), with the shade indicating significance as shown in the colour bar.

Among the genes related to photosynthesis process, we identified genes coding for five photosynthetic antenna, 13 other photosystem I and II related proteins and the small unit of the ribulose-1,5-bisphophate carboxylase (VIT_17s0000g03690), accordingly with the fact that during this time the fruit was still photosynthetically active.

Genes involved in starch synthesis were also identified, including two glycogen synthase genes (VIT_14s0108g00940; VIT_11s0065g00150), two isoforms of the phosphoenolpyruvate carboxylase (VIT_12s0028g02180; VIT_01s0011g02740) that might be responsible for organic acids – especially malate – accumulation[28], and a pyruvate kinase (VIT_05s0020g04480) and chloroplast precursor of a glyceraldehyde-3-phosphate dehydrogenase (VIT_14s0068g00680) genes both engaged in glycolysis process.

Regarding the cellular process category enrichment, we found genes involved in the synthesis and the regulation of the cytoskeleton, genes related to the cell cycle and thirty-seven transcription factor genes that could be responsible for holding on an immature transcriptional program before the maturation process will be activate. The expression of this set of genes allowed berries to grow both through cell division and cell enlargement. Finally, amino acid-tRNA synthase and ribosomal protein encoding genes represented the major part of the enriched translation process category.

Putative maturation phase-specific biomarker transcripts

Concerning the maturation phase, 44 genes were classified as putative phase-specific biomarker genes (**Table 4**). Among these genes, we found genes coding for a dehydration-responsive protein (VIT_16s0100g00570) consistent with the stress condition berries possibly undergo during ripening, a 6-phosphogluconate dehydrogenase (VIT_02s0025g00900) involved in the pentose phosphate pathway that generates energy and potentially provides precursors for the shikimate pathway and other secondary metabolic pathways occurring during this phase that generate aroma and flavour components[29]. We found also the sucrose synthase 2 (VIT_07s0005g00750) involved in sugar accumulation[30] and a pyruvate kinase (VIT_08s0007g04170), enzyme belonging to the last step of glycolysis. During this phase, pyruvate may be diverted from the TCA cycle for metabolism through aerobic ethanol fermentation instead, as the TCA cycle flux becomes limiting[31], and fermentative metabolism may be required to de-acidify the cytosol of fruit cells[32]. Interestingly, the ethylene response factor gene *VvERF045* (VIT_04s0008g06000) was also selected as putative maturation phase-biomarker gene. Previously, *VvERF045* gene expression was found specific of the ripe Pinot noir berries[33]. In the study, *VvERF045* was induced before véraison and was specific of the ripe berry, by sequence similarity it was likely a transcription activator. To functionally characterize *VvERF045*, stable transformations - over-expression and silencing – were done. Four pathogen stress response genes were significantly induced in the transgenic plants over-expressing *VvERF045*, suggesting a putative function of VvERF045 in biotic stress defence during berry ripening. Thus, the identification of *VvERF045* as putative maturation phase-biomarker gene is not surprising for field-cultivated plants. Finally, we identified also eight genes involved in transcription factor activity (VIT_13s0019g05170; VIT_05s0077g00750; VIT_07s0005g01710; VIT_08s0040g01950;

VIT_18s0076g00330; VIT_16s0050g00950; VIT_05s0020g04730; VIT_14s0060g02640) that may play a key role in the green-to-maturation phase transition.

Table 4: Putative maturation phase-specific biomarker transcripts. For each transcript, a mean FPKM value of the 10-variety samples was calculated for each growth stage. Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

O2PLS-DA Model, UV, 1+1+0, R2Y = 0.991, Q2 = 0.977

Gene_id	p(corr)	Functional annotation	P	PV	EV	H
VIT_16s0100g00570	0.9497	Dehydration-responsive protein	0.13	2.44	27.23	30.13
VIT_14s0066g01720	0.9474	Bromo-adjacenty (BAH) domain-containing protein	24.90	26.37	59.83	67.80
VIT_17s0000g08140	0.9458	MLK/Raf-related protein kinase 1	31.29	39.70	101.94	112.65
VIT_13s0019g05170	0.9456	Zinc finger (C2H2 type) family	0.76	1.26	5.92	5.88
VIT_02s0025g00900	0.9422	6-phosphogluconate dehydrogenase	28.71	35.30	73.71	85.21
VIT_18s0117g00130	0.9421	Pi starvation-induced protein	214.10	249.65	674.44	796.33
VIT_09s0002g07110	0.9416	KEG (keep on going)	13.45	20.64	80.35	112.29
VIT_10s0071g01030	0.9382	Transducin family protein / WD-40 repeat	23.75	23.30	42.55	46.24
VIT_17s0000g01950	0.9366	Dihydropyrimidine dehydrogenase	60.79	75.90	168.98	155.72
VIT_05s0020g02430	0.9359	Protein FBL4	258.52	350.32	633.10	729.20
VIT_07s0005g00750	0.9351	Sucrose synthase (SUS2)	122.87	176.12	743.50	729.28
VIT_08s0007g04170	0.9336	Pyruvate kinase, cytosolic isozyme	139.13	160.27	324.74	366.80
VIT_18s0001g07240	0.9309	14-3-3 protein GF14 kappa (GRF8)	349.26	391.28	868.19	778.08
VIT_04s0044g01060	0.9302	Leaf senescence related protein-like	1.19	1.20	2.92	3.18
VIT_10s0116g00370	0.9277	Unknown protein	2.18	2.22	24.54	24.73
VIT_13s0067g00870	0.9238	Peptidyl-prolyl cis-trans isomerase ROC7 (rotamase CyP 7)	15.89	15.26	46.42	36.62
VIT_07s0005g03310	0.9234	Cofilin	165.30	451.56	1,273.41	1,708.70
VIT_05s0077g00750	0.9233	Basic helix-loop-helix (bHLH) family	15.49	21.26	90.39	75.01
VIT_06s0004g07560	0.9219	Ribosomal protein S29 28S	11.51	12.11	52.61	61.22
VIT_04s0044g01940	0.9202	Diacylglycerol O-acyltransferase 1	18.58	25.40	49.39	48.55
VIT_07s0005g01710	0.9157	WRKY DNA-binding protein 23	9.43	19.73	125.75	87.59
VIT_00s0203g00200	0.9154	Universal stress protein (USP) family protein	63.91	70.14	114.84	144.12
VIT_13s0019g00840	0.9143	UDP-glucuronate decarboxylase.	188.78	152.65	369.67	418.42
VIT_18s0001g01120	0.9139	Proteasome 26S regulatory subunit S3 (RPN3)	19.17	19.80	42.80	49.99
VIT_08s0040g01950	0.9138	Zinc finger (C3HC4-type ring finger)	5.50	25.22	271.26	283.29
VIT_18s0076g00330	0.9133	VirE2-interacting protein (VIP1)	12.21	24.27	49.35	62.67
VIT_16s0022g00960	0.9108	Invertase/pectin methylesterase inhibitor	189.82	124.27	12,092.05	11,357.25
VIT_18s0122g00110	0.9084	Unknown protein	10.52	13.61	24.00	27.01
VIT_12s0028g01050	0.9077	Helicase in vascular tissue and TAPETUM	8.29	8.48	14.16	14.76
VIT_04s0008g06000	0.9071	VvERF045_ERF/AP2 Gene Family	6.99	36.78	317.60	265.55
VIT_13s0084g00050	0.9071	Cysteine proteinase inhibitor	100.15	120.30	310.95	308.91
VIT_08s0056g01240	0.9068	No hit	2.38	4.98	552.26	682.97
VIT_18s0076g00310	0.9066	Translation initiation factor eIF-5B	38.17	44.80	77.77	97.24
VIT_16s0050g00950	0.9064	Scarecrow-like transcription factor 8 (SCL8)	21.49	38.76	79.07	89.82
VIT_05s0062g00890	0.9063	Alpha-glucan water dikinase isoform 3, Chloroplast	28.57	31.08	56.36	57.45
VIT_05s0020g04730	0.9056	Zinc finger (C3HC4-type ring finger)	0.93	1.25	7.12	11.40
VIT_19s0015g01300	0.9044	Amino acid permease 7	11.49	10.17	158.93	127.97
VIT_16s0022g02230	0.9042	Leucine-rich repeat receptor protein kinase EXS	9.18	19.04	42.47	49.75
VIT_04s0008g01380	0.9033	Unknown protein	32.54	33.33	98.94	147.60
VIT_06s0004g03910	0.9031	Unknown protein	4.27	7.26	91.54	82.82
VIT_04s0008g06880	0.9029	ER membrane DUF1077	143.22	149.99	214.34	237.99
VIT_14s0060g02640	0.9014	Myb family	2.63	3.41	7.96	8.16
VIT_08s0007g04900	0.9009	Unknown protein	64.02	74.73	118.51	148.96
VIT_05s0020g04160	0.9000	Stem-specific protein TSJT1	122.70	104.36	210.06	197.42

CONCLUSION

The deep transcriptome shift between pre- and post-véraison phase berries was confirmed by the PCA. In addition, the unsupervised analysis showed a separation between white- and red-skinned berry samples starting from the end of véraison and more evidently at harvest particularly for Barbera, Negro amaro and Refosco full-ripe berry transcriptomes. This transcriptome separation suggested that skin colour seemed to affect transcriptome behaviours from véraison and more especially for Barbera, Negroamaro and Refosco berries. This indication is explored deeper into **chapter 8**.

However, the best O2PLS-DA model was obtained applying a four-class hypothesis corresponding to the four phenological stages. The model revealed a clear clusterization of the four-growth stage transcriptomes independently of the skin colour, suggesting that samples grouped more on the basis of the developmental stage rather than the skin colour or the variety identity. Therefore, the biomarker identification strategy was based on this information allowing the discovery of putative stage- or phase-specific positive biomarker transcripts. The huge number of putative green phase- and pea-sized berry stage-specific biomarker transcripts suggested that the first part of grape berry development seemed to be a common process among the ten studied varieties, whereas only few putative biomarker genes were selected for the two ripening stages and the maturation phase, suggesting that more genotype-dependent traits might appear during this phase.

Finally, a detailed study of each putative stage- and phase-specific biomarker gene lists showed that these transcripts stuck correctly with the known physiological and biological processes occurring during berry development but also provided new insights and then source of further investigation.

METHODS

PCA and O2PLS

Principal component analysis (PCA) was performed on the 21,746 transcript-data set using SIMCA P+ 13 software (Umetrics, USA). Orthogonal partial least squares discriminant analyses (O2PLS-DA) were used to integrate the PCA data and to find the best hypothesis of relationship between the forty transcriptomes. Our O2PLS-DA model is composed of 4 classes and was selected for its high and similar R2Y value,

which indicates how well the model explains the dataset, and cross-validated correlation Q2, representing a measure of the predictive power of the model[3]. A measure of significance for the observed group separation was calculated as a cross-validated ANOVA (CV-ANOVA) p-value. In order to define the number of latent components for O2PLS-DA models, we applied partial cross-validation and a permutation test to reveal overfitting (**Supplemental figure 1**). Multivariate data analysis was performed by using SIMCA P+ 13 (Umetrics, USA).

Biomarker transcript identification

Putative biomarker transcripts were identified by applying four distinct two-class O2PLS-DA models. In each case, the RPKM values of the ten varieties at one developmental stage (10 samples) were used as a reference class and the remaining values from the three other stages (30 samples) were grouped in one unique class[7]. The same approach was used for the green and the maturation phases by grouping the two stages corresponding to each phase. Then, an S-plot[8] was used to select putative biomarkers with a p(corr) higher than 0.9 and lower than -0.9.

Expression profile of putative biomarker transcripts

For each putative positive biomarker transcript, an average of the 10 variety-FPKM values was calculated at each stage and $\log_2$-transformed ($\log_2$(FPKM average + 1)). The resulting expression profiles were plotted in six different graphs describing the peculiar trend of the putative positive biomarker transcripts using the following R script (Example for harvest stage).

```
>data <- read.delim("Harv.txt")

>make.data.plot <- function(data) {
    >data.2 <- data[,1:dim(data)[2]-1]
    >colors <- data[,dim(data)[2]]
    >print(colors[1])

>plot(1:dim(data.2)[2],data.2[1,],type="l",col=as.character(colors[1]),ylim=c(0,max(dat
a.2)),ylab="Log2 FPKM",cex.lab=1.3,cex.axis=1.3,lwd=1,lab=c(4,8,7))
```

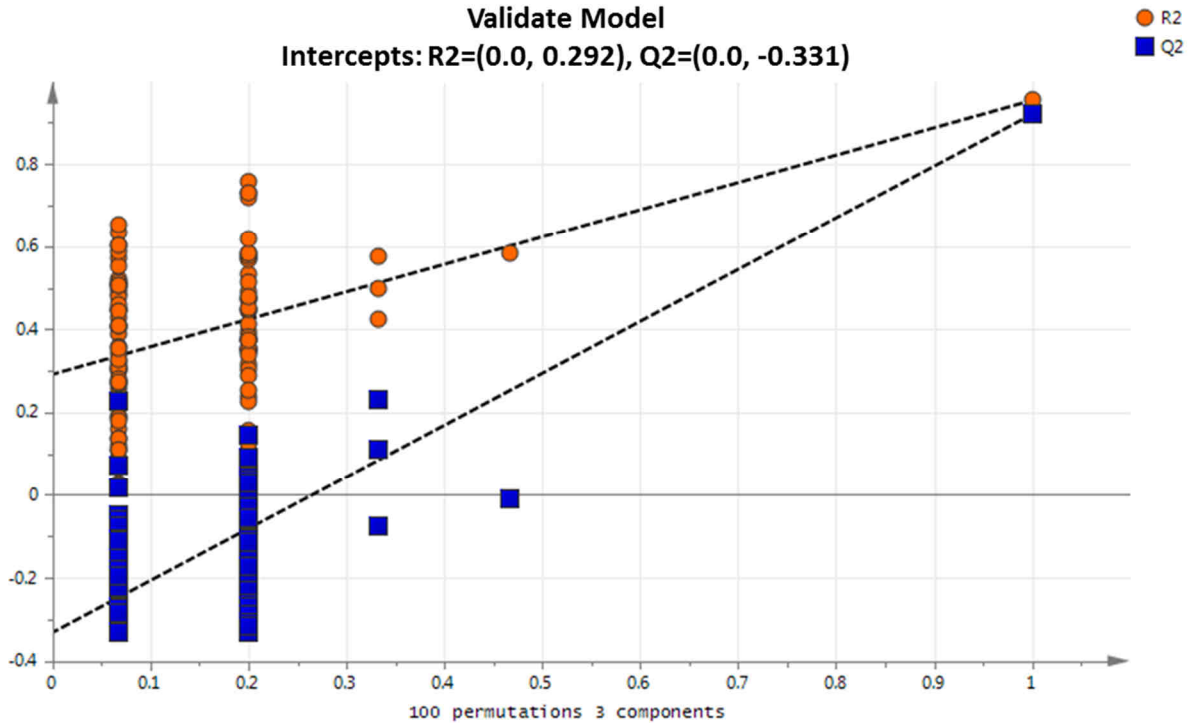
```
        >for ( i in 1:17) {
      print(colors[i])
      lines(1:dim(data.2)[2],data.2[i,],col=as.character(colors[i]),lwd=1)
   }

        for ( i in 18:18) {
      print(colors[i])
      lines(1:dim(data.2)[2],data.2[i,],col=as.character(colors[i]),lwd=4)
   }
}

>make.data.plot(data)
```

GO Enrichment analysis

GO enrichment analysis on putative pea-sized stage- and green phase-specific biomarker genes was performed by using BiNGO 2.4 plug-in tool in Cytoscape version 3.1.0 with PlantGOslim categories, as described by Maere et al.[9]. Overrepresented PlantGOslim categories were identified using a hypergeometric test with a significance threshold of 0.01 after a Benjamini & Hochberg False-Discovery Rate (FDR) correction[34].

SUPPLEMENTARY MATERIAL

Supplementary Figure 1: Validation of O2PLS-DA model. The three-latent-component O2PLS-DA model in **Figure 2** was partially cross-validated and a permutation test (100 permutations) was used to highlight putative overfitting.



REFERENCES

1. Madsen R., Lundstedt T., Trygg J. Chemometrics in metabolomics – a review in human disease diagnosis, Anal. Chim. Acta. 659(1-2):23-33 (2010)
2. Eriksson L., Byrne T., Johansson E., Trygg J., Vikström C. Multi- and Megavariate Data Analysis Basic Principles and Applications (Third revised edition). MKS Umetrics AB (2013)
3. Wheelock A. & Wheelock C. Trials and tribulations of 'omics data analysis: assessing quality of SIMCA-based multivariate models using examples from pulmonary medicine. Mol. BioSyst. 9(11):2589-2596 (2013)
4. Trygg J. O2-PLS for qualitative and quantitative analysis in multivariate calibration. Journal of Chemometrics 16:283-293 (2002)
5. Trygg J. & Wold S. O2-PLS, a two-block (X–Y) latent variable regression (LVR) method with an integral OSC filter. Journal of Chemometrics 17:53-64 (2003)
6. Bylesjö M., Rantalainen M., Cloarec O., Nicholson J.K., Holmes E., Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. Journal of Chemometrics 20(8-10):341-351 (2006)
7. Zamboni A., Di Carli M., Guzzo F., et al. Identification of putative stage-specific berry biomarkers and omics data integration into networks. Plant Physiol. 154(3):1439-1459 (2010)
8. Wiklund S., Johansson E., Sjöström L., Mellerowicz E.J., Edlund U., Shockcor J.P., Gottfries J., Moritz T., Trygg, J. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Anal. Chem. 80:115-122 (2008)
9. Maere S., Heymans K., Kuiper M. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21:3448-3449 (2005)
10. Dokoozlian N.K. Grape berry growth and development. In LP. Christensen (Ed.), Raisin production manual (pp. 30-37). Oackland, CA. Agricultural and Natural Resources Publication (2000)
11. De Veylder L., Beeckman T., Inzé D. The ins and outs of the plant cell cycle. Nature Reviews Molecular Cell Biology 8:655-665 (2007)
12. Jin J, Cai Y., Li B., et al. In and out: histone variant exchange in chromatin. Trends in Biochemical Sciences 30:680-687 (2005)
13. Miyawaki K., Tarkowski P., Matsumoto-Kitano M., Kato T., Sato S., Tarkowska D., Tabata S., Sandberg G., Kakimoto T. Roles of Arabidopsis ATP/ADP isopentenyltransferases and tRNA isopentenyltransferases in cytokinin biosynthesis. Proc. Natl. Acad. Sci. U.S.A. 103:16598-16603 (2006)
14. Ejiri S. Moonlighting functions of polypeptide elongation factor 1: from actin bundling to zinc finger protein R1-associated nuclear localization. Biosci Biotechnol Biochem. 66:1-21 (2002)
15. Carpita N.C. & Gibeaut D.M. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. The Plant Journal 3:1-30 (1993)
16. Leubner-Metzger G. & Meins F. Jr. Functions and regulation of plant β-1,3-glucanases (PR-2). In Pathogenesis-Related Proteins in Plants (Datta, S.K. & Muthukrishnan, S., eds), pp. 49-76. CRC Press LLC, Boca Raton, USA (1999)

17. Peumans W.J., Barre A., Derycke V., Rougé P., Zhang W., May G.D., Delcour J.A., Van Leuven F., Van Damme E.J. Purification, characterization and structural analysis of an abundant beta-1,3-glucanase from banana fruit. Eur J Biochem. 267(4):1188-95 (2000)
18. Schlosser J., Olsson N., Weis M., Reid K., Peng F., Lund S., Bowen P. Cellular expansion and gene expression in the developing grape (Vitis vinifera L.). Protoplasma 232:255-265 (2008)
19. Heller W. & Forkmann G. Biosynthesis. In: Harborne JB (ed.), The Flavonoids: Advances in Research since 1980. Chapman and Hall, London (1988)
20. Miyazaki J.H. & Yang S.F. Metabolism of 5-methylthioribose to methionine. Plant Physiol. 84:277-281 (1987)
21. Kende H. Ethylene biosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. 44:283-307 (1993)
22. Roje S. S-Adenosyl-L-methionine: Beyond the universal methyl group donor. Phytochemistry 67:1686-1698 (2006)
23. Gonda I., Lev S., Bar E., et al. Catabolism of L-methionine in the formation of sulfur and other volatiles in melon (*Cucumis melo* L.) fruit. The Plant Journal 74:458-472 (2013)
24. Gonda I., Bar E., Portnoy V., et al. Branched-chain and aromatic amino acid catabolism into aroma volatiles in *Cucumis melo* L. fruit. J. Exp. Bot. 61:1111-1123 (2010)
25. Goyer A., Collakova E., Shachar-Hill Y., Hanson A.D. Functional characterization of a methionine-lyase in Arabidopsis and its implication in an alternative to the reverse trans-sulfuration pathway. Plant Cell Physiol. 48(2):232-242 (2007)
26. Alandete-Saez M., Ron M., Leiboff S., McCormick S. Arabidopsis thaliana GEX1 has dual functions in gametophyte development and early embryogenesis. Plant J. 68(4):620-32 (2011)
27. Sheen J. Ca2+-dependent protein kinases and stress signal transduction in plants. Science 274:1900-1902 (1996)
28. Diakou P., Svanella L., Raymond P., Gaudillère J.P., Moing A. Phosphoenolpyruvate carboxylase during grape berry development: protein level, enzyme activity and regulation. Aust. J. Plant. Physiol. 27(3):221-229 (2000)
29. Kruger N.J. & von Schaewen A. The oxidative pentose phosphate pathway: structure and organisation. Plant Biology 6(3):236-246 (2003)
30. Hawker J.S. Changes in the activities of enzymes concerned with sugar metabolism during the development of grape berries. Phytochemistry 8(1):9-17 (1969)
31. Ollat N. & Gaudillere J.P. Investigation of assimilate import mechanisms in berries of *Vitis vinifera* var. Cabernet Sauvignon. Acta Hortic. 427:141-149 (1996)
32. Longhurst T., Lee E., Hinde R., Brady C., Speirs J. Structure of the tomato Adh2 gene and Adh2 pseudogenes, and a study of Adh2 gene expression in fruit. Plant Mol Biol. 26(4):1073-1084 (1994)
33. Dal Ri A. Functional and molecular characterization of ethylene responsive factor genes involved in grape ripening. PhD Thesis, Alma Mater Studiorum Università di Bologna (2011)
34. Klipper-Aurbach Y., Wasserman M., Braunspiegel-Weintrob N., Borstein D., Peleg S., Assa S., Karp M., Benjamini Y., Hochberg Y., and Laron Z. Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. Med. Hypotheses 45:486-490 (1995)

Chapter 6

Common and skin colour-specific differentially expressed genes

INTRODUCTION

The most common use of transcriptome profiling is in the search for differentially expressed genes (DEGs), namely genes that show differences in expression level between conditions or in other ways are associated with given predictors or responses. In the past decades, DNA microarrays have been used extensively to quantify the abundance of mRNA corresponding to different genes, and more recently, high-throughput sequencing of cDNA (RNA-Seq) has emerged as a powerful competitor. RNA-Seq offers several advantages over microarrays for differential expression analysis, such as an increased dynamic range and a lower background level, and the ability to detect and quantify the expression of previously unknown transcripts and isoforms[1-3]. Indeed RNA-seq allows detecting transcripts generated by alternative splicing and understanding whether there is any tissue- or stage-specific isoform expression.

In this work, DEGs between each subsequent pair of developmental stages (**chapter 4**) were compared among the ten varieties allowing commonly up- and down-regulated genes to be related to the physiological and biological changes already known at each growth stage[4,5]. Furthermore, the DEG comparison permitted to find out DEGs specific of each skin colour group and to gain new insights into skin colour-specific transcriptomic traits.

RESULTS AND DISCUSSION

Differentially expressed gene (DEG) comparison

In **chapter 4**, Cuffdiff program, a part of the Cufflinks package[6,7], was applied. This program used the aligned reads from two subsequent developmental stages (i.e. Pea-sized berry-Prevéraison, Prevéraison-End of véraison, End of véraison-Harvest) of each variety and reported transcripts resulting differentially expressed by rigorous statistical analysis. In this work, a comparison of these DEGs was done in order to obtain the genes that were commonly up- and down-regulated both in all the ten variety-berries, and in the five red- and white-skinned berries separately (**Fig. 1**).

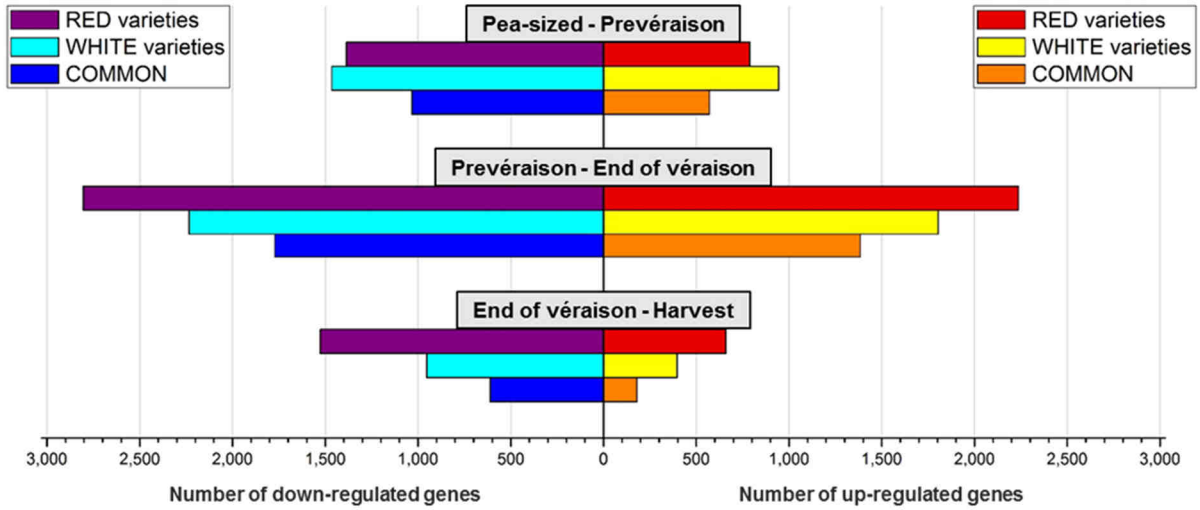


Figure 1: Differentially expressed genes between each subsequent pair of developmental stages. The histogram represents the number of the commonly down- (on the left side) and up-regulated genes (on the right side) in all the red-coloured varieties (RED varieties), all the white-coloured varieties (WHITE varieties) and in all the ten varieties (COMMON) between each subsequent pair of developmental stages.

In **Figure 1**, a high transcriptional modulation is observed after véraison with more genes down- rather than up-regulated during berry development which could explain the decline of transcripts after véraison observed in **chapter 4**. Furthermore, red-skinned berries seemed to have a higher transcriptional modulation after véraison whereas white-skinned berries seemed to be slightly more transcriptionally active during the green development.

Commonly up-regulated genes at each growth stage change

Pea-sized berries - Prevéraison

571 genes were identified as up-regulated between pea-sized berry stage and prevéraison. In order to find out the biological processes, which these 571 genes belong to, an analysis for overrepresented functions was done using the gene ontology (GO) enrichment tool BiNGO[8] in Cytoscape and the GoSlim Plants annotation. Nonetheless, no GO category was identified as significantly overrepresented, showing that no specific process or ontological category was induced during the green phase. However, the modulation of some interesting genes was noticed, such as the up-regulation of Myb divaricata (*DIV*). DIV corresponds to a two-MYB domain protein, which is involved in growth and in the definition of cell types. Galego and colleagues[9] found that it is expressed mainly in a single layer of cells of ventral and lateral petals

in Antirrhinum flowers. In tomato, MYBI, orthologous to DIV, was described to interact with Fruit SANT/MYB Binding protein1 (FSB1) allowing differentially fruit cell expansion during tomato development[10]. Consequently, we can suppose that the gene Myb divaricata could be involved in cell enlargement during green berry development. We observed also the presence of sixty-two transcription factor activity-related genes and fifty transcripts involved in transport as an auxilin-like protein. In a study, Suetsugu and co-workers[11] showed that an auxilin-like J-domain protein, JAC1, regulates phototropin-mediated chloroplast movement in *Arabidopsis thaliana*. Indeed, chloroplasts change their position in a cell in response to environmental light conditions[12]. Hence, the up-regulation of this auxilin-like gene at prevéraison could be related to a hypothetical increase of chloroplast movement within berry pericarp cells at this stage.

Prevéraison - End of véraison

In total, 1,384 transcripts were classified as commonly up-regulated in all the 10-variety berries between prevéraison and end of véraison stages. In order to gain insights into biological processes, these 1,384 genes were analysed for overrepresented functions using the GO enrichment tool BiNGO[8] and the GoSlim Plants annotation, only referred to biological processes. In **Figure 2**, the BiNGO-derived graph illustrates the most highly significant enrichment of specific functions. Basically, node size is positively correlated with the number of genes belonging to the category. Non-coloured nodes are not overrepresented, but they may be the parents of overrepresented terms. Coloured nodes represent GO terms that are significantly overrepresented (corrected p-value < 0.01[13]), with the shade indicating significance as shown in the colour bar. Prevéraison up-regulated genes were significantly enriched for genes involved in generation of precursor metabolites and energy (GO:0006091), transport (GO:0006810), carbohydrate metabolic process (GO:0005975) and catabolic process (GO:0009056).

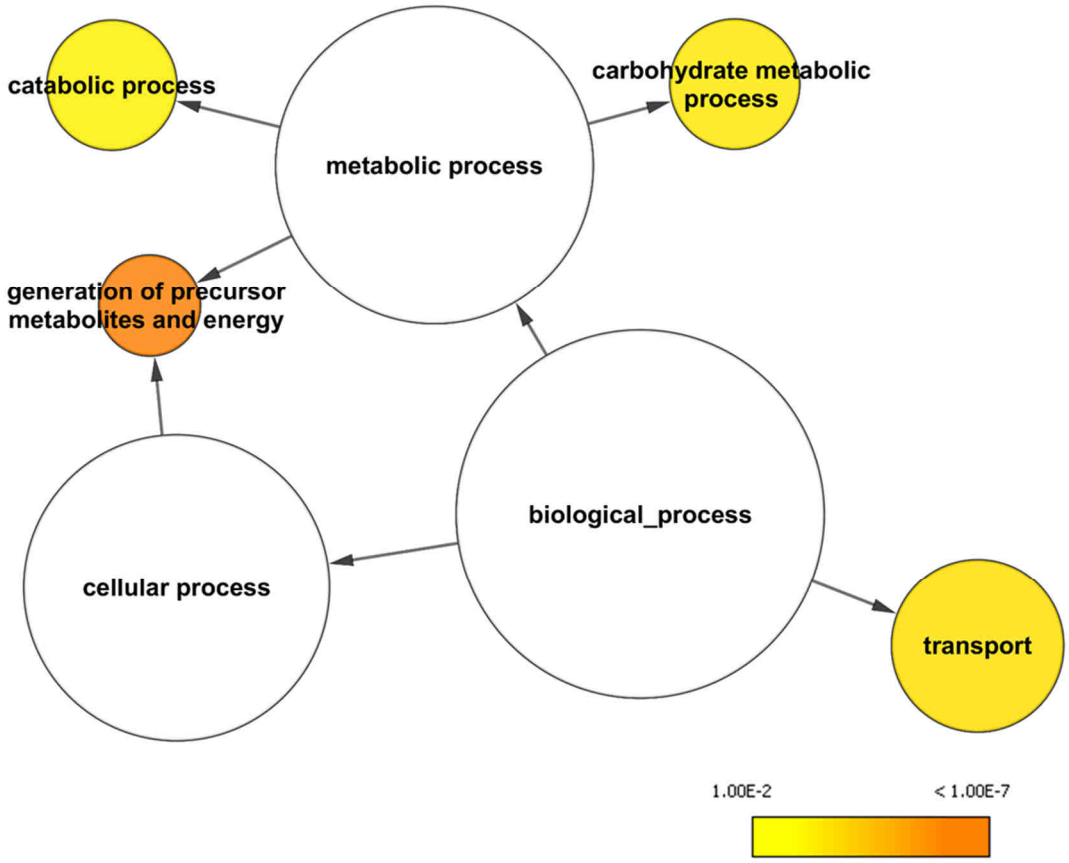


Figure 2: Enriched GO terms in up-regulated genes between prevéraison and end of véraison stages. The network graph shows BiNGO visualization of the over-represented GO terms for the 1,384 up-regulated genes between prevéraison and end of véraison stages.

Among the genes belonging to the generation of precursor metabolites and energy GO category, we identified several genes related to glycolysis as a glyceraldehyde-3-phosphate dehydrogenase gene (VIT_01s0010g02460), a cytosolic phosphoglycerate kinase gene (VIT_19s0085g00370) and three pyruvate kinase (VIT_08s0007g04170; VIT_02s0012g01170; VIT_10s0071g01060) genes, and genes involved in citric acid cycle: two isocitrate dehydrogenase genes (VIT_03s0038g03120; VIT_04s0079g00530), a succinate dehydrogenase gene (VIT_14s0219g00150) and a malate dehydrogenase gene (VIT_10s0003g01000). The up-regulation of the glycolysis-related genes confirmed the hypothesis of Famiani *et al.*[14] that glycolysis is actually not inhibited during ripening and that sugars, and not malate, may provide the bulk of the deficit in substrate utilised by respiration and ethanolic fermentation in the pericarp of grape berries during most of ripening. Accordingly, the up-regulation of genes coding for enzymes involved in Krebs cycle, especially the malate dehydrogenase, supported the above-stated hypothesis that after the onset of ripening, malate is mainly used as

an intermediate in the conversion of sugars in pyruvate to support respiration and alcoholic fermentation.

Concerning transport category, sugar transporter genes were found, as two tonoplast monosaccharide transporter genes *VvTMT1*, also named *VvHT6*, and *VvTMT2* (VIT_18s0122g00850 and VIT_03s0038g03940 respectively), the sucrose transporter gene *VvSUC11* (VIT_18s0001g08210), also called *VvSUT1*, and the hexose transporter gene *VvHT13* (VIT_11s0016g03400) suggesting their involvement in sugar influx and accumulation, which becomes remarkable throughout the ripening process[15,16]. Furthermore, two genes coding for aquaporins – one situated on the plasma membrane (PIP) (VIT_08s0040g01890) and another one on the tonoplast (TIP) (VIT_13s0019g00330) – and two potassium transporters (VIT_00s0125g00190; VIT_17s0000g01930) were also identified as up-regulated, and might be related to water influx contributing to berry enlargement and softening[17,18].

As Deluc and colleagues[19] described, the up-regulation of transcripts involved in starch degradation was also observed at the end of véraison, including an α-amylase (VIT_03s0063g00400) and two water dikinase genes (VIT_05s0062g00890; VIT_05s0062g00900). In the carbohydrate metabolic process category, genes related to the polysaccharide composition of the cell wall were found as four xyloglucan endotransglycosylase genes (VIT_02s0012g02220; VIT_05s0062g00480; VIT_06s0061g00550; VIT_01s0011g06250) and three endo-1,4-beta-glucanase genes (VIT_00s0340g00050; VIT_00s2526g00010; VIT_07s0005g00740) both known to participate in cell wall loosening, by depolymerization and solubilization of hemicellulose and pectin[20].

Finally, among the catabolic process category, the up-regulation of five GSDL-motif lipase genes (VIT_09s0002g00580; VIT_14s0066g00250; VIT_03s0038g03580; VIT_18s0001g14790; VIT_18s0001g14800) was noticed. GDSL esterases/lipases are a subclass of lipolytic enzymes that are very important research subjects because of their multifunctional properties, such as broad substrate specificity. It was reported that GDSL family from *Arabidopsis thaliana* consists of 108 members[21], and *Vitis vinifera* contain 96 members[22]. The ability of GDSL-lipases to act on a wide range of substrates[23] makes it difficult to identify their natural substrates and consequently complicates the characterization of their biochemical activity *in planta*. In fact, the involvement of GDSL-lipase genes in plant defence was suggested, based on their induction by salicylic acid and pathogens[24-26] and the demonstration that their activity modulates disease susceptibility[25,26]. Similarly, GDSL-lipases were shown to be

induced by various abiotic elicitors[26-29], and their ability to increase plant tolerance to salt and other environmental stresses was shown[26,28]. In addition, GDSL-lipases were indicated as participating in hormonal pathways related to growth processes[30,31]. There is also evidence for a correlation between the expression of GDSL-lipases and other developmental processes[29,32-35]. Finally, the involvement of GDSL-lipases in the metabolism of cutin and wax, which form the cuticle secreted by the epidermis in the aerial parts of the plant, is also becoming apparent[36-42]. Consequently, we can suppose that the five GDSL lipase genes might be involved either in response to abiotic or biotic stress response, or in response to hormonal stimulus, or in cutin and wax metabolism during ripening phase.

End of véraison - Harvest

Only 179 genes were identified as commonly up-regulated between the end of véraison and harvest, confirming that transcriptional modulation seems to be genotype-dependent during maturation phase (**chapters 4 and 5**). These genes were analysed for overrepresented functions using the GO enrichment tool BiNGO[8] and the GoSlim Plants annotation (**Fig. 3**). Four GO categories were found significantly enriched: biosynthetic process (GO:0009058) including genes belonging to the secondary metabolic process (GO:0019748) and transcription (GO:0006350) categories, and response to biotic stimulus (GO:0009607).

Among the enriched secondary metabolic process category, we found genes coding for a 4-coumarate-CoA ligase, a phenylalanine ammonium lyase (PAL) and eight stilbene synthases, related to the phenylpropanoid metabolic pathway and stilbene biosynthesis[43] (**Supplementary table 1**). However, gene expression intensity of the PAL and the stilbene synthase genes was largely higher in full-ripe red-skinned berries. Indeed, it was shown in several studies that the red grapes contain generally higher levels of stilbenes than the white berries[44,45]. Moreover, the greater expression of members of the PAL gene family was consistent with the production of substrates for anthocyanin biosynthesis that is the fingerprint of red grape varieties (**chapter 9**).

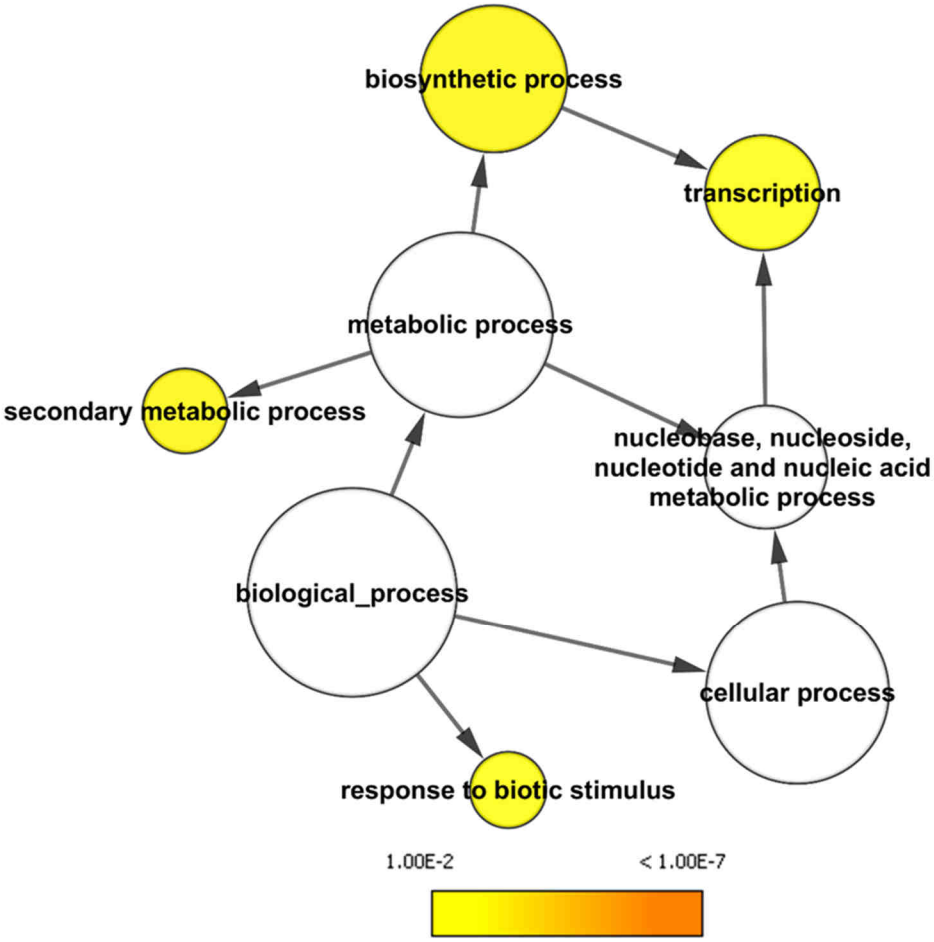


Figure 3: Enriched GO terms in up-regulated genes between end of véraison and harvest stages. The network graph shows BiNGO visualization of the over-represented GO terms for the 179 up-regulated genes between end of véraison and harvest stages.

Seventeen transcription factor genes were identified among the enriched transcription category including *VvNAC03* (VIT_00s0375g00040), *VvNAC11* (VIT_14s0108g01070) and *VvNAC37* (VIT_10s0003g00350). NAC (NAM/ATAF1/2/CUC2) domain-containing proteins comprise one of the largest plant-specific TF families, with more than 100 members identified in both Arabidopsis[46] and rice[46,47], which are expressed in several tissues and developmental stages. NACs play important roles in plant development, including the formation and maintenance of shoot apical meristems[48], floral development and morphogenesis[49], embryo development[50], hormone signalling[51], and formation of secondary walls[52]. Recently, a tomato NAC transcription factor, *SlNAC4*, was found to function as a positive regulator of fruit ripening and carotenoid accumulation[53]. In grapevine, Wang and colleagues[54] performed a genome-wide identification of *NAC* domain genes and investigation of

their chromosome locations, gene structures, phylogeny and expression profiles. They found out for each grape NAC gene the corresponding most similar gene in *Arabidopsis thaliana*, and predicted the function of each NAC gene according to its orthologue. Regarding this analysis, *VvNAC03* and *VvNAC11* might be involved in multicellular organism development, while *VvNAC37* might be also related to the regulation of secondary cell wall biogenesis. Consequently, the up-regulation of *VvNAC37* in full-ripe berries could suggest the formation of secondary cell wall maybe related to the cell aging.

The response to biotic stimulus category was also classified as significantly enriched, including three pathogenesis related-protein genes (VIT_05s0077g01560; VIT_05s0077g01540; VIT_05s0077g01550) and three harpin-inducing protein genes (VIT_04s0069g01010; VIT_08s0007g02360; VIT_16s0098g00890). The term Pathogenesis Related Proteins (PR) covers all antimicrobial plant proteins that are induced after pathogens attack[55]. Harpins are glycine-rich and heat-stable proteins that are secreted through type III secret ion system in gram-negative plant-pathogenic bacteria. Furthermore, some harpins were shown to have virulence activity[56,57]. Hence, the up-regulation of these six transcripts could be explained by the fact that plants were grown in field, where they were likely responding to biotic and abiotic stresses.

Commonly down-regulated genes during berry development

Pea-sized berry stage - Prevéraison

Differentially expression analysis permitted to identify 1,034 down-regulated genes in all varieties between pea-sized berry stage and prevéraison. After analysis of these genes using the GO enrichment tool BiNGO[8] and the GoSlim Plants annotation, four significantly overrepresented categories were observed: cellular component organization (GO:0016043), secondary metabolic process (GO:0019748), cell cycle (GO:0007049) and carbohydrate metabolic process (GO:0005975) (**Fig. 4**).

The down-regulation of genes belonging to the cell cycle category, as cyclin and cyclin-dependent genes, showed that cell division was slowing down at prevéraison, establishing the final number of cells before switching to fruit growth by only cell expansion.

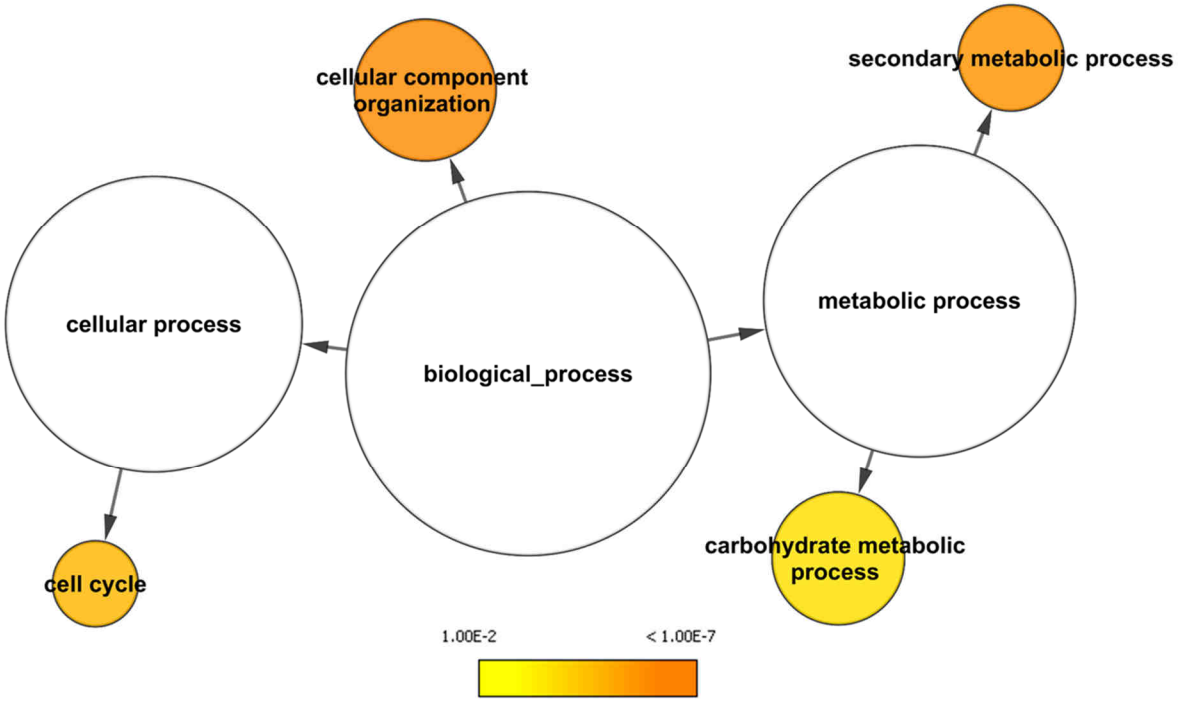


Figure 4: Enriched GO terms in down-regulated genes between pea-sized berry stage and prevéraison stage. The network graph shows BiNGO visualization of the over-represented GO terms for the 1,034 down-regulated genes between pea-sized berry stage and prevéraison stage.

Among the cellular component organization category, thirty-four genes related to cell wall metabolism were identified as three cellulose synthase-like genes (VIT_04s0069g00780; VIT_00s0469g00010; VIT_08s0007g03990), the cellulose synthase A4, A7 and A8 genes (VIT_07s0005g04110; VIT_11s0037g00530; VIT_10s0003g01560 respectively) involved in secondary cell wall metabolism in *Arabidopsis thaliana*[58,59], three endo-1,4-beta-glucanase genes (VIT_18s0001g14040; VIT_07s0005g00740; VIT_19s0090g01050) and four expansin genes: *VvEXPA6* (VIT_06s0004g04860), *VvEXPA13* (VIT_13s0019g01650), *VvEXPA16* (VIT_14s0108g01020) and *VvEXPB2* (VIT_12s0059g00190). Furthermore, genes related to cell wall polysaccharide synthesis were also included in the enriched carbohydrate metabolic process category. These results coincide with the decrease of the cell division and the expansion of xylem vascular bundles at prevéraison stage. However, it might also be a first sign of berry softening as discussed by Thomas *et al.*[60] and Matthews *et al.*[61].

Regarding the secondary metabolic process category, we found down-regulated genes involved in the general phenylpropanoid pathway: two phenylalanine ammonia-lyase genes *VvPAL1* and *VvPAL2* (VIT_08s0040g01710 and VIT_13s0019g04460 respectively) and three 4-coumarate-CoA ligase genes (VIT_01s0010g02740;

VIT_11s0052g01090; VIT_16s0039g02040), but also genes related to flavonoid biosynthetic pathway as a chalcone synthase gene *VvCHS2* (VIT_14s0068g00920), three chalcone isomerase genes (VIT_07s0151g01060; VIT_13s0067g02870; VIT_13s0067g03820) and a leucoanthocyanidin dioxgenase gene (VIT_02s0025g04720). The down-regulation of these genes could be associated with a decline of the proanthocyanidin production[62]. In fact, the gene *VvANR*, coding for the anthocyanidin reductase protein responsible for proanthocyanidin synthesis from cyanidin, was statistically down-regulated in all the prevéraison berries except in Sangiovese. In addition, ten glutathione S-transferase genes were identified as down-regulated at prevéraison that we can hypothesize to be involved in tannin export from the site of synthesis in the cytoplasm into the vacuole similarly to anthocyanin transport during the maturation phase[63,64].

Finally, among the carbohydrate metabolic process category, we found down-regulated genes related to starch synthesis usually laid down as granules in chloroplasts in photosynthetic tissues. Hence, this could be associated with the down-regulation of thirteen genes involved in the photosynthetic process, indicating the beginning of photosynthesis shutdown and the breaking down of chloroplast structures, namely the slowdown of the green development began at prevéraison stage.

Prevéraison - End of véraison

Between prevéraison and the end of véraison, 1,772 genes were classified as statistically down-regulated in all ten variety berries. In order to gain insights into biological processes, these genes were analysed for overrepresented functions using the GO enrichment tool BiNGO[8] (**Fig. 5**). Four categories were found significantly enriched: the transcription (GO:0006350), the photosynthesis (GO:0015979), the response to endogenous stimulus (GO:0009719) and the generation of precursor metabolites and energy (GO:0006091) mainly composed of photosynthesis-related genes.

Concerning the overrepresented transcription category, 122 transcription factor genes were identified (**Supplemental table 2**) as genes coding for fourteen basic helix-loop-helix proteins, twenty-five Myb domain proteins and nine WRKY DNA-binding proteins. These results suggest that green-to-maturation transition seems to correspond to a shift of transcriptional program engaging both up-regulation of some transcription factors and the decrease in the expression of other ones.

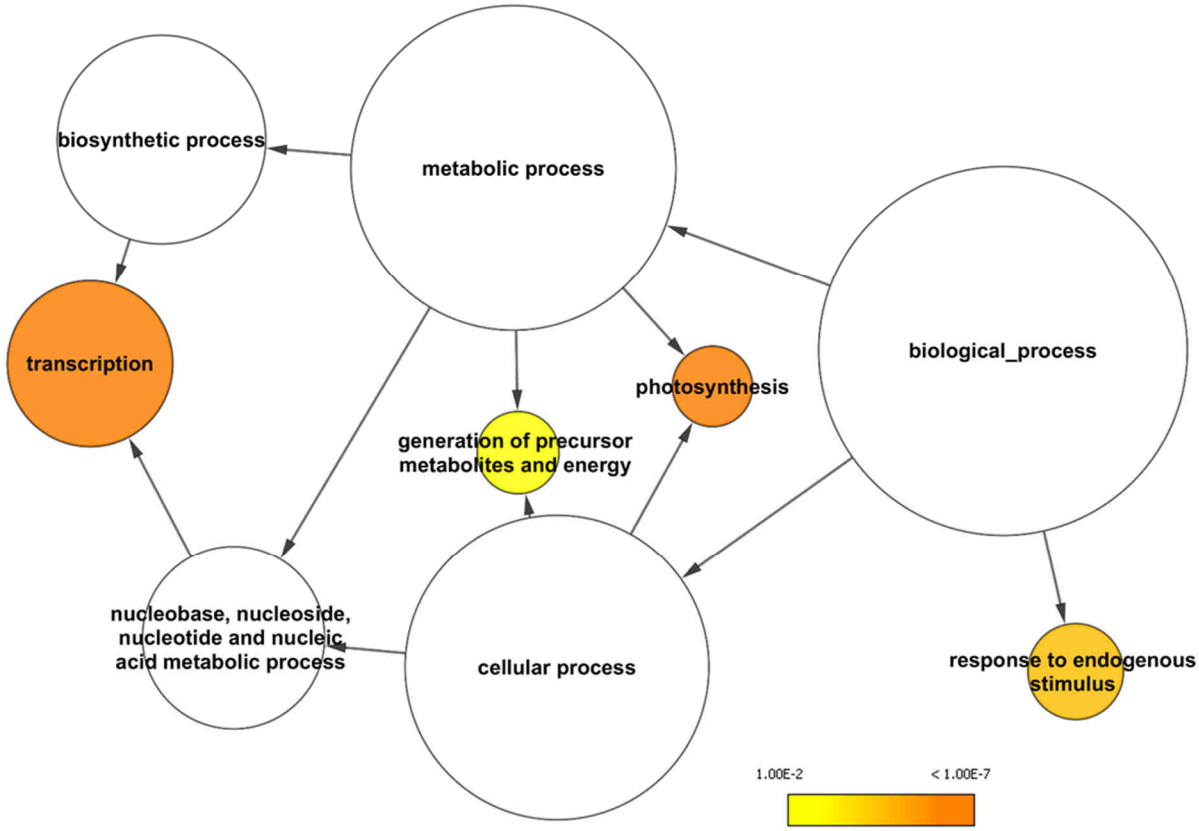


Figure 5: Enriched GO terms in down-regulated genes between prevéraison and the end of véraison. The network graph shows BiNGO visualization of the over-represented GO terms for the 1,772 down-regulated genes between prevéraison and the end of véraison.

The enrichment of the photosynthesis and generation of precursor metabolites and energy categories corroborated the well-known shut-off of the photosynthesis process occurring after véraison.

Finally, among the enriched response to endogenous stimulus category, we identified an indole-3-acetic acid (IAA)-amido synthetase gene (VIT_07s0129g00660), also named *GH3.2* (Gretchen Hagen protein)[65], involved in the synthesis of IAA-Asp conjugates and linked to IAA degradation[66]. Twenty-two genes related to auxin signalling were also found as ten auxin response factor (ARF) and four auxin/IAA protein (Aux/IAA) genes that are transcription factors playing key roles in regulating auxin-responsive transcription in plants[67]. Numerous reports indicated that indole acetic acid levels are much early in berry development, after which they decline steadily to be very low at véraison[68-72]. The high levels early in berry development are in agreement with the proposed role for auxin in cell division and expansion. Moreover, the application of natural or synthetic auxins during the pre-ripening stage of fruit development has often been found to delay ripening or some part of the ripening

process, such as the accumulation of sugars and anthocyanins, as well as the reduction of acidity and chlorophyll levels[73] and therefore auxins are widely viewed as ripening inhibitors[72]. Consequently, the down-regulation of these auxin-related genes at the end of véraison is not surprising if we consider the physiological state of the berries at this stage. Furthermore, fourteen genes related to ethylene signalling were identified as seven ethylene response factor (AP2/ERF) genes: *VvERF10* (VIT_11s0016g00670), *VvERF11* (VIT_04s0008g02230), *VvERF80* (VIT_16s0013g00950), *VvERF82* (VIT_16s0013g00980), *VvERF85* (VIT_16s0013g01070), *VvERF92* (VIT_16s0013g01120) and *VvERF117* (VIT_03s0063g00460). Grape, which is a non-climacteric fruit, is not very sensitive to ethylene; however, ethylene appears to be necessary for normal fruit ripening[74-76]. Ethylene concentrations rise slightly after fruit set, peak just before véraison and then decline to low levels by maturity[77]. Therefore, the down-regulation of ethylene-related genes in ripening berries correlates with the decline of ethylene synthesis during maturation phase. Moreover, Licausi and co-workers[78] demonstrated that grapevine AP2/ERF transcription factor superfamily was composed of 149 genes. In this study, ERF family members were subdivided according to their respective group (I to X) by comparing their expression patterns in the aerial vegetative and reproductive tissues of *Vitis vinifera* Cv. Corvina. Regarding this analysis, *VvERF80*, *VvERF82*, *VvERF85* and *VvERF92* genes belong to the group IX. In Arabidopsis, tobacco (*Nicotiana tabacum*) and tomato (*Solanum lycopersicum*), ERFs belonging to group IX are involved in the response to pathogens[79,80]. Thus, these ERF genes might be more involved in response to biotic stresses rather in organ development regulation. Nonetheless, further investigations are necessary to confirm this hypothesis.

End of véraison - Harvest

Between the end of véraison and harvest stage, 610 down-regulated genes were identified. After analysis for overrepresented functions using the GO enrichment tool BiNGO[8], two significantly overrepresented GO categories were observed: the photosynthesis (GO:0015979) and the generation of precursor metabolites and energy (GO:0006091) mostly composed of photosynthesis-related genes (**Fig. 6**). In fact, these overrepresented categories were mainly composed of genes encoding proteins involved in photosystem I and II structures and functions. As shown in the previous sections, photosynthesis-related genes started being down-regulated right after pea-

sized berry stage and almost all the other ones were completely off by harvest, when berries have entered the maturation phase and utilized exclusively photosynthesis-derived compounds – sucrose for the most – from the leaves. This result suggests that photosynthesis process slowed down gradually throughout the maturation phase.

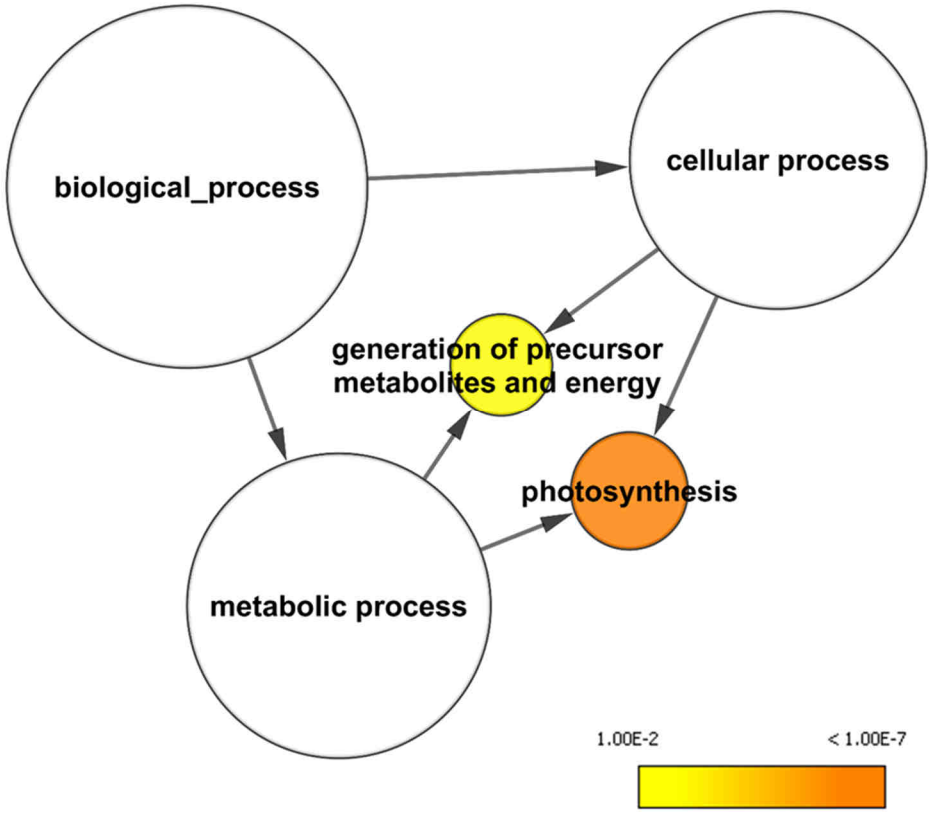


Figure 6: Enriched GO terms in down-regulated genes between the end of véraison and harvest. The network graph shows BiNGO visualization of the over-represented GO terms for the 610 down-regulated genes between the end of véraison and harvest.

Furthermore, down-regulation of genes involved in cell wall expansion and loosening was observed, as a cellulose synthase A1 gene (VIT_04s0008g05220) required for the synthesis of primary cell wall[81], two endo-1,4-D-beta-glucanase genes (VIT_02s0025g01380; VIT_12s0059g01250), three pectinesterase genes (VIT_04s0044g01000, VIT_04s0044g01010, VIT_04s0044g01020; VIT_07s0005g00730; VIT_09s0002g00330) and five expansin genes: *VvEXPA4* (VIT_04s0079g00420), *VvEXPA5* (VIT_06s0004g00070), *VvEXPA11* (VIT_08s0007g00440), *VvEXPA16* (VIT_14s0108g01020) and *VvEXPA18* (VIT_17s0053g00990), suggesting a diminution of cell expansion at harvest. Interestingly, down-regulation of expansin genes was observed above between pea-sized berry stage and prevéraison, indicating that some grape expansin genes, as *VvEXPA6*,

VvEXA13 and *VvEXPB2*, were predominantly expressed during the green phase – probably involved in cell enlargement – whereas other ones were mainly expressed during the maturation phase, e.g. *VvEXPA1* (VIT_01s0026g02620), *VvEXPA5* and *VvEXPA17*, and might be related to berry softening. On the other hand, *VvEXPA16* was identified as significantly down-regulated in prevéraison berries, up-regulated in ripening berries at the end of véraison and then down-regulated in full-ripe berries; this expression pattern suggests its involvement in both developmental phases as previously suggested also by Dal Santo *et al.*[82].

Skin colour-specific DEGs

The DEG comparison allowed identifying 126 and 11 DEGs as specific to red- and white-skinned berries respectively, i.e. genes that were differentially expressed between two subsequent developmental stages in either the five red- or five white-skinned berry samples (**Tables 1 and 2**). These results show that, besides their skin colour, red-skinned berries seemed to have more specific DEGs than white-skinned berries. In addition, the skin colour-specific DEGs between pea-sized berry and prevéraison stages were less numerous, indicating that skin colour seemed to have a greatest influence on the transcriptome during the maturation phase, when indeed berry colouring and other specific compounds biosynthesis happened.

Table 1: Red-skinned berry specific up- and down-regulated genes between each subsequent pair of growth stages. Abbreviations for the four developmental stages: PV (Prevéraison), EV (End of véraison), H (Harvest). An average of the FPKM values of the five red- and five white-skinned berries was done for each stage. A log-transformed fold change (LogFC) was calculated using the FPKM average values.

			Red-skinned berries			White-skinned berries		
Gene_ID	**Functional annotation**		**P**	**PV**	**Log_2-FC**	**P**	**PV**	**Log_2-FC**
VIT_09s0002g08460	8-amino-7-oxononanoate synthase	↓	4.61	2.00	**-1.20**	4.05	3.03	**-0.42**
VIT_09s0002g05570	ABC transporter g family pleiotropic drug resistance 12	↓	1.10	0.48	**-1.19**	0.76	0.63	**-0.28**
VIT_08s0040g02660	RARE-cold-inducible 2A	↓	8.31	0.95	**-3.13**	2.11	0.94	**-1.16**

(Table continues on following page)

Table 1 (Continued from previous page)

Gene_ID	Functional annotation		Red-skinned berries PV	Red-skinned berries EV	Red-skinned berries Log_2-FC	White-skinned berries PV	White-skinned berries EV	White-skinned berries Log_2-FC
VIT_01s0010g03460, VIT_01s0010g03470, VIT_01s0010g03480	Caffeoyl-CoA O-methyltransferase	↑	0.48	27.18	**5.83**	0.67	0.04	**-3.97**
VIT_01s0010g03490	Caffeoyl-CoA O-methyltransferase (AOMT)	↑	0.49	24.30	**5.63**	0.53	0.02	**-4.55**
VIT_01s0010g03510	Caffeoyl-CoA O-methyltransferase (AOMT)	↑	2.11	199.92	**6.57**	1.54	0.09	**-4.04**
VIT_01s0026g02400	Glutathione S-transferase 10	↑	0.27	3.67	**3.74**	0.26	0.28	**0.13**
VIT_03s0017g00870	Transferase family / Anthocyanin Acyl-transferase	↑	0.39	43.83	**6.80**	0.62	0.02	**-5.03**
VIT_03s0017g01020	Methyl-CpG-binding domain 8 MBD08	↑	14.12	17.09	**0.28**	13.79	12.56	**-0.14**
VIT_03s0038g00020	Vacuolar protein sorting 45	↑	7.05	9.82	**0.48**	6.94	6.85	**-0.02**
VIT_04s0079g00690	Glutathione S-transferase (GST4)	↑	0.74	1428.47	**10.92**	0.10	0.15	**0.55**
VIT_05s0136g00260	Chalcone synthase (CHS3)	↑	65.73	981.35	**3.90**	86.73	37.13	**-1.22**
VIT_06s0004g02620	Phenylalanine ammonia-lyase	↑	7.88	60.07	**2.93**	8.68	3.20	**-1.44**
VIT_06s0004g03740	DNA-directed RNA polymerase III subunit C1	↑	6.33	7.59	**0.26**	8.42	6.57	**-0.36**
VIT_06s0004g08150	Trans-cinnamate 4-monooxygenase	↑	46.16	90.20	**0.97**	30.73	8.76	**-1.81**
VIT_06s0009g02810	Flavonoid 3',5'-hydroxylase	↑	5.89	43.92	**2.90**	4.31	0.13	**-5.10**
VIT_06s0009g02830	Flavonoid 3',5'-hydroxylase	↑	10.07	91.81	**3.19**	7.73	0.43	**-4.18**
VIT_06s0009g02840	Flavonoid 3',5'-hydroxylase	↑	5.01	45.95	**3.20**	4.69	0.09	**-5.63**
VIT_06s0009g02860	Flavonoid 3',5'-hydroxylase (F3'5'Hf)	↑	1.67	8.81	**2.39**	1.21	0.11	**-3.44**
VIT_06s0009g02880	Flavonoid 3',5'-hydroxylase	↑	1.25	38.17	**4.93**	1.08	0.02	**-5.87**
VIT_06s0009g02920	Flavonoid 3',5'-hydroxylase	↑	2.80	25.10	**3.17**	2.28	0.17	**-3.75**
VIT_06s0009g02970	Flavonoid 3',5'-hydroxylase	↑	3.32	41.54	**3.65**	2.47	0.05	**-5.68**
VIT_06s0009g03000	Flavonoid 3',5'-hydroxylase	↑	13.18	68.93	**2.39**	10.65	0.94	**-3.50**
VIT_06s0009g03010	Flavonoid 3',5'-hydroxylase	↑	14.07	74.40	**2.40**	11.13	1.28	**-3.12**
VIT_06s0009g03530	Homeodomain protein 14	↑	8.78	10.24	**0.22**	8.71	7.61	**-0.19**
VIT_07s0005g03730	Protein phosphatase 2, regulatory subunit B''	↑	25.18	27.73	**0.14**	24.61	21.34	**-0.21**
VIT_08s0007g03560	Anthocyanin membrane protein 1 (Anm1)	↑	0.15	15.85	**6.71**	0.32	0.05	**-2.70**
VIT_08s0040g00440	Unknown protein	↑	9.99	34.04	**1.77**	10.81	2.67	**-2.02**
VIT_10s0042g01390	Pentatricopeptide (PPR) repeat-containing protein	↑	1.02	1.65	**0.69**	1.25	1.25	**0.00**
VIT_11s0016g03450	Unknown protein	↑	12.13	14.84	**0.29**	13.26	12.09	**-0.13**
VIT_13s0019g04460	Phenylalanine ammonia-lyase 2 (PAL2)	↑	4.54	63.59	**3.81**	5.63	2.31	**-1.28**
VIT_15s0046g02620	Beta-amylase 9 BMY9	↑	7.36	9.65	**0.39**	8.25	7.36	**-0.16**
VIT_16s0039g02230	UDP-glucose:flavonoid 3-O-glucosyltransferase	↑	0.06	123.04	**11.05**	0.01	0.03	**1.78**
VIT_18s0001g06750	Steroleosin-B	↑	0.32	2.23	**2.80**	0.87	0.07	**-3.55**
VIT_18s0001g14540	DEAD-box ATP-dependent RNA helicase 28	↑	16.71	19.11	**0.19**	21.74	16.34	**-0.41**
VIT_19s0015g00830, VIT_19s0015g00840	Proliferating cell nuclear protein P120	↑	55.92	64.59	**0.21**	59.48	49.63	**-0.26**
VIT_19s0090g01650	LOP1 TRN1 (LOPPED 1, TORNADO 1)	↑	0.65	1.78	**1.45**	0.73	0.79	**0.12**
VIT_00s0838g00020	Ribosomal protein S5	↓	83.56	45.22	**-0.89**	74.75	50.53	**-0.57**
VIT_02s0012g01800	Shikimate kinase	↓	12.68	5.71	**-1.15**	11.36	7.53	**-0.59**
VIT_03s0091g00050	Triacylglycerol/steryl ester hydrolase	↓	21.11	9.98	**-1.08**	18.35	11.35	**-0.69**
VIT_04s0023g03860	Rac-like GTP-binding protein ARAC1 precursor	↓	32.49	15.96	**-1.03**	30.05	21.16	**-0.51**
VIT_05s0020g02460	Cobalt ion transporter	↓	6.54	2.81	**-1.22**	6.43	3.83	**-0.75**
VIT_07s0005g02540, VIT_07s0005g02550	Protoporphyrinogen oxidase (PPOX)	↓	21.24	11.49	**-0.89**	17.81	13.07	**-0.45**
VIT_07s0031g02460	Unknown protein	↓	22.54	12.00	**-0.91**	23.73	16.54	**-0.52**
VIT_07s0129g00090	Ribosomal protein L26 (RPL26A) 60S	↓	455.94	248.09	**-0.88**	418.45	313.53	**-0.42**
VIT_08s0007g05040	Unknown protein	↓	25.50	11.20	**-1.19**	24.38	15.41	**-0.66**
VIT_08s0056g01690	LrgB-like family protein	↓	9.27	4.79	**-0.95**	8.13	7.75	**-0.07**
VIT_14s0068g00750	C2 domain-containing protein	↓	37.13	20.08	**-0.89**	34.61	24.44	**-0.50**
VIT_15s0048g02190	PAP/fibrillin family	↓	8.27	4.32	**-0.94**	7.88	5.14	**-0.62**
VIT_16s0039g02260	RelA/SpoT protein (RSH3)	↓	21.85	13.05	**-0.74**	20.36	15.22	**-0.42**
VIT_16s0039g02530	No hit	↓	41.84	19.49	**-1.10**	41.92	33.72	**-0.31**
VIT_16s0039g02540	Rubredoxin	↓	68.76	32.62	**-1.08**	62.67	50.85	**-0.30**
VIT_17s0000g02160	Unknown protein	↓	34.21	16.98	**-1.01**	22.82	16.44	**-0.47**
VIT_17s0000g07790	N-hydroxythioamide S-beta-glucosyltransferase	↓	79.49	45.02	**-0.82**	79.72	56.52	**-0.50**
VIT_18s0089g00020	Phosphatidylinositol N-acetylglucosaminyl-transferase subunit P	↓	12.57	5.99	**-1.07**	10.85	6.46	**-0.75**

(Table continues on following page)

Table 1 (Continued from previous page)

			Red-skinned berries			White-skinned berries		
Gene_ID	**Functional annotation**		**EV**	**H**	**Log_2-FC**	**EV**	**H**	**Log_2-FC**
VIT_00s0225g00050	MATE efflux family protein	↑	0.19	0.91	**2.24**	0.09	0.25	**1.56**
VIT_00s0226g00100	Regulator of chromosome condensation 1	↑	1.17	5.54	**2.25**	0.40	0.56	**0.50**
VIT_01s0011g06140	Avr9/Cf-9 rapidly elicited protein	↑	2.07	18.16	**3.13**	0.79	0.53	**-0.56**
VIT_01s0150g00190	Transaldolase	↑	162.24	210.71	**0.38**	151.45	165.31	**0.13**
VIT_02s0025g00530	Unknown	↑	0.08	2.45	**4.95**	0.05	0.22	**2.16**
VIT_03s0038g02400	Protein kinase	↑	1.44	5.14	**1.84**	1.21	0.50	**-1.29**
VIT_03s0063g02250	VvPMT5 Polyol transporter 5	↑	2.60	27.85	**3.42**	1.26	1.08	**-0.22**
VIT_03s0063g02650	Hydrolase, alpha/beta fold	↑	0.03	12.68	**8.54**	0.01	0.08	**2.59**
VIT_04s0069g00920	WRKY DNA-binding protein 11	↑	6.56	26.91	**2.04**	4.30	3.27	**-0.39**
VIT_06s0004g03100	MLO1	↑	18.53	40.89	**1.14**	12.36	12.59	**0.03**
VIT_06s0004g04620	Unknown	↑	2.23	15.54	**2.80**	1.58	1.41	**-0.17**
VIT_06s0004g06750	Armadillo/beta-catenin repeat	↑	7.64	12.69	**0.73**	6.05	6.45	**0.09**
VIT_06s0004g07120	Gamma interferon responsive lysosomal thiol reductase family protein	↑	23.57	38.48	**0.71**	33.85	34.13	**0.01**
VIT_07s0005g04620	Unknown	↑	0.52	6.63	**3.67**	0.55	0.62	**0.18**
VIT_07s0031g01320	TGA-type basic leucine zipper protein TGA1.1	↑	3.25	5.51	**0.76**	2.98	3.00	**0.01**
VIT_07s0031g01980	VvERF113_ERF/AP2 Gene Family	↑	0.19	2.70	**3.86**	0.13	0.17	**0.39**
VIT_07s0031g03060	NSL1 (necrotic spotted lesions 1)	↑	10.44	45.92	**2.14**	8.04	9.13	**0.18**
VIT_07s0104g00850	Pto kinase interactor	↑	0.75	14.97	**4.33**	0.28	0.43	**0.59**
VIT_08s0007g00030	Cation/hydrogen exchanger 20 (CHX20)	↑	0.33	2.87	**3.10**	0.18	0.12	**-0.61**
VIT_08s0007g01230	Lectin protein kinase	↑	0.61	2.68	**2.13**	0.18	0.28	**0.61**
VIT_09s0002g05600	ABC transporter g family pleiotropic drug resistance 12	↑	58.10	208.72	**1.85**	49.68	48.63	**-0.03**
VIT_12s0028g00610	Aspartyl protease	↑	2.61	14.87	**2.51**	1.55	1.46	**-0.09**
VIT_12s0055g00340	WRKY DNA-binding protein 9	↑	0.10	1.48	**3.88**	0.01	0.10	**4.22**
VIT_13s0019g00310	ABC transporter B member 20	↑	14.75	23.77	**0.69**	13.05	14.26	**0.13**
VIT_13s0019g01710	Scarecrow transcription factor 14 (SCL14)	↑	0.24	2.31	**3.25**	0.16	0.20	**0.32**
VIT_14s0066g01370	No hit	↑	2.35	5.63	**1.26**	3.68	4.69	**0.35**
VIT_14s0068g01960	Acetolactate synthase III, chloroplast precursor	↑	27.76	42.58	**0.62**	28.49	27.22	**-0.07**
VIT_17s0000g04370	Wall-associated kinase	↑	1.88	18.18	**3.27**	1.04	1.25	**0.26**
VIT_17s0000g06930	Unfertilized embryo sac 10 UNE10	↑	0.37	1.22	**1.72**	0.12	0.27	**1.14**
VIT_18s0001g14410	Unknown protein	↑	10.01	22.57	**1.17**	9.06	8.69	**-0.06**
VIT_18s0001g14690	Protein kinase	↑	0.67	5.28	**2.98**	0.53	0.18	**-1.57**
VIT_18s0166g00190	U-box domain-containing protein	↑	1.83	15.75	**3.11**	1.08	0.84	**-0.36**
VIT_00s0322g00030	Alpha-6-galactosyltransferase	↓	255.98	142.55	**-0.84**	190.64	231.37	**0.28**
VIT_01s0010g03460, VIT_01s0010g03470, VIT_01s0010g03480	Caffeoyl-CoA O-methyltransferase	↓	27.18	4.35	**-2.64**	0.04	0.00	**-3.65**
VIT_01s0010g03490	Caffeoyl-CoA O-methyltransferase (AOMT)	↓	24.30	4.22	**-2.53**	0.02	0.01	**-0.91**
VIT_01s0010g03510	Caffeoyl-CoA O-methyltransferase (AOMT)	↓	199.92	29.37	**-2.77**	0.09	0.02	**-2.56**
VIT_01s0011g01290	N-carbamyl-L-amino acid amidohydrolase	↓	56.32	39.79	**-0.50**	54.73	51.04	**-0.10**
VIT_02s0012g01440	Zinc finger (C3HC4-type ring finger)	↓	8.53	3.17	**-1.43**	6.72	7.17	**0.09**
VIT_02s0025g04380	ATPP2-A13	↓	25.73	14.12	**-0.87**	17.53	15.40	**-0.19**
VIT_02s0033g00380	Myb VvMYBA2	↓	262.43	107.73	**-1.28**	74.61	136.24	**0.87**
VIT_02s0033g00390	Myb domain protein 113	↓	158.06	62.83	**-1.33**	57.84	101.01	**0.80**
VIT_02s0033g00410	Myb VvMYBA1	↓	71.21	18.66	**-1.93**	9.86	15.27	**0.63**
VIT_03s0017g00870	Transferase family / Anthocyanin Acyl-transferase	↓	43.83	9.06	**-2.27**	0.02	0.01	**-1.64**
VIT_03s0088g00260	Serine carboxypeptidase S10	↓	17.06	3.30	**-2.37**	0.28	0.21	**-0.44**
VIT_04s0008g00650	3'-5' exonuclease domain-containing protein	↓	310.05	204.42	**-0.60**	324.95	301.13	**-0.11**
VIT_04s0079g00690	Glutathione S-transferase GST4	↓	1428.47	621.73	**-1.20**	0.15	0.18	**0.33**
VIT_05s0020g00600	1-cys peroxiredoxin	↓	5.84	1.59	**-1.87**	1.21	0.95	**-0.35**
VIT_05s0051g00860	PTAC5 (plastid transcriptionally active5)	↓	8.29	4.33	**-0.94**	9.71	10.00	**0.04**
VIT_06s0004g03250	Protein disulfide isomerase	↓	66.17	48.32	**-0.45**	58.50	54.20	**-0.11**
VIT_06s0009g02810	Flavonoid 3',5'-hydroxylase	↓	43.92	15.86	**-1.47**	0.13	0.02	**-2.71**
VIT_06s0009g02840	Flavonoid 3',5'-hydroxylase	↓	45.95	14.66	**-1.65**	0.09	0.03	**-1.83**
VIT_06s0009g02860	flavonoid 3',5'-hydroxylase (F3'5'Hf)	↓	8.81	2.70	**-1.71**	0.11	0.04	**-1.55**
VIT_06s0061g00230	ABC transporter G member 7	↓	28.74	17.93	**-0.68**	23.85	26.89	**0.17**
VIT_07s0104g01430	Phototropic-responsive NPH3	↓	25.94	13.65	**-0.93**	22.76	18.59	**-0.29**
VIT_08s0007g00530	3-beta hydroxysteroid dehydrogenase/isomerase protein	↓	105.21	53.56	**-0.97**	112.42	117.93	**0.07**
VIT_08s0007g01640	Fructokinase (FRK2)	↓	9.53	3.63	**-1.39**	10.20	10.02	**-0.02**
VIT_08s0007g03560	Anthocyanin membrane protein 1 (Anm1)	↓	15.85	4.06	**-1.97**	0.05	0.02	**-1.31**
VIT_08s0056g00790	Phagocytosis and cell motility protein ELMO1	↓	20.86	14.38	**-0.54**	20.05	18.63	**-0.11**
VIT_10s0003g04180	Early-responsive to dehydration stress protein 4	↓	349.28	186.66	**-0.90**	246.84	303.52	**0.30**
VIT_11s0052g00680	Unknown protein	↓	50.20	33.21	**-0.60**	59.33	53.42	**-0.15**
VIT_13s0064g01280	SWAP (Suppressor-of-White-APricot)	↓	149.86	109.45	**-0.45**	144.08	136.22	**-0.08**
VIT_14s0066g01780	Heavy-metal-associated domain-containing protein	↓	25.49	14.10	**-0.85**	31.93	24.54	**-0.38**
VIT_14s0068g02040	Gda-1	↓	188.55	118.34	**-0.67**	173.79	167.99	**-0.05**
VIT_16s0039g02230	UDP-glucose:flavonoid 3-O-glucosyltransferase	↓	123.04	35.91	**-1.78**	0.03	0.06	**0.86**
VIT_16s0098g00980	Unknown	↓	9.01	5.25	**-0.78**	7.78	6.19	**-0.33**
VIT_17s0000g01080	HVA22E (HVA22-like protein E)	↓	199.24	83.53	**-1.25**	130.18	136.53	**0.07**
VIT_17s0000g09200	Unknown protein	↓	32.29	19.51	**-0.73**	33.87	27.87	**-0.28**
VIT_18s0001g09400	Cytochrome b5 DIF-F (VvCytoB5)	↓	295.75	76.84	**-1.94**	3.24	3.71	**0.20**
VIT_18s0072g00380	Abscisic stress ripening protein 2 (ASR2)	↓	12228.74	5078.71	**-1.27**	13071.60	9677.34	**-0.43**
VIT_18s0157g00200	Steroleosin-B	↓	12.59	7.07	**-0.83**	17.99	17.34	**-0.05**
VIT_19s0140g00210	SOUL heme-binding	↓	157.55	76.52	**-1.04**	209.60	220.81	**0.08**

Regarding red-skinned berry-specific DEGs, the more marked down-regulation concerned a gene orthologous of RARE-cold-inducible 2A (*RSI2A*) of *Arabidopsis thaliana* in prevéraison red grapes in comparison to the white ones. RSI2A is characterized by a small molecular size and an extreme degree of hydrophobicity[83]. In a study, Medina and co-workers[84] suggested that *RCI2A* and its homologue *RCI2B* were regulated at transcriptional level during Arabidopsis development and in response to different environmental stimuli and treatments including low temperature, dehydration, and exogenous abscisic acid. The specific down-regulation of this gene, and more particularly its higher expression level in pea-sized berries of red grape varieties, suggests that red grapes underwent more drought stress at pea-sized berry stage.

Most of the up-regulated genes between prevéraison and the end of véraison (**Table 1**) were involved in phenylpropanoid, flavonoid and anthocyanin biosynthetic pathways with the high fold change up-regulation of the UDP-glucose:flavonoid 3-*O*-glucosyl transferase gene *VvUFGT* (VIT_16s0039g02230) responsible for anthocyanin synthesis[85,86], the glutathione S-transferase gene *VvGST4* (VIT_04s0079g00690) involved in anthocyanin transport to the vacuole[87,88]. Nevertheless, we noticed the absence of the transcription factor gene *VvMybA1* which expression is normally specific to red grape varieties and crucial for anthocyanin production[89,90]. In fact, due to a mismapping issue, *VvMybA2* reads were attributed to *VvMybA1* in white-skinned berries. This trouble will be explained exhaustively in **chapter 9**. Interestingly, observing the gene expression profiles, the UDP-glucose:flavonoid 3-*O*-glucosyl transferase *VvUFGT*, the Glutathione S-transferase gene *VvGST4*, three flavonoid 3',5'-hydroxylase genes (VIT_06s0009g02840; VIT_06s0009g02860; VIT_06s0009g02810) and two anthocyanin *O*-methyltransferase genes (VIT_01s0010g03490; VIT_01s0010g03510) were found down-regulated in all the five red-skinned berries between the end of véraison and harvest, suggesting a decrease of anthocyanin biosynthesis and transport between the two ripening stages. On the other hand, it is possible that transcriptional down-regulation would not correspond to a direct slowdown of the pathway and that translated enzymes keep working longer and metabolites get accumulated.

Concerning white-skinned berry-specific DEGs (**Table 2**), the down-regulation of a phytosulfokine gene (*PSK4*) was noticed between the end of véraison and harvest. The autocrine growth factor phytosulfokine (PSK) is a 5-amino acid disulfated peptide of the sequence Tyr(SO3H)-Ile-Tyr(SO3H)-Thr-Gl. PSK is produced from pre-proteins

encoded by five genes in *Arabidopsis thaliana* that are expressed throughout the plant life cycle, suggesting that PSK plays a ubiquitous role in plant growth and development[91-94]. *In planta*, PSK signalling participates in the control of root and shoot growth. Root growth and hypocotyl elongation are promoted by PSK, mainly through the signalling of cell elongation rather than cell division[94-96]. However, Igarashi and co-workers[97] demonstrated that PSK attenuates pattern-triggered immunity signalling in Arabidopsis. Therefore, we can hesitate about the role of this growth hormone during berry ripening, which could be involved in either cell elongation or response to biotic stress and thus gene functional analyses seem to be necessary to understand the biological role of *PSK4* down-regulation in white-skinned berries.

Table 2: White-skinned berry specific up- and down-regulated genes between each subsequent pair of developmental stages. Abbreviations for the four developmental stages: PV (Prevéraison), EV (End of véraison), H (Harvest). An average of the FPKM values of the five red- and five white-skinned berries was done for each stage. A log-transformed fold change (LogFC) was calculated using the FPKM average values.

			White-skinned berries			Red-skinned berries		
Gene_ID	**Functional annotation**		**P**	**PV**	**Log_2-FC**	**P**	**PV**	**Log_2-FC**
VIT_01s0010g02750	Lipoxygenase	↓	29.87	21.56	**-0.47**	25.67	24.73	**-0.05**

			White-skinned berries			Red-skinned berries		
Gene_ID	**Functional annotation**		**PV**	**EV**	**Log_2-FC**	**PV**	**EV**	**Log_2-FC**
VIT_04s0008g05270	ATPase AFG1	↑	13.12	14.83	**0.18**	13.49	11.38	**-0.24**
VIT_09s0002g05990	6-4 photolyase	↑	8.56	11.72	**0.45**	10.09	4.93	**-1.03**
VIT_13s0067g02560	Unknown protein	↑	39.61	58.59	**0.56**	50.77	35.06	**-0.53**
VIT_01s0010g03510	Caffeoyl-CoA O-methyltransferase	↓	1.54	0.09	**-4.04**	2.11	199.92	**6.57**
VIT_07s0005g02660	ABC transporter B member 13	↓	10.13	4.91	**-1.05**	8.61	8.96	**0.06**

			White-skinned berries			Red-skinned berries		
Gene_ID	**Functional annotation**		**EV**	**H**	**Log_2-FC**	**EV**	**H**	**Log_2-FC**
VIT_00s1291g00030	Receptor protein kinase PERK1	↑	24.75	39.62	**0.68**	38.28	24.96	**-0.62**
VIT_13s0084g00240	Steroid sulfotransferase	↑	16.26	31.06	**0.93**	12.78	10.02	**-0.35**
VIT_00s1328g00020	Phytosulfokine PSK4	↓	9.84	3.31	**-1.57**	8.50	24.03	**1.50**
VIT_04s0044g00050	Protein kinase APK1B	↓	4.95	2.61	**-0.93**	5.30	7.41	**0.48**
VIT_06s0061g00860	Zinc-binding protein	↓	11.22	6.88	**-0.71**	11.78	12.82	**0.12**

CONCLUSION

Comparing the differentially expressed genes (DEGs) identified for each variety in **chapter 4** allowed selecting the genes that were commonly modulated – both up- and down-regulated – in all varieties, or exclusively in the five red- or white-skinned berries. From this comparison, a remarkable transcriptional modulation after véraison was observed with more genes being down- rather than up-regulated during berry development. Moreover, red-skinned berries seemed to have a higher transcriptional

modulation after véraison whereas white-skinned berries seemed to be slightly more transcriptionally active during the green development.

The analysis of the commonly up- and down-regulated genes permitted to highlight the main biological processes occurring during berry development and to identify the genes involved in these biological functions. For example, gene expression modulation of photosynthesis-related genes indicated that photosynthesis process started to decrease right after pea-sized berry stage, then dramatically declined after the onset of ripening, and finally slowed down gradually throughout the maturation phase, confirming that after véraison berries utilized exclusively photosynthesis-derived compounds – sucrose for the most – from the leaves. Moreover, the up-regulation of sucrose and hexose transporter genes at the end of véraison corroborates with the remarkable sugar accumulation occurring during the ripening phase, which is coupled with the up-regulation of aquaporin genes involved in water influx – responsible for cell expansion and berry softening. Interestingly, glycolysis-related genes were found as up-regulated between prevéraison and the end of véraison, confirming that glycolysis could actually not be inhibited during ripening and that sugars, and not malate, might provide the bulk of the deficit in substrate utilised by respiration and ethanolic fermentation in the pericarp of grape berries during most of ripening[14]. A future work could consist in carrying out berry phenotype comparison in order to observe the influence of expression level variation of selected genes – identified as involved in the main biological processes – on berry physiology.

Finally, obtaining skin colour specific DEGs permitted to find out that red-skinned berries, besides their skin colour, had more specific DEGs than white-skinned berries throughout development. In addition, skin colour-specific DEGs between pea-sized berry and prevéraison stages were few numerous, indicating that skin colour seemed to have a greatest influence on the transcriptome during the maturation phase. This point will be discussed exhaustively in **chapter 8**.

METHODS

GO Enrichment analysis

GO enrichment analysis on up- and down-regulated genes at each growth stage change was performed using BiNGO 2.4 plug-in tool in Cytoscape version 3.1.0 with PlantGOslim categories, as described by Maere *et al.*[8]. Overrepresented PlantGOslim categories were identified using a hypergeometric test with a significance threshold of 0.01 after a Benjamini & Hochberg False-Discovery Rate (FDR) correction[13].

SUPPLEMENTARY MATERIAL

Supplementary table 1: Up-regulated genes between the end of véraison and harvest among the significantly enriched secondary metabolic process category. An average of the FPKM values of the five red- and five white-skinned berries was done for the end of véraison stage (EV) and harvest stage (H).

		Red-skinned berries		White-skinned berries	
Gene_ID	**Functional annotation**	**EV**	**H**	**EV**	**H**
VIT_02s0109g00250	4-coumarate-CoA ligase	0.83	2.18	0.76	3.80
VIT_16s0039g01170	Phenylalanine ammonium lyase	2.90	86.09	0.49	3.93
VIT_16s0100g00780	VvSTS10_Stilbene synthase	2.63	142.57	0.63	3.73
VIT_16s0100g00900	VvSTS20_Stilbene synthase	3.33	233.59	1.41	6.57
VIT_16s0100g00910	VvSTS21_Stilbene synthase	5.57	266.02	1.15	5.73
VIT_16s0100g01000	VvSTS28_Stilbene synthase	6.34	323.93	2.65	11.07
VIT_16s0100g01030	VvSTS31_Stilbene synthase	2.88	202.79	1.03	5.91
VIT_16s0100g01070	VvSTS35_Stilbene synthase	3.65	125.64	0.69	5.18
VIT_16s0100g01200	VvSTS48_Stilbene synthase	7.20	270.00	1.68	13.31
VIT_16s0100g00750	VvSTS7_Stilbene synthase	4.94	91.62	0.84	5.31

Supplementary table 2: Down-regulated transcription factor genes between prevéraison and the end of véraison.

Gene_ID	Funtional annotation
VIT_15s0021g01880	Arabidopsis thaliana homeobox protein 2
VIT_17s0000g08480	ATMYB66/WER/WER1 (WEREWOLF 1)
VIT_04s0008g01870	ATMYB66/WER/WER1 (WEREWOLF 1)
VIT_09s0002g01380	ATMYB66/WER/WER1 (WEREWOLF 1)
VIT_08s0007g07870	Basic helix-loop-helix (bHLH) family
VIT_00s0274g00070	Basic helix-loop-helix (bHLH) family
VIT_13s0067g01350	Basic helix-loop-helix (bHLH) family
VIT_00s1312g00010	Basic helix-loop-helix (bHLH) family
VIT_12s0028g03550	Basic helix-loop-helix (bHLH) family
VIT_01s0010g00740	Basic helix-loop-helix (bHLH) family
VIT_05s0124g00240	Basic helix-loop-helix (bHLH) family
VIT_19s0014g04670	Basic helix-loop-helix (bHLH) family
VIT_13s0067g02280	Basic helix-loop-helix (bHLH) family
VIT_11s0052g00100	Basic helix-loop-helix (bHLH) family
VIT_00s1314g00010	Basic helix-loop-helix (bHLH) family
VIT_05s0020g04780	Basic helix-loop-helix (bHLH) family
VIT_15s0046g02560	basic helix-loop-helix (VvMycA1)
VIT_02s0012g01450	Basic helix-loop-helix BHLH071
VIT_18s0001g13040	Basic leucine-zipper 11 GBF6
VIT_02s0025g00200	BEL1 (BELL 1)
VIT_06s0009g00410	BLH1 (embryo sac development arrest 29)
VIT_04s0008g02750	BZIP transcription factor BZO2H2
VIT_05s0077g01240	Calmodulin-binding transcription activator 6
VIT_03s0038g02130	Cold shock protein-1
VIT_06s0004g06710	Cys-3-His zinc finger protein
VIT_13s0067g01070	Cys-3-His zinc finger protein
VIT_04s0023g01610	DDM1 (decreased DNA methylation 1)
VIT_16s0098g01420	DOF affecting germination 1
VIT_09s0002g02490	Dof zinc finger protein 1
VIT_10s0003g00040	Dof zinc finger protein DOF1.4
VIT_02s0025g03450	Enhancer of GLABRA3
VIT_19s0014g02240	Ethylene responsive element binding factor 4
VIT_08s0007g07550	GATA transcription factor 11
VIT_03s0038g00490	GATA transcription factor 12
VIT_15s0021g02510	GATA transcription factor 2
VIT_12s0028g03100	GPRI1 (GOLDEN2 1)
VIT_16s0098g01080	Growth-regulating factor 7
VIT_17s0000g00270	GT2-like trihelix DNA-binding protein
VIT_11s0016g03940	Heat shock transcription factor C1
VIT_09s0002g03740	Homeobox gene 8
VIT_08s0007g06670	Homeobox-leucine zipper protein 14
VIT_00s0732g00010	Homeobox-leucine zipper protein 22
VIT_18s0001g08410	Homeobox-leucine zipper protein 22
VIT_16s0098g01170	Homeobox-leucine zipper protein HB-12
VIT_01s0026g01950	Homeobox-leucine zipper protein HB13
VIT_10s0003g04670	Homeobox-leucine zipper transcription factor PHABULOSA
VIT_16s0100g00670	Homeodomain GLABROUS1
VIT_01s0026g01550	Homeodomain leucine zipper protein HB-1
VIT_17s0000g05630	Homeodomain leucine zipper protein HB-1
VIT_14s0060g02440	Indeterminate(ID)-domain 2
VIT_11s0118g00800	KNAT2 (knotted1-like homeobox gene 3)
VIT_14s0108g00760	Mini zinc finger 2 MIF2
VIT_19s0085g00890	Myb APL (altered phloem development)
VIT_10s0116g00500	Myb caprice CPC
VIT_12s0059g02360	Myb caprice CPC
VIT_00s0203g00070	Myb domain protein 102
VIT_12s0134g00490	Myb domain protein 14
VIT_17s0000g09080	Myb domain protein 4
VIT_17s0000g02650	Myb domain protein 7
VIT_09s0002g01400	Myb domain protein 7
VIT_07s0129g01050	Myb domain protein 73
VIT_18s0001g11170	Myb domain protein 73
VIT_01s0026g01910	Myb domain protein 88
VIT_08s0007g00410	Myb domain protein 91
VIT_14s0083g00120	Myb domain protein 91
VIT_03s0180g00210	Myb domain protein R1
VIT_01s0026g01050	Myb family
VIT_01s0011g03110	Myb family
VIT_09s0054g01620	Myb family
VIT_07s0197g00060	Myb family
VIT_04s0023g01910	Myb family
VIT_02s0033g00300	Myb family
VIT_04s0043g00340	Myb KAN1 (KANADI 1)
VIT_16s0050g02530	Myb Triptychon
VIT_11s0016g01300	MYBPAR
VIT_06s0004g01870	No transmitting tract
VIT_11s0016g01480	Nuclear transcription factor Y subunit A-3
VIT_07s0031g01460	Nuclear transcription factor Y subunit B-3
VIT_19s0015g00440	Nuclear transcription factor Y subunit B-3
VIT_18s0001g11310	OBF binding protein 1
VIT_00s0253g00060	OBF binding protein 2
VIT_14s0068g00330	PTF1 (plastid transcription factor 1) TCP13
VIT_08s0007g03820	PTL (PETAL LOSS)
VIT_13s0019g01710	Scarecrow transcription factor 14 (SCL14)
VIT_06s0004g04980	Scarecrow transcription factor 14 (SCL14)
VIT_13s0019g01220	Scarecrow transcription factor 3 (SCL3)
VIT_06s0004g04960	Scarecrow-like transcription factor 14 SCL14
VIT_04s0023g01660	Scarecrow-like transcription factor 7 (SCL7)
VIT_10s0003g03910	TCP family transcription factor 24
VIT_12s0028g02520	TCP family transcription factor 4
VIT_10s0042g00170	TCP family transcription factor 7
VIT_17s0000g06020	TCP family transcription factor TCP14
VIT_14s0068g01690	TCP family transcription factor TCP20
VIT_12s0035g00690	TCP family transcription factor TCP23
VIT_07s0031g00780	Transcriptional factor B3
VIT_05s0020g03880	TSO1 (chinese for 'ugly')
VIT_07s0104g01290	TUBBY like protein 3 TLP3
VIT_05s0020g01930	TUBBY like protein 3 TLP3
VIT_18s0001g12660	TUBBY like protein 6 TLP6
VIT_19s0014g01250	TUBBY like protein 7 TLP7
VIT_04s0023g03710	VvMYB4b
VIT_14s0006g01620	VvMYBC2-L3
VIT_07s0031g02610	VvNAC39_NAC domain-containing protein
VIT_03s0038g03410	VvNAC41_NAC domain-containing protein
VIT_15s0048g02280	VvNAC54_NAC domain-containing protein
VIT_06s0080g00780	VvNAC74_NAC domain-containing protein
VIT_12s0028g03350	VvSBP11_Squamosa promoter-binding protein
VIT_19s0014g02350	VvSBP18_Squamosa promoter-binding protein
VIT_01s0010g03710	VvSBP2_Squamosa promoter-binding protein
VIT_08s0007g06270	VvSBP8_Squamosa promoter-binding protein
VIT_10s0003g00050	VvSBP9_Squamosa promoter-binding protein
VIT_04s0069g00920	WRKY DNA-binding protein 11
VIT_07s0031g00080	WRKY DNA-binding protein 15
VIT_15s0046g02190	WRKY DNA-binding protein 22
VIT_08s0058g00690	WRKY DNA-binding protein 33
VIT_09s0018g00240	WRKY DNA-binding protein 40
VIT_15s0046g01140	WRKY DNA-binding protein 53
VIT_02s0025g01280	WRKY DNA-binding protein 53
VIT_19s0090g00840	WRKY DNA-binding protein 6
VIT_08s0058g01390	WRKY DNA-binding protein 70
VIT_18s0122g01140	Wuschel-related homeobox 13
VIT_04s0044g00150	Wuschel-related homeobox 13
VIT_06s0004g03070	Zinc finger (C2H2 type) JACKDAW
VIT_04s0023g01880	Zinc finger homeobox 31

REFERENCES

1. Oshlack A., Robinson M.D., Young M.D. From RNA-seq reads to differential expression results. Genome Biol. 11:220 (2010)
2. Agarwal A., Koppstein D., Rozowsky J., Sboner A., Habegger L., Hillier L.W., Sasidharan R., Reinke V., Waterston R.H., Gerstein M. Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. BMC Genomics 11:383 (2010)
3. Bradford J.R., Hey Y., Yates T., Li Y., Pepper S.D., Miller C.J. A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. BMC Genomics 11:282 (2010)
4. Kennedy J. Understanding grape berry development. Practical Winery & Vineyard (2002)
5. Conde C., Silva P., Fontes N., et al. Biochemical changes throughout grape berry development and fruit and wine quality. Food 1:1-22 (2007)
6. Trapnell C., Hendrickson D.G., Sauvageau M., Goff L., Rinn JL., et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 31:46-53 (2012)
7. Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 28:511-515 (2010)
8. Maere S., Heymans K., Kuiper M. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21:3448-3449 (2005)
9. Galego L. & Almeida J. Role of DIVARICATA in the control of dorsoventral asymmetry in Antirrhinum flowers. Genes Dev. 16(7):880-891 (2002)
10. Machemer K., Shaiman O., Salts Y., Shabtai S., Sobolev I., Belausov E., Grotewold E., Barg R. Interplay of MYB factors in differential cell expansion, and consequences for tomato fruit development. Plant J. 68(2):337-350 (2011)
11. Suetsugu N., Kagawa T., Wada M. An auxilin-like J-domain protein, JAC1, regulates phototropin-mediated chloroplast movement in Arabidopsis. Plant Physiol. 139(1):151-162 (2005)
12. Wada M., Kagawa T., Sato Y. Chloroplast movement. Annu Rev Plant Biol. 54:455-468 (2003)
13. Klipper-Aurbach Y., Wasserman M., Braunspiegel-Weintrob N., Borstein D., Peleg S., Assa S., Karp M., Benjamini Y., Hochberg Y., Laron Z. Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. Med. Hypotheses 45:486-490 (1995)
14. Famiani F., Farinelli D., Palliotti A., Moscatello S., Battistelli A., Walker R.P. Is stored malate the quantitatively most important substrate utilised by respiration and ethanolic fermentation in grape berry pericarp during ripening? Plant Physiol Biochem. 76:52-57 (2014)
15. Davies C., Wolf T., Robinson S.P. Three putative sucrose transporters are differentially expressed in grapevine tissues. Plant Sci. 147:93-100 (1999)
16. Afoufa-Bastien D., Medici A., Jeauffre J., Coutos-Thévenot P., Lemoine R., Atanassova R., Laloi M. The *Vitis vinifera* sugar transporter gene family: phylogenetic overview and macroarray expression profiling. BMC Plant Biology 10:245 (2010)

17. Choat B., Gambetta G.A., Shackel K.A., Matthews M.A. Vascular function in grape berries across development and its relevance to apparent hydraulic isolation. Plant Physiol. 151(3):1677-1687 (2009)
18. Rogiers S., Keller M., Holzapfel B.P., Virgona J.M. Accumulation of potassium and calcium by ripening berries on field vines of *Vitis vinifera* (L) cv. Shiraz. Aust J Grape Wine R. 6(3):240-243 (2000)
19. Deluc L.G., Grimplet J., Wheatley M.D., et al. Transcriptomic and metabololite analyses of Cabernet Sauvignon grape berry development. BMC Genomics 8:429 (2007)
20. Goulao L.F., Oliveira C.M. Cell wall modifications during fruit ripening: when a fruit is not the fruit. Trends Food Sci Technol. 19(1):4-25 (2008)
21. Ling H. Sequence analysis of GDSL lipase gene family in *Arabidopsis thaliana*. Pak J Biol Sci. 11(5):763-767 (2008)
22. Volokita M., Rosilio-Brami T., Rivkin N., Zik M. Combining comparative sequence and genomic data to ascertain phylogenetic relationships and explore the evolution of the large GDSL-lipase family in land-plants. Mol Biol Evol. 28(1):551-565 (2010)
23. Akoh C.C., Lee G.C., Liaw Y.C., Huang T.H., Shaw J.F. GDSL family of serine esterases/lipases. Prog Lipid Res. 43:534-552 (2004)
24. Lee K.A. & Cho T.J. Characterization of a salicylic acid- and pathogen-induced lipase-like gene in chinese cabbage. J Biochem Mol Biol. 36:433-441 (2003)
25. Oh I.S., Park A.R., Bae M.S., et al. Secretome analysis reveals an Arabidopsis lipase involved in defense against *Alternaria brassicicola*. Plant Cell 17:2832-2847 (2005)
26. Hong J.K., Choi H.W., Hwang I.S., Kim D.S., Kim N.H., Choi du S., Kim Y.J., Hwang B.K. Function of a novel GDSL-type pepper lipase gene, CaGLIP1, in disease susceptibility and abiotic stress tolerance. Planta 227:539-558 (2008)
27. Taipalensuu J., Andreasson E., Eriksson S., Rask L. Regulation of the wound-induced myrosinase-associated protein transcript in Brassica napus plants. Eur J Biochem. 247:963-971 (1997)
28. Naranjo M.A., Forment J., Roldan M., Serrano R., Vicente O. Overexpression of *Arabidopsis thaliana LTL1*, a salt-induced gene encoding a GDSL-motif lipase, increases salt tolerance in yeast and transgenic plants. Plant Cell Environ. 29:1890-1900 (2006)
29. Riemann M., Gutjahr C., Korte A., et al. *GER1*, a GDSL motif-encoding gene from rice is a novel early light- and jasmonate-induced gene. Plant Biol. 9:32-40 (2007)
30. Kiba T., Naitou T., Koizumi N., Yamashino T., Sakakibara H., Mizuno T. Combinatorial microarray analysis revealing Arabidopsis genes implicated in cytokinin responses through the His->Asp phosphorelay circuitry. Plant Cell Physiol. 46:339-355 (2005)
31. Cao D., Cheng H., Wu W., Soo H.M., Peng J. Gibberellin mobilizes distinct DELLA-dependent transcriptomes to regulate seed germination and floral development in Arabidopsis. Plant Physiol. 142:509-525 (2006)
32. Dickstein R., Prusty R., Peng T., Ngo W., Smith M.E. *ENOD8*, a novel early nodule-specific gene, is expressed in empty alfalfa nodules. Mol Plant Microbe Interact. 6:715-721 (1993)

33. Mayfield J.A., Fiebig A., Johnstone S.E., Preuss D. Gene families from the *Arabidopsis thaliana* pollen coat proteome. Science 292:2482-2485 (2001)
34. Kondou Y, Nakazawa M, Kawashima M, et al. RETARDED GROWTH OF EMBRYO1, a new basic helix-loop-helix protein, expresses in endosperm to control embryo growth. Plant Physiol. 147:1924-1935 (2008)
35. Updegraff E.P., Zhao F., Preuss D. The extracellular lipase EXL4 is required for efficient hydration of Arabidopsis pollen. Sex Plant Reprod. 22:197-204 (2009)
36. Broun P., Poindexter P., Osborne E., Jiang C.Z., Riechmann J.L. WIN1, a transcriptional activator of epidermal wax accumulation in Arabidopsis. Proc Natl Acad Sci USA. 101:4706-4711 (2004)
37. Kannangara R., Branigan C., Liu Y., et al. The transcription factor WIN1/SHN1 regulates cutin biosynthesis in *Arabidopsis thaliana*. Plant Cell 19:1278-1294 (2007)
38. Reina J.J., Guerrero C., Heredia A. Isolation, characterization, and localization of AgaSGNH cDNA: a new SGNH-motif plant hydrolase specific to *Agave americana* L. leaf epidermis. J Exp Bot. 58:2717-2731 (2007)
39. Cominelli E., Sala T., Calvi D., Gusmaroli G., Tonelli C. Over-expression of the Arabidopsis *AtMYB41* gene alters cell expansion and leaf surface permeability. Plant J. 53:53-64 (2008)
40. Mintz-Oron S., Mandel T., Rogachev I., et al. Gene expression and metabolism in tomato fruit surface tissues. Plant Physiol. 147:823-851 (2008)
41. Panikashvili D., Shi J.X., Bocobza S., Franke R.B., Schreiber L., Aharoni A. The Arabidopsis DSO/ABCG11 transporter affects cutin metabolism in reproductive organs and suberin in roots. Mol Plant. 3:563-575 (2010)
42. Takahashi K., Shimada T., Kondo M., Tamai A., Mori M., Nishimura M., Hara-Nishimura I. Ectopic expression of an esterase, which is a candidate for the unidentified plant cutinase, causes cuticular defects in *Arabidopsis thaliana*. Plant Cell Physiol. 51:123-131 (2010)
43. Sparvoli F., Martin C., Scienza A., Gavazzi G., Tonelli C. Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (*Vitis vinifera* L.). Plant Mol Biol. 24(5):743-55 (1994)
44. Romero-Pérez A.I., Ibern-Gómez M., Lamuela Raventos R.M., De la Torre-Boronat M.C. Piceid, the major resveratrol derivative in grape juices. J Agric Food Chem. 47:1533-1536 (1999)
45. Bavaresco L., Pezzutto S., Gatti M., Mattivi F. Role of the variety and some environmental factors in grape stilbenes. Vitis 46:57-61 (2007)
46. Ooka H., Satoh K., Doi K., et al. Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. DNA Res 10:239-247 (2003)
47. Fang Y., You J., Xie K., Xie W., Xiong L. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. Mol Genet Genomics 280:535-546 (2008)
48. Souer E., van Houwelingen A., Kloos D., Mol J., Koes R. The no apical meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordial boundaries. Cell 85:159-170 (1996)

49. Sablowski R.W. & Meyerowitz E.M. A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA. Cell 92:93-103 (1998)
50. Duval M., Hsieh T., Kim S., Thomas T. Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. Plant Mol Biol. 50:237-248 (2002)
51. Fujita M., Fujita Y., Maruyama K., Seki M., Hiratsu K., Ohme-Takagi M., Tran L.S.P, Yamaguchi-Shinozaki K., Shinozaki K. A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. Plant J. 39:863-876 (2004)
52. Zhong R., Demura T., Ye Z.H. SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis. Plant Cell 18:3158-3170 (2006)
53. Zhu M., Chen G., Zhou S., Tu Y., Wang Y., Dong T., Hu Z. A new tomato NAC (NAM/ATAF1/2/CUC2) transcription factor, SlNAC4, functions as a positive regulator of fruit ripening and carotenoid accumulation. Plant Cell Physiol. 55(1):119-135 (2014)
54. Wang N., Zheng Y., Xin H., Fang L., Li S. Comprehensive analysis of NAC domain transcription factor gene family in *Vitis vinifera*. Plant Cell Rep. 32(1):61-75 (2013)
55. Bowles D.J. Defence-related proteins in higher plants. Annu Rev Biochem. 59:873-907 (1990)
56. Yang C.H., Gavilanes-Ruiz M., Okinaka Y., Vedel R., Berthuy I., Boccara M., Chen J. W., Perna N.T., Keen N.T. Hrp genes of *Erwinia chrysanthemi* 3937 are important virulence factors. Mol. Plant-Microbe. Interact. 15:472-480 (2002)
57. Sgro G.G., Ficarra F.A., Dunge G., Scarpeci T.E., Valle E.M., Cortadi A., Orellano E.G., Gottig, N., Ottado J. Contribution of a harpin protein from *Xanthomonas axonopodis* pv. citri to pathogen virulence. Mol. Plant Pathol. 13:1047-1059 (2012)
58. Taylor N.G., Scheible W.R., Cutler S., Somerville C.R., Turner S.R. The irregular xylem3 locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell. 11(5):769-780 (1999)
59. Taylor N.G., Laurie S., Turner SR. Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. Plant Cell. 12(12):2529-2540 (2000)
60. Thomas T.R., Shackel K.A., Matthews M.A. Mesocarp cell turgor in *Vitis vinifera* L. berries throughout development and its relation to firmness, growth, and the onset of ripening. Planta 228:1067-1076 (2008)
61. Matthews M.A., Thomas T.R., Shackel K.A. Fruit ripening in *Vitis vinifera* L.: possible relation of veraison to turgor and berry softening. Australian Journal of Grape and Wine Research 15:278-283 (2009)
62. Bogs J., Downey M., Harvey J.S., Ashton A.R., Tanner G.J, Robinson S.P. Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves. Plant Physiol. 139:652–63 (2005)
63. Gagné S., Cédric Saucier C., Gény L. Composition and cellular localization of tannins in Cabernet Sauvignon skins during growth. J. Agric. Food Chem. 54(25):9465-9471 (2006)
64. Zhang W., Conn S., Franco C. Characterisation of anthocyanin transport and storage in *Vitis vinifera* L. cv Gamay Fréaux cell suspension cultures. J Biotechnol. 131:196-210 (2007)

65. Staswick P.E., Serban B., Rowe M., Tiryaki I., Maldonado M.T., Maldonado M.C., Suza W. Characterization of an Arabidopsis enzyme family that conjugates amino acids to indole-3-acetic acid. Plant Cell 17(2):616-627 (2005)
66. Östin A., Kowalyczk M., Bhalerao R.P., Sandberg G. Metabolism of indole-3-acetic acid in Arabidopsis. Plant Physiol. 118(1):285-296 (1998)
67. Wang S., Hagen G., Guilfoyle T.J. ARF-Aux/IAA interactions through domain III/IV are not strictly required for auxin-responsive gene expression. Plant Signaling & Behavior 8:6 (2013)
68. Inaba A., Ishida M., Sobajima Y. Changes in endogenous hormone concentrations during berry development in relation to ripening of Delaware grapes. J Japan Soc Hort Sci. 45:245-252 (1976)
69. Cawthon D.L. & Morris J.R. Relationship of seed number and maturity to berry development, fruit maturation, hormonal changes, and uneven ripening of Concord (*Vitis labrusca* L.) grapes. J Am Soc Hort Sci. 107:1097-1104 (1982)
70. Baydar N.G. & Harmankaya N. Changes in endogenous hormone levels during the ripening of grape cultivars having different berry set mechanisms. Turk J Agric For. 29:205-210 (2005)
71. Zhang X.R., Luo G.G., Wang R.H., Wang J., Himelrick D.G. Growth and developmental responses of seeded and seedless grape berries to shoot girdling. J Am Soc Hort Sci. 128:316-323 (2003)
72. Böttcher C., Keyzers R.A., Boss P.K., Davies C. Sequestration of auxin by the indole-3-acetic acid-amido synthetase GH3-1 in grape berry (*Vitis vinifera* L.) and the proposed role of auxin conjugation during ripening. J Exp Bot. 61:3615-3625 (2010)
73. Böttcher C. & Davies C. Hormonal control of grape berry development and ripening. In The Biochemistry of the Grape Berry, vol. 1. Edited by Gerós H., Chaves M.M., Delrot S. Sharjah: Bentham Science pp.194-217 (2012)
74. Chervin C., Tira-Umphon A., Terrier N., Zouine M., Severac D., Roustan J.P. Stimulation of the grape berry expansion by ethylene and effects on related gene transcripts, over the ripening phase. Physiol Plant. 134(3):534-546 (2008)
75. El-Kereamy A., Chervin C., Roustan J.P., Cheynier V., Souquet J.M., Moutounet M., Raynal J., Ford C., Latche A., Pech J.C., Bouzayen M. Exogenous ethylene stimulates the long-term expression of genes related to anthocyanin biosynthesis in grape berries. Physiol Plant. 119:175-182 (2003)
76. Tesniere C., Pradal M., El-Kereamy A., Torregrosa L., Chatelet P., Roustan J.P., Chervin C. Involvement of ethylene signalling in a non-climacteric fruit: new elements regarding the regulation of ADH expression in grapevine. J Exp Bot. 55:2235-2240 (2004)
77. Chervin C., El-Kereamy A., Roustan J.P., Latché A., Lamon J., Bouzayen M. Ethylene seems required for the berry development and ripening in grape, a non-climacteric fruit. Plant Sci. 167(6):1301-1305 (2004)
78. Licausi F., Giorgi F.M., Zenoni S., Osti F., Pezzotti M., Perata P. Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*. BMC Genomics 11:719 (2010)
79. Berrocal-Lobo M., Molina A., Solano R. Constitutive expression of ETHYLENE-RESPONSE-FACTOR 1 in Arabidopsis confers resistance to several necrotrophic fungi. Plant J. 29:23-32 (2002)

80. Gu Y.Q., Wildermuth M.C., Chakravarthy S., Loh Y.T., Yang C., He X., Han Y., Martin G.B. Tomato transcription factor Pti4, Pti5, and Pti6 activate defence responses when expressed in Arabidopsis. Plant Cell 14:817-831 (2002)
81. Desprez T., Juraniec M., Crowell E.F., Jouy H., Pochylova Z., Parcy F., Höfte H., Gonneau M., Vernhettes S. Organization of cellulose synthase complexes involved in primary cell wall synthesis in Arabidopsis thaliana. Proc Natl Acad Sci. USA. 104:15572-15577 (2007)
82. Dal Santo S., Vannozzi A., Tornielli G.B., Fasoli M., Venturini L., Pezzotti M., Zenoni S. Genome-wide analysis of the expansin gene superfamily reveals grapevine-specific structural and functional characteristics. PLoS One 16;8(4):e62206 (2013)
83. Capel J., Jarillo J.A., Salinas J., Martinez-Zapater J.M. Two homologous low-temperature-inducible genes from Arabidopsis encode highly hydrophobic proteins. Plant Physiol. 115:569-576 (1997)
84. Medina J., Catalá R., Salinas J. Developmental and stress regulation of RCI2A and RCI2B, two cold-inducible genes of arabidopsis encoding highly conserved hydrophobic proteins. Plant Physiol. 125(4):1655-1666 (2001)
85. Larson R.L., Coe E.H. Gene-dependent flavonoid glucosyltransferase in maize. Biochem Genet. 15:153-156 (1977)
86. Boss P.K., Davies C., Robinson S.P. Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L. cv. Shiraz grape berries and the implications for pathway regulation. Plant Physiol. 111:1059-1066 (1996)
87. Mueller L.A., Goodman C.D., Silady R.A., Walbot V. AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. Plant Physiol. 123:1561-1570 (2000)
88. Gomez C., Conejero G., Torregrosa L., Cheynier V., Terrier N., Ageorges A. In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST. Plant J. 67(6):960-970 (2011)
89. Kobayashi S., Ishimaru M., Hiraoka K., Honda C. Myb-related genes of the Kyoho grape (Vitis labruscana) regulate anthocyanin biosynthesis. Planta 215:924-933 (2002)
90. Kobayashi S., Yamamoto N.G., Hirochika H. Association of VvmybA1 gene expression with anthocyanin production in grape (Vitis vinifera) skin - color mutants. J Japan Soc Hort Sci. 74:196-203 (2005)
91. Matsubayashi Y., Ogawa M., Morita M., Sakagami Y. Phytosulfokine, sulfated peptides that induce the proliferation of single mesophyll cells of Asparagus officinalis L. Proc. Natl Acad. Sci. USA. 93:7623-7627 (1997)
92. Yang H., Matsubayashi Y., Hagani H., Sakagami Y. Phytosulfokine-a, a peptide growth factor found in higher plants: its structure, functions and receptors. Plant Cell Physiol. 41:825-830 (2000)
93. Yang H., Matsubayashi Y., Hagani H., Nakamura K., Sakagami Y. Diversity of Arabidopsis genes encoding precursors for phytosulfokine, a peptide growth factor. Plant Physiol. 127:842-851 (2001)
94. Stührwohldt N., Dahlke R.I., Steffens B., Johnson A., Sauter M. Phytosulfokine-α controls hypocotyl length and cell expansion in Arabidopsis thaliana through phytosulfokine receptor 1. PLoS ONE 6:e21054 (2011)

95. Matsubayashi Y., Ogawa M., Kihara H., Niwa M., Sakagami Y. Disruption and overexpression of Arabidopsis phytosulfokine receptor gene affects cellular longevity and potential for growth. Plant Physiol. 142:45-53 (2006)
96. Kutschmar A., Rzewuski G., Stuhrwohldt N., Beemster G.T.S., Inze D., Sauter M. PSK-α promotes root growth in Arabidopsis. New Phytol. 181:820-831 (2009)
97. Igarashi D., Tsuda K., Katagiri F. The peptide growth factor, phytosulfokine, attenuates pattern-triggered immunity. Plant J. 71(2):194-204 (2012)

Chapter 7

Grape berry development transcriptomic route

INTRODUCTION

Multicellular organism development depends of global transcriptional program, i.e. the expression of different sets of genes throughout development. Since two decades, transcriptomic analyses have been developed in order to identify these sets of genes and to follow their expression profiles throughout tissue development. Even if the last transcriptomic technology, the RNA-Sequencing, has already been used to study fruit development of many crops[1-5] including grape berry[6-9], no experiment was carried out on different *Vitis vinifera* varieties. In fact, the comparison of transcriptomes of grape berry from different varieties could provide both common and specific transcriptomic traits throughout fruit development. To represent common transcriptomic traits, we introduced the new term “transcriptomic route” corresponding to the genes having a similar expression pattern during development. On the other hand, specific gene expression and modulation might be the source of information to define for a molecular point of view the well-known traits that make the ten varieties so different for each other.

Clusterization is a statistical approach suitable to subdivide a set of items (in our case, transcripts) in such a way that similar items fall into the same cluster, whereas dissimilar items fall in different clusters. The two most important classes of clustering methods are hierarchical clustering and partitioning. In hierarchical clustering, each cluster is subdivided into smaller clusters, forming a tree-shaped data structure or dendrogram. K-means clustering, a partitioning method, subdivides the genes into a predetermined number (k) of clusters[10]. Nevertheless, the performance of cluster analyses depends upon selecting a good distance function over input data set - clusters are determined based on distances between items, thus selecting the distance between pairs of objects to be employed by the clustering method is at least as important as selecting the clustering method itself[11]. One of the most commonly used distance measure for gene expression data is the Euclidean distance, which measures the geometric distance between two items.

In this work, in order to establish berry development transcriptomic route, a novel coupled clustering analysis (a hierarchical clustering confirmed by a k-mean clustering) was carried out on 4,613 commonly modulated genes and a selection of constitutive genes was performed by classifying transcripts in function of their coefficient of variation among the 40-berry samples.

RESULTS AND DISCUSSION

Selection of the commonly up- and down-regulated genes

After identifying the differentially expressed genes (DEGs) with a pairwise analysis of subsequent developmental stages (i.e. Pea-sized berry-Prevéraison, Prevéraison-End of véraison, End of véraison-Harvest) as described in **chapters 4 and 6**, we selected the 4,613 genes that were commonly up- and down-regulated in at least one comparison in all the ten varieties and exhibiting an expression intensity above 1 FPKM in all varieties (**Fig. 1**).

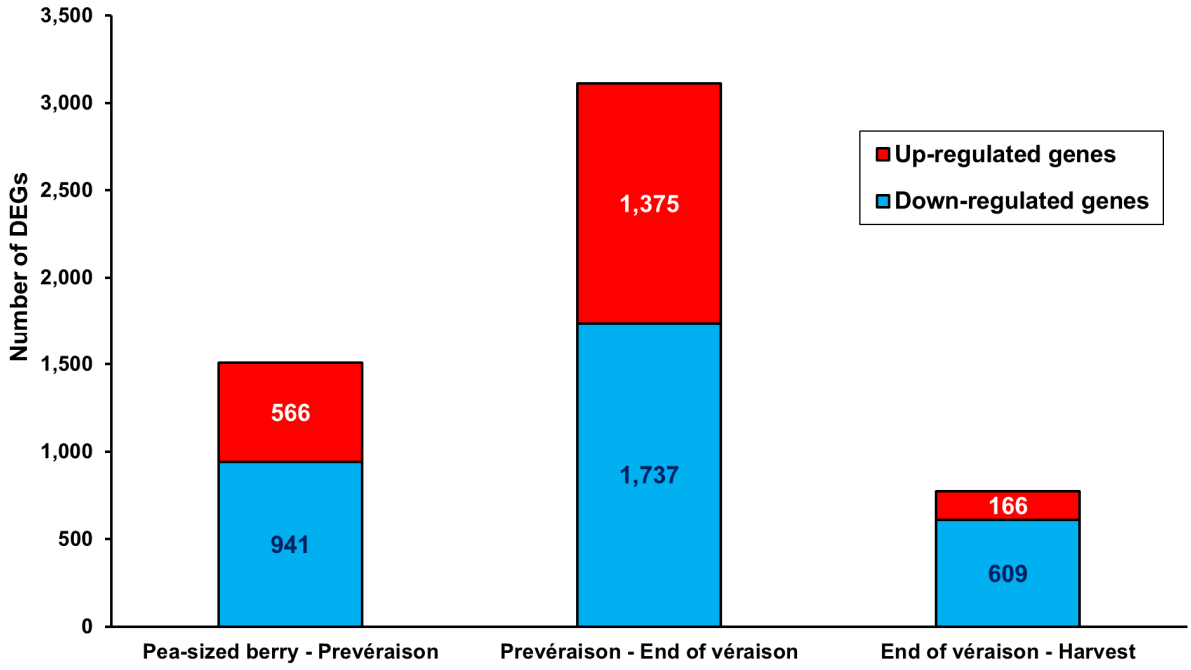


Figure 1: Commonly DEGs between each consecutive pair of developmental stages. The histogram represents the number of commonly down- (blue) and up-regulated genes (red) in all the 10-variety berries between each subsequent pair of developmental stages. Only transcripts showing a gene expression value above 1 FPKM at least at one growth stage in all the 10-variety berries were selected.

In **Figure 1**, we can see that the most intense transcriptional modulation occurred at véraison. We can also observe that more genes were down- rather than up-regulated during berry development, which confirms previous results of Palumbo and colleagues[12] (**chapter 11**), suggesting that the transition to mature growth seemed to predominantly involve the suppression of vegetative pathways rather than the activation of processes typical of maturity.

Berry development transcriptomic route

Clustering analysis

In order to establish the grape berry development transcriptomic route, corresponding to the genes having a similar pattern of expression independently on the variety, a coupled clustering analysis was performed. This analysis allowed identifying 913 genes having the same expression pattern among the 4,613 commonly modulated and grouping them into six clusters containing 542, 94, 120, 20, 134, 3 transcripts, respectively (**Fig. 2**). In **Figure 2**, the first three clusters contain genes showing a higher expression intensity at either one early stage or during the entire green phase, whereas the last three clusters are composed of transcripts exhibiting a peak of expression at either one ripening stage or throughout maturation phase.

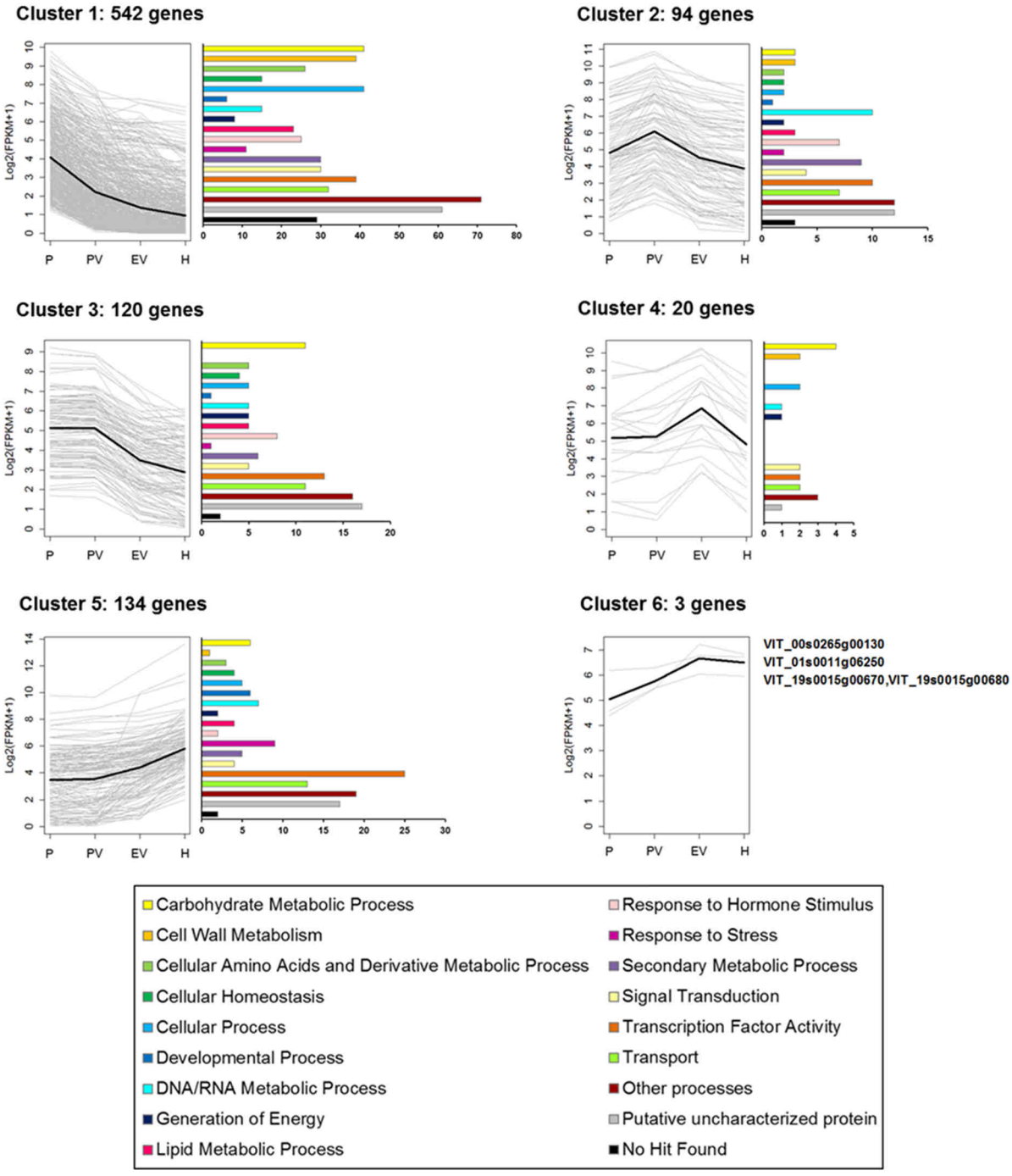


Figure 2: Gene expression profiles and Gene Ontology (GO) categories distribution of the six expression clusters. Clusters were obtained by a coupled clustering analysis (hierarchical clustering followed by a k-mean clustering analysis as confirmation) using the 4,613 commonly modulated genes. Each single line represents the log_2-transformed average of the mean FPKM values for an individual transcript. Histogram representation of the GO categories distribution is expressed as the amount of genes belonging to the cluster. P, Pea-sized berry; PV, Prevéraison; EV, End of véraison; H, Harvest.

Cluster 1

Cluster 1 contained 542 transcripts showing a decreasing trend throughout berry development. The genes were analysed for overrepresented functions using the GO enrichment tool BiNGO[13] (**Fig. 3**) and four GO categories were found significantly overrepresented: cellular component organization (GO:0016043), secondary metabolic process (GO:0019748), carbohydrate metabolic process (GO:0005975) and cell cycle (GO:0007049).

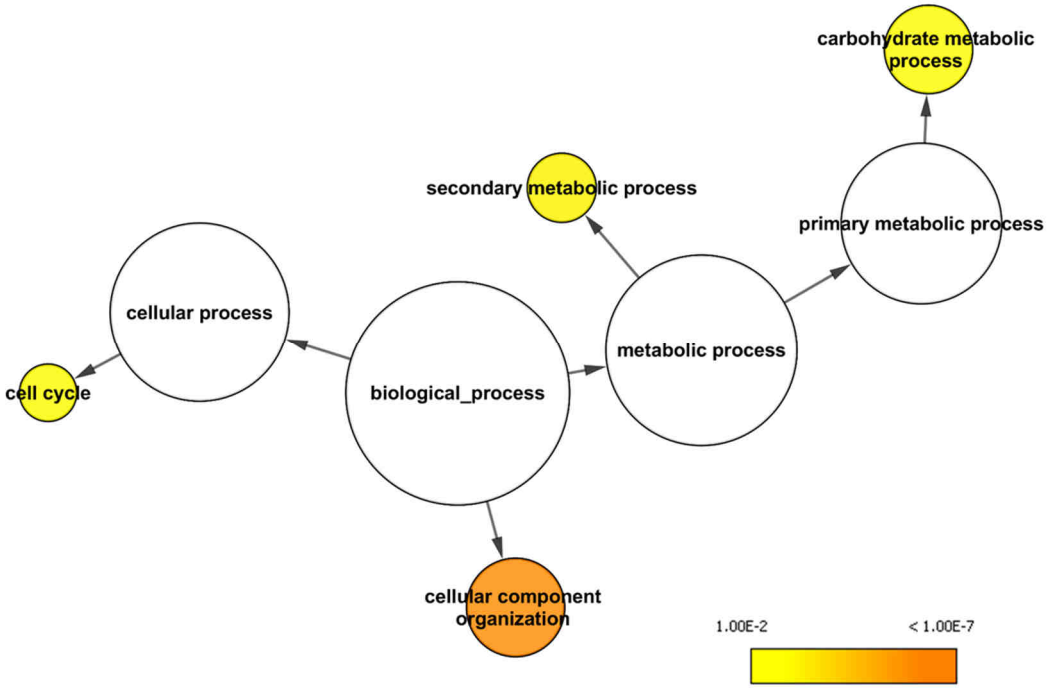


Figure 3: Enriched GO terms in cluster 1 transcripts.

The network graph shows BiNGO visualization of the overrepresented GO terms for the 542 transcripts composing cluster 1. Node size is positively correlated with the number of genes belonging to the category. Non-coloured nodes are not overrepresented, but they may be the parents of overrepresented terms. Coloured nodes represent GO terms that are significantly overrepresented (Benjamini and Hochberg corrected p-value < 0.01), with the shade indicating significance as shown in the colour bar.

Among the cellular component organization category, we identified genes coding for a tubulin α (VIT_14s0108g00440) and three tubulin β chain proteins (VIT_02s0025g03680; VIT_06s0004g05870; VIT_08s0040g00980) involved in the formation of microtubules, which undergo dynamic reorganization to form arrays of cortical microtubules, the pre-prophase band and the citokinetic phragmoplast during plant cell division[14]. In addition, twenty-five genes related to cell wall metabolism were identified as two cellulose synthase-like A genes (VIT_04s0069g00780; VIT_08s0007g03990) predicted to encode β-mannan synthases[15], three expansin genes: *VvEXPA6*

(VIT_06s0004g04860), *VvEXPA16* (VIT_14s0108g01020) and *VvEXPB2* (VIT_12s0059g00190), and five pectin methylesterase genes (VIT_07s0005g00730; VIT_11s0016g00290; VIT_11s0016g00330; VIT_11s0016g03020; VIT_14s0060g01960) involved in cell wall loosening[16]. Furthermore, genes related to cell wall polysaccharide synthesis were found among the enriched carbohydrate metabolic process category as a cellulose synthase A3 gene (VIT_08s0007g08380) required for the synthesis of primary cell wall[17], the cellulose synthase A4, A7 and A8 genes (VIT_07s0005g04110; VIT_11s0037g00530; VIT_10s0003g01560 respectively) involved in secondary cell wall metabolism in *Arabidopsis thaliana*[18,19] and three xyloglucan endotransglucosylase/hydrolase genes (VIT_05s0062g00240; VIT_10s0003g02440; VIT_12s0134g00160). It has been postulated that xyloglucan endotransglucosylase/hydrolases carry out various functions[20]; these include wall loosening[21], wall strengthening[22], integrating new xyloglucans into the wall[23], and hydrolysing xyloglucans[24,25] particularly during xylem formation[26,27]. The peak of expression of these cell wall-related genes at pea-sized berry stage coincides with the cytokinesis phenomenon[28] occurring during cell division and the formation of xylem allowing a symplasmic unloading pathway during this early stage[29]. In particular, Dal Santo and colleagues[30] found that *VvEXPA6*, *VvEXPA16* and *VvEXPB2* were expressed mainly in grapevine vegetative organs supporting the hypothesis that this cluster included only genes related to green development. In this case, expansin genes might be related to the cell expansion and primary cell wall deposition, helping cellulose network to be built. Moreover, *VvEXPA6* was found expressed in root tissue, while *VvEXPA16* and *VvEXPB2* were expressed in swelling bud, suggesting their involvement in secondary cell wall/xylem formation.

Another overrepresented GO category was the cell cycle function including genes coding for two types of cyclin gene necessary to the plant cell-cycle machinery[31]: a cyclin B gene (VIT_08s0040g00930) and the corresponding cyclin-dependent kinase (VIT_18s0122g00550) and five cyclin D genes (VIT_00s0199g00210; VIT_18s0001g07220; VIT_07s0129g01100; VIT_18s0001g09920; VIT_00s0203g00160) supporting the fact that pericarp cells were undergoing cell division during pea-sized berry stage.

Among the enriched secondary metabolic process GO category, genes involved in lignin biosynthesis were identified, as a caffeic acid methyltransferase gene (VIT_18s0001g02610), cinnamyl alcohol dehydrogenase gene[32] (VIT_00s0615g00020) and five laccase genes (VIT_08s0007g01910; VIT_06s0004g04050; VIT_13s0067g01970; VIT_08s0007g06460; VIT_04s0023g01960). Laccases, or p-diphenol:O2 oxidoreductases,

are copper-containing glycoproteins found in a wide range of living organisms including bacteria, fungi, insects, and plants. In a recent study, a laccase was found as necessary for lignin polymerization during vascular development in *Arabidopsis thaliana*[33]. Thus, we can hypothesize that these laccase genes could be involved in xylem formation at this early growth stage.

Furthermore, among the carbohydrate metabolic process category, we found two sucrose synthase genes *VvSUS3* and *VvSUS5*[34] (VIT_17s0053g00700 and VIT_04s0079g00230 respectively) and a tonoplast monosaccharide transporter *VvTMT3*[35] (VIT_07s0031g02270). During berry development, sucrose synthases and invertases are mainly involved in the hydrolysis of the sucrose – produced by photosynthesis in the leaves – into glucose and fructose[36]. Until véraison, most of the sugar imported into the berry is metabolised to organic acids and almost no storage occurs. In tomato, sucrose synthase activity has been showed to control the unloading kinetics of the very young fruit determining then the growth rate[37]. Consequently, the down-regulation of these two sucrose genes during berry development suggests that, among the sucrose synthase gene family, these two genes are green phase-specific and might be crucial for berry formation.

Finally, 225 positive biomarker transcripts of pea-sized berry stage (out of the 326 discussed in **chapter 5**) were identified among cluster 1 (**Supplementary table 1**) showing that the most part of the putative pea-sized berry stage-specific positive biomarker transcripts could be identified by a clustering analysis as they have the same profile during berry development.

Cluster 2

Cluster 2 was composed of 94 transcripts as two cellulose synthase-like genes (VIT_00s0975g00020; VIT_00s1349g00010), a cellulose synthase A3 gene and an invertase/pectin methylesterase inhibitor gene (VIT_00s0323g00060) shown to regulate the activity of pectin methylesterase at post-translational level by binding to this protein[38].

Among cluster 2, we observed also a high proportion of genes involved in DNA/RNA metabolic process and transcription factor activity (**Fig. 2**). In fact, six transcripts coding for ribosomal proteins were identified: S26 (VIT_18s0001g06290), L13A (VIT_05s0029g00020), L21 (VIT_08s0007g06890) and two L26 isoforms (VIT_07s0129g00160;

VIT_07s0129g00090) (S for small-unit and L for large unit) and ten transcription factor genes.

Furthermore, thirteen putative prevéraison stage-specific positive biomarker transcripts were found among cluster 2 (**Table 1**) as an elongation factor 1-alpha gene (VIT_08s0040g02330) which encoded protein catalyses the first step of the elongation cycle during peptide synthesis, interacting with aminoacyl-tRNA to bring it to the acceptor site of the ribosome[39].

Table 1: List of the thirteen putative prevéraison stage-specific positive biomarker transcripts having a similar expression profile during development. For each transcript, a mean FPKM value of the 10-variety samples was calculated for each growth stage. Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

Gene_ID	Functional annotation	P	PV	EV	H
VIT_02s0087g00720	Receptor protein kinase	3.21	35.40	3.05	1.11
VIT_04s0023g03880	Unknown	106.33	278.99	107.03	64.77
VIT_05s0020g02490	Unknown protein	38.78	62.76	38.53	37.84
VIT_06s0004g01390	S-receptor kinase	13.42	45.89	12.89	5.59
VIT_06s0004g06650	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase	1.77	4.38	1.11	1.00
VIT_07s0191g00180	Homeobox-leucine zipper protein ATHB-6	142.34	323.47	131.04	94.77
VIT_08s0040g02330	Elongation factor 1-alpha	434.52	928.30	343.55	251.53
VIT_08s0040g02360	Calmodulin binding protein	81.07	254.58	89.55	69.97
VIT_14s0068g01690	TCP family transcription factor TCP20	23.71	42.40	20.10	18.58
VIT_15s0107g00270	Mechanosensitive ion channel	8.24	28.04	1.88	0.70
VIT_18s0001g05800	Dehydration-responsive protein	141.82	312.83	172.43	123.44
VIT_18s0117g00260	Co-chaperone-curved DNA binding protein A	10.85	26.07	5.70	3.91
VIT_19s0014g01250	TUBBY like protein 7 TLP7	69.54	120.96	58.64	52.10

Cluster 3

Among the 120 transcripts composing cluster 3, the photosynthesis process (GO:0015979) was identified as significantly overrepresented GO category (p-value = 1.2079E-5), using the enrichment tool BiNGO[13]. This GO category included photosystem I light harvesting complex gene 3 and 5 (VIT_15s0024g00040 and VIT_18s0001g10550 respectively), a photosystem II oxygen-evolving complex precursor gene (VIT_12s0028g01080) and a phosphoenolpyruvate carboxylase gene (VIT_01s0011g02740). The photosynthetic activity of the berry has been shut down from either pea-sized berry stage (cluster 1) or at the end of the green phase, as the decreasing trend of photosynthesis-related genes pointed out in cluster 3. Moreover, phosphoenolpyruvate carboxylase gene decrease of expression was coherent with the transition to the maturation phase concerning the end in accumulation of malate, which PEPC was suggested to be necessary to[40].

Genes involved in auxin transport and gene regulation were also found, as two auxin efflux carrier genes (VIT_04s0044g01850; VIT_04s0044g01870), a auxin/indole acetic acid protein (Aux/IAA) gene (VIT_18s0001g08090) and four auxin response factor genes[41] (VIT_02s0025g01740; VIT_06s0004g03130; VIT_12s0035g01800; VIT_18s0089g00910). Indole acetic acid levels are high early in berry development – in agreement with the proposed role for auxin in cell division and expansion – after which they decline steadily to be very low at véraison[42,43]. The down-regulation of these auxin transport/signalling/sensing-related genes after véraison seems to be important for initiating grape berry maturation and confirms the hypothetic role of auxins as ripening inhibitors[43].

Interestingly, one hundred transcripts of cluster 3 were classified among the putative positive green phase-specific biomarker transcripts (**chapter 5**), confirming the robustness of both analyses (**Supplementary table 2**).

Cluster 4

Cluster 4 was composed of only twenty transcripts (**Table 2**), including the two sugar transporter genes: *VvHT6/TMT2* (VIT_18s0122g00850) and a SWEET gene (VIT_17s0000g00830). The hexose transporter *VvHT6/TMT2* transcripts are highly accumulated at véraison[44,45], suggesting that this tonoplastic transporter may be responsible for vacuolar accumulation of hexose at the inception of ripening[46]. *SWEET* genes codify a new class of sugar transporters found in many species and mediating glucose efflux across the cell membrane[47]. Sixteen homologues of *AtSWEET1* were preliminary identified in grapevine and the SWEET gene (VIT_17s0000g00830) was indeed represented among those three genes significantly expressed over berry development[46]. An aquaporin TIP1-3 (VIT_13s0019g00330) was also comprised in cluster 4 suggesting that it might facilitate the transport of water and/or small neutral solutes across tonoplast[48] during the maturation phase as berry enlarges. With regard to cell expansion and wall modification, two cell wall loosening-related genes were found: a pectin methyl-esterase PME1 gene[16] (VIT_09s0002g00330) and an expansin *VvEXPA18* gene[30,49] (VIT_17s0053g00990). The peak of expression of these transcripts at the end of véraison (**Fig. 2**) could be easily associated with the physiological processes occurring at this stage as sugar accumulation coupled with water inflow, cell expansion and berry softening.

Table 2: Cluster 4 gene list.

Gene_ID	Functional annotation
VIT_00s0220g00070	Latex cyanogenic beta glucosidase
VIT_03s0063g01870	Cold acclimation protein WCOR413 protein beta form
VIT_06s0004g05810	TPR-containing protein kinase
VIT_06s0009g02500	Plastocyanin domain-containing protein
VIT_09s0002g00330	Pectinesterase PME1
VIT_12s0034g01160	Blue (type 1) copper domain
VIT_12s0055g00910	Kelch repeat-containing F-box protein
VIT_13s0019g00330	Aquaporin TIP1-3
VIT_13s0064g00460	Unknown protein
VIT_13s0064g01260	DNA-damage-repair/toleration protein (DRT100)
VIT_13s0067g00120, VIT_13s0067g00140	GCN5 N-acetyltransferase (GNAT)
VIT_13s0067g01980	Proteinase inhibitor (LUTI)
VIT_14s0060g00280	RARE-cold-inducible 2B
VIT_14s0060g00760	Galactinol synthase
VIT_15s0048g02870	Homeobox-leucine zipper protein HB-7
VIT_17s0000g00830	SWEET transporter
VIT_17s0000g05570	Receptor protein kinase
VIT_17s0053g00990	VvEXPA18
VIT_18s0001g05300	Trehalose-6-phosphate phosphatase
VIT_18s0122g00850	VvHT6/TMT1 Tonoplast monosaccharide transporter2

Cluster 5

Cluster 5 contained 134 transcripts that were characterized by an increasing expression profile from the end of véraison to full-ripening stage. Among these genes, it is worth to notice the presence of the expansin gene *VvEXPB4* (VIT_15s0021g02700), likely involved in cell wall expansion and modification typical of the maturation phase of berry development. Interestingly, Dal Santo and colleagues[30] found that *VvEXPB4* was expressed in both flesh and skin of developing Corvina berries throughout ripening, starting from véraison. Moreover this group comprises the germacrene D synthase gene *VvTPS07*[50] (VIT_18s0001g04280) that was described to be involved in sesquiterpene production at ripening in Gewürztraminer grapes by Martin *et al.*[51]. A wax synthase transcript (VIT_19s0090g01420, VIT_19s0090g01430) was also found. Aerial organs of higher plants are covered by the cuticle at the interface between the plant tissue and the environment. In grape berry, the cuticular membrane shield the underlying berry tissues from desiccation, infection by pathogenic bacteria and fungi, insect attack, injuries due to wind, physical abrasion, frost and radiation[52]. The cuticle is mainly formed by a structural component called cutin, an amorphous biopolyester of hydroxy-fatty acids. Associated with this biopolymer are waxes or soluble cuticular lipids[53]. Waxes can be embedded within the cutin – intracuticular waxes – or deposited

on the outer surface of the plant cuticle – epicuticular waxes. Comménil and co-workers[54] observed that the content of cutine per unit surface decreases more than 2.5-fold between berry fruit set and véraison – causing a decrease of cuticle thickness – and then stays stationary during ripening. In contrast, a measurable increase of wax weight per unit surface was noticed between pea-sized berry stage and harvest – from about 1.2 to 2.2 µg.mm^{-2}. Moreover, they observed that the epicuticular wax layer appeared strongly contrasted, continuous, and thicker in full-ripe berries than in young grapes in relation with the increase of wax content described before. In fact, study of the grape berry surface using electron microscopy showed that wax grains fused and the waxes appeared in the form of platelets densely distributed over the surface of the cuticle, increasing in size and complexity during berry development[55]. Therefore, the wax synthase gene, identified as highly expressed during ripening phase, could be related to the epicuticular wax layer formation. Consequently, the expression profiles (**Fig. 2**) of the above-described transcripts well correlated with the associated physiological process and their role in berry maturation.

In addition, the transcription (GO:0006350) was identified as a significantly overrepresented GO category (p-value = 1.4521E-4), including twenty-five transcription factor genes, using the enrichment tool BiNGO[13] and the GoSlim Plants annotation (**Table 3**). Among them, we found *VvNAC03*, *VvNAC11* and *VvNAC37*, which were identified as statistically up-regulated in all ten varieties between the end of véraison and harvest in **chapter 6**, confirming their belonging to cluster 5. By similarity with their orthologues in *Arabidopsis thaliana*[56], *VvNAC03* and *VvNAC11* might be involved in multicellular organismal development, while *VvNAC37* might be also related to the regulation of secondary cell wall biogenesis, suggesting the formation of secondary cell wall in full-ripe berries likely related to the cell aging. Moreover, we identified two WRKY (WRKYGQK DNA binding domain sequence) genes: *VvWRKY6* and *VvWRKY32*[57]. In *Arabidopsis thaliana*, *AtWRKY6* was described to have a role as a transcription factor functioning in several cell death-related processes, including leaf senescence, floral organ abscission, pathogen defence, and wound response[58]. Hence, we can suppose that *VvWRKY6* could be involved in the loss of cell vitality characteristic of berry maturity[59]. Consequently, we can hypothesize that the peak of expression of these twenty-five transcription factor genes at harvest (**Fig. 2**) and the expression of their target genes might be necessary to complete berry maturation.

Table 3: List of the 25 transcription factor transcripts included in cluster 5.

Gene_ID	Functional annotation
VIT_18s0001g10300	Basic helix-loop-helix (bHLH) family
VIT_01s0011g03720	BEE1 (BR Enhanced expression 1)
VIT_12s0055g00420	BZIP transcription factor
VIT_14s0030g01600	DNA-binding storekeeper protein
VIT_08s0040g00610	GTB1 (global transcription factor group B1)
VIT_01s0011g04870	Homeobox-leucine zipper protein 17 (HB-17)
VIT_15s0048g00830	LOB domain-containing 18 (Asymmetric leaves 2-like protein 20)
VIT_19s0090g00590	Myb domain protein 101
VIT_14s0066g01090	Myb domain protein 24
VIT_16s0039g01900	Myb KAN2 (KANADI 2)
VIT_03s0038g00610	PHD finger transcription factor
VIT_08s0007g07440	PUMILIO 12 (APUM12)
VIT_06s0004g05210	Transcription initiation factor TFIIA large subunit
VIT_00s0375g00040	VvNAC03_NAC domain-containing protein
VIT_14s0108g01070	VvNAC11_NAC domain-containing protein
VIT_10s0003g00350	VvNAC37_NAC domain-containing protein
VIT_11s0037g00150	WRKY DNA-binding protein 32
VIT_12s0059g00880	WRKY DNA-binding protein 6
VIT_13s0047g01130	Zfwd2 protein (ZFWD2)
VIT_08s0007g04770	Zinc finger (C2H2 type) family
VIT_06s0061g00760	Zinc finger (C2H2 type) family
VIT_08s0040g02160	Zinc finger (C3HC4-type ring finger)
VIT_14s0060g02280	Zinc finger (C3HC4-type ring finger)
VIT_08s0032g00920	Zinc finger (C3HC4-type ring finger) CIC7E11
VIT_01s0026g00850	Zinc finger protein 5

Furthermore, twelve of cluster 5 transcripts were identified as putative harvest stage-specific positive biomarker transcripts (**chapter 5**) showing that the most part of these biomarker transcripts (12/17) follow a similar expression pattern during development (**Table 4**).

Table 4: List of the twelve putative harvest stage-specific positive biomarker transcripts having a similar expression profile during development. For each transcript, a mean FPKM value of the 10-variety samples was calculated for each growth stage. Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

Gene_ID	Functional annotation	P	PV	EV	H
VIT_00s1916g00010	DNA binding	37.0	34.9	36.9	99.2
VIT_01s0011g05810	Adagio protein 1	2.3	4.2	10.2	22.7
VIT_02s0109g00250	4-coumarate-CoA ligase	0.4	0.4	0.8	3.0
VIT_05s0020g03870	MLK/Raf-related protein kinase 1	1.6	2.9	12.9	32.0
VIT_05s0020g04080	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 2	343.3	441.2	910.1	1,848.2
VIT_05s0102g00250	Unknown protein	2.3	2.9	3.3	9.8
VIT_08s0217g00030	Ran-binding protein 1 RanBP1	99.0	103.0	162.1	286.4
VIT_11s0118g00560	Nodulin	5.3	6.9	19.6	56.3
VIT_12s0057g00310	Gamete expressed1 (GEX1)	0.2	0.3	8.7	63.5
VIT_16s0098g01230	Cytochrome b5	184.6	220.4	226.4	385.9
VIT_18s0001g10480	Unknown protein	1.1	2.6	10.3	27.4
VIT_18s0001g11910	1-acyl-sn-glycerol-3-phosphate acyltransferase 4	0.4	0.2	1.0	4.3

Cluster 6

Cluster 6 contained only three genes coding for an uncharacterized protein, a xyloglucan endo-transglycosylase responsible for cell wall-loosening[21] and a predicted E3 ubiquitin-protein ligase HUWE1. E3 ubiquitin-protein ligases HUWE1 are animal HECT ubiquitin-protein ligase known to mediate ubiquitination and subsequent proteasomal degradation of target proteins[60]. Nevertheless, a recent study showed that E3 ubiquitin-protein ligases HUWE1 are phylogenetically closed to plant subfamily V HETC proteins, having both UBA domains[61]. The similarity of sequence and the absence of plant Subfamily V HETC proteins in *Arabidopsis thaliana*[61] could explain this prediction issue.

Constitutive genes throughout berry development

Finally, in order to identify genes having a constant expression level throughout berry development in all the ten-variety berries, transcripts were classified by their coefficient of variation among the 40 samples (**Supplemental table 3**).

Among the 10 first top-coefficient-of-variation-scored genes, we found transcripts encoding the ubiquitin-specific protease 12, the Actin 7 (ACT7) / actin 2, the beta 1 subunit of the adaptor-related protein complex 2 and a phosphatidylinositol 4-kinase with an expression intensity of approximatively 58, 1,295, 86, and 14 FPKM respectively (**Fig. 4**). Hence, their different expression levels could be useful to produce a large range of reference genes for real-time qPCR. Furthermore, actin transcript was already described as one of the most stable reference gene for real-time RT-PCR by Reid *et al.*[62]. With our results, we can confirm it but we can also claim that actin gene could be used as reference gene to compare gene expression between pericarps from different varieties.

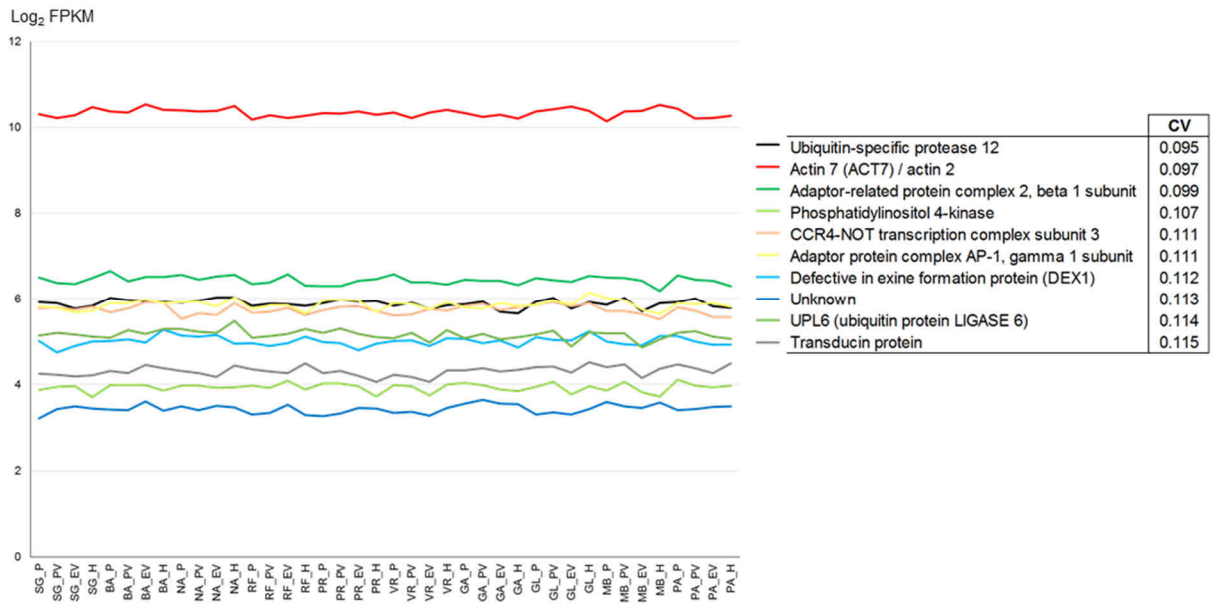


Figure 4: Expression profiles of the 10 first top-coefficient-of-variation-scored genes in the 40-sample FPKM dataset. Coefficient of variation (CV) was calculated for each transcript by dividing the standard deviation of the 40-sample FPKM values by the mean of these values. Each single line represents the log_2-transformed mean FPKM values for an individual transcript.
Abbreviations for each variety: SG (Sangiovese), BA (Barbera), NA (Negro amaro), RF (Refosco), PR (Primitivo), VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina) and for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

Moreover, the comparison of the selected 80 constitutive genes – with a coefficient of variation below than 0.1 among the 40-sample data set (**Supplemental table 3**) – with the 76 non-plastic and constitutive transcripts of Corvina berries from véraison to full maturity[63] – transcripts presenting neither variation among the 11 vineyards nor among the three developmental stages – showed that only two transcripts were common to the two experiments. The two genes: VIT_00s0174g00150 and VIT_04s0008g02930 encode the ubiquitin-specific protease 5, which is a deubiquitinating enzyme (DUB), and an oligouridylate binding protein 1B (UBP1B) respectively. DUBs are proteases that reverse the modification of proteins by ubiquitin. DUBs are divided into two general groups, ubiquitin C-terminal hydrolases and ubiquitin-specific proteases based on their amino acid sequence and substrate specificity[64,65]. DUBs act in the release of ubiquitin moieties from the polyubiquitin chain[66], which is essential for the 26S proteasome recognition, and leads to the subsequent degradation of the polyubiquitinated substrate[67]. On the other hand,

UBP1s are an heterogeneous nuclear RNA-binding-like protein associated with the poly(A)$^+$ RNA in the cell nucleus[68] that enhance intron recognition, splicing, and mRNA accumulation, and assemble into heat-triggered stress granules in transiently transfected tobacco protoplasts[68-71]. Consequently, these two housekeeping genes seem to be good candidates as reference genes for quantitative gene expression analysis on berry pericarps. Nevertheless, further experiments including pre-véraison growth stages and several cultivars – red and white grapes – are necessary to confirm this.

Finally, the selected 80 constitutive genes and the 80 first top-coefficient-of-variation-scored genes among the 54 samples of the grape gene expression atlas[72] – from 0.0136 until 0.0206 – share only one gene (VIT_02s0012g00910) coding for a µ2 subunit of the adaptor-related protein complex 2 – homologous to the medium subunit 2 of the mammalian ADAPTOR PROTEIN COMPLEX 2 (AP2)[73,74] – involved in the clathrin-mediated endocytosis in *Arabidopsis thaliana*[73-78]. Remarkably, this gene showed constitutive expression also in the 11-vineyard data set[63] and in the transcriptome of all grapevine organs[72] (**Fig. 5**). Therefore, this result shows that this gene could be the best candidate as reference gene for quantitative gene expression analysis carried out on all grapevine organs independently of the environment.

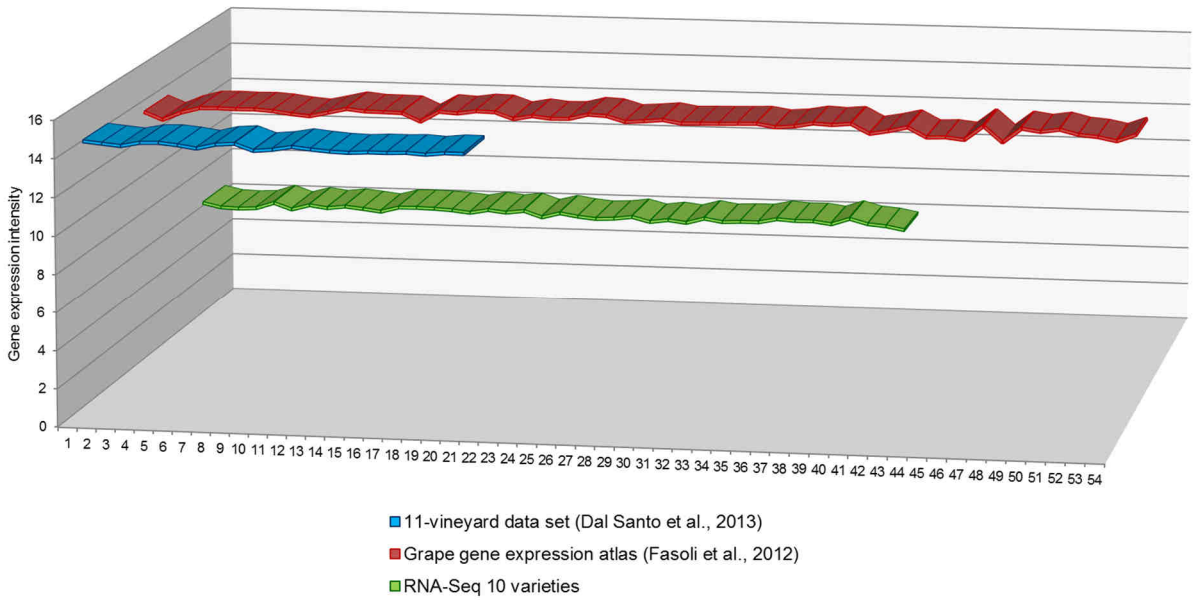


Figure 5: Expression profiles of µ2 subunit of the adaptor-related protein complex 2 gene (VIT_02s0012g00910) in the 11-vineyard data set (21 samples)[63], in the whole grapevine expression atlas (54 samples)[72] in $\log_2$-fluorescence intensity and in this RNA-Seq assay (40 samples) in $\log_2$-FPKM.

CONCLUSION

The elaboration of a novel coupled clustering analysis allowed to group 913 differentially expressed transcripts into six clusters. The first three clusters were composed of 542, 120 and 94 transcripts respectively having a higher expression intensity during green phase, whereas the three last clusters containing 134, 3, 20 transcripts respectively exhibiting a peak of expression during the maturation phase. In fact, the comparison of the cluster gene lists with the stage- and phase-specific positive biomarker transcripts obtained in **chapter 5** showed that most of the biomarker transcripts were included in this analysis, confirming their profiles throughout berry development. A detailed study of cluster gene lists showed that these transcripts stuck correctly with the known physiological processes occurring during berry development. For example, the decreasing expression profiles throughout berry development of two sucrose synthase genes *VvSUS3* and *VvSUS5* and a tonoplast monosaccharide transporter *VvTMT3* suggested their involvement during berry formation, whereas the peak of expression of the hexose transporter gene *VvHT6*/*TMT2* and a SWEET gene at the end of véraison indicated their involvement in sugar accumulation during maturation phase. Moreover, the water inflow occurring during this phase, responsible for berry enlargement and softening, could be linked to the up-regulation of the aquaporin TIP1-3 gene just after véraison. Concerning the photosynthetic activity of the berry, photosynthesis-related gene expression was found to be shut down from either pea-sized berry stage (cluster 1, data not shown) or at the end of the green phase, as the decreasing trend of photosynthesis-related genes pointed out in cluster 3. Furthermore, the down-regulation of auxin-related genes after véraison confirmed the role of auxins as enhancers of berry formation and inhibitors of ripening. During this study, many genes related to cell wall formation/modification were identified, showing that each growth event – berry formation by rapid cell division and cell enlargement, berry expansion and softening during maturation phase – requested a distinct set of cell wall-related genes and also different genes belonging to large gene family as the expansin gene family. Interestingly, no real transcription of genes related to secondary metabolism was found, except genes involved in lignin biosynthesis and polymerisation – related genes involved in xylem formation during the green phase and in the formation of secondary cell wall related to cell aging at full-maturity – and the germacrene D synthase gene *VvTPS07*[50], involved in sesquiterpene production at

ripening in Gewürztraminer grapes[51], exhibiting an increasing gene expression profile during berry ripening. Consequently, this clustering analysis allowed the identification of sets of genes involved in the several biological process occurring at the different stages or during each phase in all 10-variety berries, but also provided new research perspectives. For example, the resulting cluster genes could help to restrain the target gene list of the identified transcription factor genes and to better characterize genes having no hit or devoid of functional annotation.

Furthermore, 80 transcripts were selected by their coefficient of variation among the 40-berry FPKM values. This analysis provided new possibilities for the selection of reference genes to compare gene expression between pericarps from different varieties using real-time qPCR technology. Nevertheless, this experiment was carried out on grapevine varieties cultivated in the same environmental conditions. The comparison of these genes with the non-plastic and constitutive genes of Corvina berry pericarps[63] showed that only two transcripts were shared by the two experiments, encouraging performing further experiments in order to obtain the best reference genes, i.e. the non-plastic and constitutive genes throughout whole berry development independently of berry skin colour. Moreover, expression profiles of the shared constitutive gene with the grapevine atlas[72] confirmed that it is the best candidate for reference gene for quantitative gene expression analysis carried out on all grapevine organs. Nevertheless, further experiments are necessary to confirm this hypothesis.

Finally, we established the first grape berry development transcriptomic route composed of 993 genes. This result represents a solid database for all the scientific community and more particularly for future research on grape berry development.

METHODS

Gene Clustering

Firstly, a hierarchical clustering analysis was performed on the 4,613 transcripts that were commonly up- or down-regulated between at least two subsequent developmental stages in all the ten varieties and showing an expression intensity above 1 FPKM, using Ward agglomeration method and Euclidean distance as the metric[79] (R package). In order to select the right number of clusters, a measure of the benefit deriving from the use of an additional cluster was calculated (**Fig. 5**).

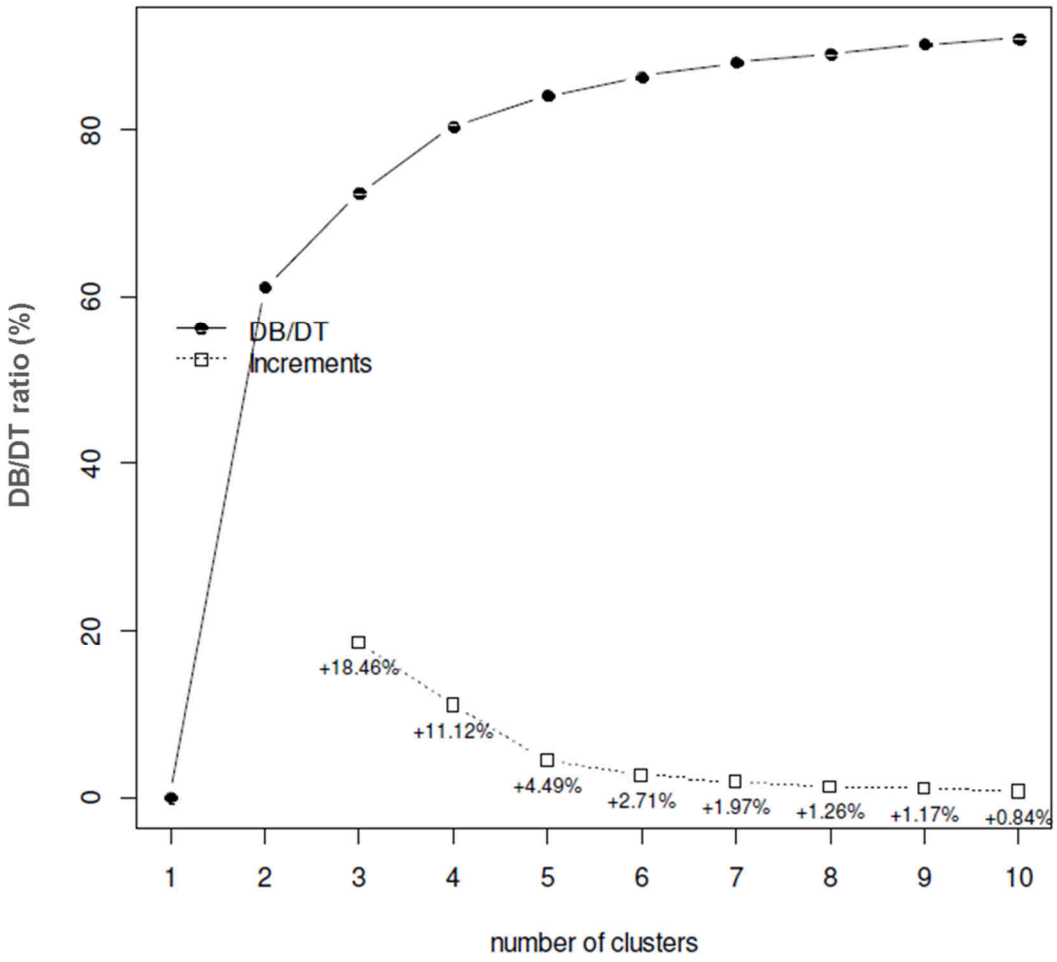


Figure 5: Plot of the benefit deriving from the use of an additional cluster. Solid black line shows the relationship between the number of clusters and the ratio between the deviance between clusters (DB) and the total deviance (DT). Dotted line shows the increase (in percentage terms) of the DB/DT giving a measure of the benefit resulting from an additional cluster (R software).

In **Figure 5**, the solid black line shows the relationship between the number of clusters and the ratio between the deviance between clusters (DB) and the total deviance (DT). Total deviance corresponds to the sum of DB and the deviance within clusters. The DB/DT ratio explains the percentage of total variation by clustering. Basically, higher is the DB/DT ratio, higher is the variability between clusters respect to the variability within clusters. The dotted line shows the increase (in percentage terms) of the DB/DT giving a measure of the benefit resulting from an additional cluster. When the solid line is flat (i.e. the DB/DT ratio is almost constant), the dotted line is close to zero and the benefit is almost null. Consequently, we decided to determine six as the number of clusters representing 1,197 transcripts with a DB/DT ratio of approximately 86%.

Then, the six resulting clusters were confirmed by k-means cluster analysis using Hartigan and Wong algorithm with a maximum number of 100 iterations and a

number random sets equal to 10. Dr. Marco Sandri realized the above-described biostatistics analyses.

Functional Category Distribution and GO Enrichment Analysis

All cluster transcripts were annotated against the V1 version[80] of the 12x draft annotation of the grapevine genome (http://genomes.cribi.unipd.it/DATA/). This was verified manually and integrated using gene ontology (GO) classifications. Transcripts were then grouped into the sixteen most represented functional categories: carbohydrate metabolic process (GO:0005975), cell wall metabolism (GO:0044036), cellular amino acids and derivative metabolic process (GO:0006520), cellular homeostasis (GO:0019725), cellular process (GO:0009987), developmental process (GO:0032502), DNA/RNA metabolic process (GO:0090304), generation of energy (GO:0006091), lipid metabolic process (GO:0006629), response to hormone stimulus (GO:0009725), response to stress (GO:0006950), secondary metabolic process (GO:0019748), signal transduction (GO:0007165), transcription factor activity (GO:0051090), transport (GO:000681), other processes (GO:0008150) based on GO biological processes. Genes with unknown functions or with a “no hit” annotation were also included. The distribution of functional categories is represented in a histogram showing the percentage of the genes in each cluster (**Fig. 2**).

GO enrichment analysis was performed by using BiNGO 2.4 plug-in tool in Cytoscape version 3.1.0 with PlantGOslim categories, as described by Maere *et al.*[13]. Overrepresented PlantGOslim categories were identified using a hypergeometric test with a significance threshold of 0.01 after a Benjamini & Hochberg False-Discovery Rate (FDR) correction[81].

SUPPLEMENTARY MATERIAL

Supplementary table 1: List of the 225 putative pea-sized berry stage-specific positive biomarker transcripts having a similar expression profile during development. For each transcript, a mean FPKM value of the 10 variety samples was calculated for each growth stage. Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

Gene_ID	Functional annotation	P	PV	EV	H
VIT_00s0131g00340	Annexin ANN4	3.81	0.47	0.19	0.03
VIT_00s0179g00340	Histone H2A variant 1 HTA11	5.45	0.22	0.09	0.01
VIT_00s0199g00210	Cyclin delta-2 (CYCD2)	7.07	1.92	0.03	0.00
VIT_00s0201g00060	Unknown protein	2.08	0.75	0.79	0.82
VIT_00s0203g00160	Cyclin D-type	13.31	2.49	0.18	0.10
VIT_00s0259g00170	Unknown protein	5.38	0.56	0.38	0.08
VIT_00s0323g00050	Invertase/pectin methylesterase inhibitor	11.31	2.43	1.02	1.58
VIT_00s0333g00050	DIR1 (defective IN induced resistance 1)	49.07	4.52	0.07	0.03
VIT_00s0480g00070	Polyphenol oxidase II, chloroplast precursor	156.67	27.69	7.93	0.72
VIT_00s0480g00080	Polyphenol oxidase II, chloroplast precursor	415.47	74.26	20.16	2.00
VIT_00s0499g00020	Unknown protein	12.45	1.03	0.26	0.02
VIT_00s0615g00020	Cinnamyl alcohol dehydrogenase	24.96	7.28	1.60	1.54
VIT_00s0665g00030	Unknown protein	26.38	5.93	1.02	0.21
VIT_01s0011g03210	Aspartyl protease	21.77	5.17	0.44	0.08
VIT_01s0011g03450	Alpha-glucosidase	43.40	6.52	3.58	1.32
VIT_01s0011g03860	RKL1 (Receptor-like kinase 1)	3.18	0.40	0.35	0.24
VIT_01s0011g04080	Zinc finger (C3HC4-type ring finger)	249.91	58.32	0.45	0.17
VIT_01s0011g06240	Cytochrome B561	9.10	1.63	1.12	0.35
VIT_01s0011g06590	Protease inhibitor/seed storage/lipid transfer protein (LTP)	12.06	1.56	0.09	0.02
VIT_01s0011g06610	Glutamate decarboxylase	4.15	0.58	0.23	0.14
VIT_01s0026g00350	Basic helix-loop-helix (bHLH) family	194.65	26.07	0.08	0.07
VIT_01s0026g00570	Bet v I allergen	134.83	7.55	1.08	0.07
VIT_01s0026g00970	Cupin family protein	4.77	0.64	0.16	0.08
VIT_01s0026g02340	Ankyrin protein kinase	266.48	106.07	44.16	33.40
VIT_01s0127g00400	Polygalacturonase GH28	13.02	0.68	0.01	0.00
VIT_01s0127g00850	Polygalacturonase BURP	34.42	4.75	0.09	0.02
VIT_01s0137g00390	Unknown protein	4.39	0.40	0.15	0.02
VIT_01s0137g00720	Lipase GDSL	2.33	0.36	0.04	0.00
VIT_01s0150g00480	Protein binding protein	19.75	4.93	0.37	0.08
VIT_02s0012g01350	No hit	15.89	1.52	0.19	0.01
VIT_02s0012g01870	No hit	56.85	13.87	6.21	2.54
VIT_02s0025g01040	Receptor protein kinase	9.63	2.08	0.21	0.07
VIT_02s0025g02620	BDG1 (BODYGUARD1) hydrolase	25.06	1.72	3.80	2.23
VIT_02s0025g02720	Unknown protein	2.96	0.30	0.17	0.07
VIT_02s0025g03190	2-oxoglutarate-dependent dioxygenase	13.50	1.32	1.15	0.32
VIT_02s0033g00260	Pinoresinol-lariciresinol reductase	2.01	0.15	0.09	0.02
VIT_03s0017g01380	Disease resistance protein (CC-NBS-LRR class)	2.04	0.42	0.31	0.08
VIT_03s0017g02240	Glucan endo-1,3-beta-glucosidase precursor	6.38	0.70	0.60	0.09
VIT_03s0017g02280	Myb family	2.62	0.31	0.24	0.06
VIT_03s0038g01410	Aquaporin PIP PIP1A	27.77	3.02	0.98	0.46
VIT_03s0038g01830	Proline-rich protein 4	103.34	5.87	0.84	0.19
VIT_03s0038g02170	Thaumatin	79.76	30.81	7.74	4.04
VIT_03s0038g03220	Chitin elicitor-binding CEBIP LysM domain-containing	41.71	9.36	2.66	1.10
VIT_03s0038g04000	Cysteine endopeptidase, papain-type (XCP1)	13.97	1.19	0.23	0.33
VIT_03s0063g00390	Lag one homologue 2	6.07	2.65	2.31	1.49
VIT_03s0063g01120	Kinesin phragmoplast orienting kinesin 2	2.06	0.72	0.43	0.19
VIT_03s0063g02130	UPF0326 protein hag1	2.04	0.30	0.07	0.01
VIT_03s0091g00450	Progesterone 5-beta-reductase	77.01	8.73	0.48	0.63
VIT_04s0008g01850	Trihelix DNA-binding protein (GT2)	13.97	3.66	0.88	0.13
VIT_04s0008g05150	Zinc finger (Ran-binding)	4.80	0.26	0.01	0.00

(Table continues on following page)

Supplementary table 1 (Continued from previous page)

Gene_ID	Functional annotation	P	PV	EV	H
VIT_04s0008g05450	Phospholipase D alpha 1 precursor (PLD 1)	2.01	0.22	0.15	0.14
VIT_04s0008g05830	Armadillo/beta-catenin repeat	64.28	11.82	3.73	1.41
VIT_04s0008g06670	Plastocyanin domain-containing protein	8.74	0.64	0.23	0.03
VIT_04s0008g06790	Biopterin transport protein BT1	11.77	4.20	4.32	2.32
VIT_04s0023g01960	Laccase (diphenol oxidase)-like protein	1.90	0.15	0.06	0.02
VIT_04s0023g03650	Galactose mutarotase-like	31.15	11.15	1.75	0.37
VIT_04s0023g03750	Ascorbate peroxidase	5.53	0.69	0.46	0.07
VIT_04s0044g01670	Ankyrin	74.26	20.97	8.65	7.44
VIT_04s0069g01120	Acyl-activating enzyme 7	3.07	0.13	0.00	0.00
VIT_04s0079g00230	Sucrose synthase (SUS5)	3.94	0.42	0.36	0.09
VIT_05s0020g00420	Polygalacturonase GH28	11.25	1.02	0.12	0.01
VIT_05s0020g00730	Receptor-like protein kinase	50.63	17.95	9.25	1.78
VIT_05s0020g02470	AWPM-19 membrane	6.01	0.70	0.37	0.09
VIT_05s0020g02480	Glutamine synthetase	80.97	29.51	5.82	2.01
VIT_05s0077g00540	No hit	29.23	3.41	0.73	0.72
VIT_05s0077g02270	Unknown protein	10.11	2.28	0.45	0.24
VIT_05s0077g02330	Transducin protein	9.21	2.80	0.33	0.15
VIT_06s0004g00050	Kinesin family member C3	6.36	1.42	0.11	0.15
VIT_06s0004g01680	Unknown protein	5.54	1.72	0.67	0.70
VIT_06s0004g02860	Unknown protein	3.87	0.24	0.08	0.12
VIT_06s0004g03050	fasciclin arabinogalactan-protein (FLA11)	4.78	0.28	0.04	0.01
VIT_06s0004g03770	Embryo-specific protein	8.61	1.11	0.86	0.13
VIT_06s0004g04370	Histone H4	38.04	3.65	2.62	3.00
VIT_06s0004g05790	Unknown protein	20.19	8.05	1.44	1.05
VIT_06s0004g05870	Tubulin beta-3 chain	8.83	0.63	0.28	0.23
VIT_06s0004g05940	Uncoupling protein 1	15.84	7.04	5.09	2.10
VIT_06s0004g06820	Unknown protein	10.26	1.46	1.78	0.39
VIT_06s0004g07460	Unknown protein	5.11	0.28	0.01	0.01
VIT_06s0004g07680	Nodulin	741.80	229.22	1.90	0.50
VIT_06s0009g01090	Heat shock HSP20 family protein	7.13	0.74	0.72	0.16
VIT_06s0009g01640	Receptor protein kinase	7.99	2.84	1.01	0.19
VIT_06s0061g00070	Ubiquitin-conjugating enzyme E2 C	3.82	0.41	0.17	0.24
VIT_06s0061g00140	Unknown protein	5.93	0.23	0.00	0.00
VIT_06s0061g00240	ABC transporter G member 7	10.87	3.24	2.02	0.44
VIT_06s0061g00520	Serine carboxypeptidase precursor	3.67	0.35	0.05	0.03
VIT_06s0061g00620	Short-chain dehydrogenase carbonyl reductase 3	42.56	10.10	1.99	0.77
VIT_06s0061g00850	Harpin-induced protein	4.44	0.62	0.45	0.07
VIT_06s0061g00890	Carboxypeptidase	15.85	3.38	0.30	0.23
VIT_06s0080g01260	Receptor-like kinase	21.60	11.41	7.13	4.73
VIT_07s0005g00730	Pectinesterase family	185.44	51.19	40.90	3.39
VIT_07s0005g01840	Patatin	9.05	0.84	0.24	0.02
VIT_07s0031g00320	Curly leaf PHCLF1	7.62	3.08	1.97	1.27
VIT_07s0031g02270	VvTMT3 Tonoplast monosaccharide transporter	12.00	1.87	0.05	0.00
VIT_07s0104g01820	Glutathione S-transferase 11 GSTF7	4.36	0.85	0.01	0.00
VIT_07s0104g01830	Glutathione S-transferase 8 GSTF8	45.80	8.99	0.46	0.06
VIT_07s0129g00190	Binding	4.42	0.83	0.26	0.15
VIT_07s0129g00560	No hit	86.41	16.34	3.79	0.85
VIT_07s0129g01070	Leucine-rich repeat protein kinase	12.24	1.88	0.11	0.03
VIT_07s0129g01100	Cyclin D	4.51	0.84	0.13	0.24
VIT_07s0151g00440	AP2 domain containing protein	6.02	1.53	0.72	0.55

(Table continues on following page)

Supplementary table 1 (Continued from previous page)

Gene_ID	Functional annotation	P	PV	EV	H
VIT_07s0151g00450	Basic helix-loop-helix (bHLH) family	3.29	0.39	0.16	0.14
VIT_08s0007g00410	Myb domain protein 91	23.79	6.39	0.70	0.13
VIT_08s0007g01850	Glycine-rich protein	5.09	1.61	0.97	0.69
VIT_08s0007g02920	ATAN11 (ANTHOCYANIN11)	3.62	0.87	0.12	0.22
VIT_08s0007g03990	Cellulose synthase CSLA09	45.31	15.43	4.52	1.43
VIT_08s0007g04150	RKL1 (Receptor-like kinase 1)	21.30	3.01	1.97	2.13
VIT_08s0007g04820	Pectate lyase	5.05	0.91	0.10	0.01
VIT_08s0007g05220	Unknown	6.55	0.95	0.48	0.04
VIT_08s0007g05520	Pentatricopeptide (PPR) repeat	4.30	1.48	0.96	0.95
VIT_08s0007g05610	Unknown protein	1.74	0.16	0.07	0.01
VIT_08s0007g08380	Cellulose synthase CESA3	16.58	4.23	1.09	0.66
VIT_08s0032g00100	Heat shock HSP20 family protein	4.90	0.38	0.25	0.00
VIT_08s0040g00930	Cyclin B-type	2.29	0.24	0.07	0.05
VIT_08s0040g03400	Short-chain dehydrogenase/reductase	67.74	13.51	2.39	0.97
VIT_08s0056g00500	Unknown protein	31.64	11.21	4.10	1.66
VIT_08s0058g01340	Leucine-rich repeat transmembrane protein kinase	5.69	2.46	1.62	0.85
VIT_09s0002g02220	Protein kinase CDG1	12.14	1.44	0.15	0.15
VIT_09s0002g03140	Lipase GDSL	120.19	8.27	3.40	0.54
VIT_09s0002g06110	Zinc finger (Ran-binding)	9.54	1.08	0.32	0.06
VIT_09s0002g06880	Beta-1,3-glucanase	6.19	0.90	0.48	0.23
VIT_09s0018g00520	Plastocyanin domain-containing protein	38.00	12.00	3.08	0.22
VIT_09s0018g01870	D-3-phosphoglycerate dehydrogenase, chloroplast precursor (3-PGDH)	99.06	24.07	5.17	2.33
VIT_10s0003g00220	Unknown	34.79	9.75	2.26	0.21
VIT_10s0003g00620	Lipase GDSL	17.67	1.00	2.94	0.51
VIT_10s0003g01050	CHUP1 (chloroplast unusual positioning 1)	11.88	1.49	0.07	0.05
VIT_10s0003g02900	LHCII type I CAB-1	894.28	184.64	13.76	0.79
VIT_10s0003g03320	Unknown	2.92	0.27	0.22	0.04
VIT_10s0003g03340	No hit	6.74	0.69	0.44	0.09
VIT_10s0003g03360	No hit	4.32	0.65	0.37	0.05
VIT_10s0003g03410	No hit	5.09	0.40	0.31	0.04
VIT_10s0003g03430	No hit	8.04	0.85	0.43	0.04
VIT_10s0003g03440	No hit	4.64	0.46	0.29	0.03
VIT_10s0003g03910	TCP family transcription factor 24	82.98	15.42	4.91	5.45
VIT_10s0071g00870	High mobility group B 6	4.18	0.25	0.17	0.36
VIT_10s0092g00360	VPS2.2 SNF7	3.04	0.86	0.42	0.59
VIT_10s0116g00300	Harpin-induced protein-related	1.93	0.22	0.26	0.23
VIT_10s0116g00560	Polyphenol oxidase II, chloroplast precursor	214.77	37.80	11.78	0.98
VIT_11s0016g00220	Glycosyl hydrolase family 17 protein	4.03	0.46	0.38	0.08
VIT_11s0016g00330	Pectinesterase family	56.34	7.04	1.75	1.13
VIT_11s0016g00470, VIT_11s0016g00480	Sucrose synthase (SUS1)	588.42	210.95	125.03	81.48
VIT_11s0016g02110	FAD linked oxidase, N-terminal	91.26	3.08	4.08	0.42
VIT_11s0016g02180	Amine oxidase	2.71	0.27	0.02	0.07
VIT_11s0016g02280	Glycine dehydrogenase 2	45.75	17.18	17.82	8.49
VIT_11s0016g02900	Leucine-rich repeat family protein / extensin	28.73	3.10	0.19	0.06
VIT_11s0016g03020	Pectinesterase family	11.85	1.15	0.02	0.04
VIT_11s0016g03640	Rac-like GTP-binding protein ARAC7 (GTPase protein ROP9)	205.95	36.20	0.79	0.27
VIT_11s0016g03750	Myb-related protein 3R-1 (Plant c-MYB-like protein 1) MYB3R1	1.81	0.42	0.26	0.36
VIT_11s0016g04540	ABC transporter G member 22	28.05	7.20	4.14	1.27
VIT_12s0028g02630	Rac-like GTP-binding protein RAC2	7.20	0.52	0.10	0.01
VIT_12s0028g03050	VvNAC34_NAC domain-containing protein	1.99	0.17	0.10	0.13

(Table continues on following page)

Supplementary table 1 (Continued from previous page)

Gene_ID	Functional annotation	P	PV	EV	H
VIT_12s0028g03190	Reticulon family protein	2.70	0.22	0.07	0.01
VIT_12s0035g01820	Proton-dependent oligopeptide transport (POT) family protein	13.87	3.20	0.37	1.76
VIT_12s0035g02190	MLO-like protein 13	3.60	0.40	0.37	1.00
VIT_12s0057g00500	Thymidine kinase	3.45	0.61	0.27	0.25
VIT_12s0059g00570	fasciclin arabinogalactan-protein (FLA7)	305.31	77.21	52.09	20.48
VIT_12s0059g01590	Lipase GDSL	108.12	28.93	10.98	1.95
VIT_12s0059g01830	Unknown	541.01	41.82	1.04	0.72
VIT_12s0059g02310	PDF2 (protodermal factor2)	6.17	2.30	1.15	1.43
VIT_12s0142g00440	ABA-responsive protein HVA22F	5.39	0.82	0.64	0.12
VIT_13s0019g02020	Thaumatin	5.57	1.47	0.23	0.36
VIT_13s0019g03280	LIM domain containing protein	137.64	70.33	54.50	41.68
VIT_13s0019g05290	Acyl-CoA binding	20.80	10.13	8.59	7.48
VIT_13s0064g00440	24-sterol C-methyltransferase	66.59	27.14	25.30	9.15
VIT_13s0064g00560	DNA topoisomerase, ATP-hydrolyzing	2.02	0.30	0.13	0.08
VIT_13s0320g00010	Lectin	2.76	0.63	0.05	0.02
VIT_14s0060g01960	Pectinesterase; Pectinesterase inhibitor	3.42	0.44	0.45	0.18
VIT_14s0066g00010	Unknown protein	1.53	0.06	0.01	0.00
VIT_14s0066g00100	Unknown protein	5.41	0.32	0.19	0.04
VIT_14s0066g00120	Unknown protein	23.68	1.12	1.16	0.14
VIT_14s0066g02130	Thaumatin	9.52	1.68	0.59	0.53
VIT_14s0068g00150	No hit	5.94	0.72	0.10	0.01
VIT_14s0068g00270	Hydroxyproline-rich glycoprotein	3.79	0.33	0.09	0.02
VIT_14s0068g02000	Ribonucleotide reductase R2	3.36	0.52	0.38	0.24
VIT_14s0108g00440	Tubulin alpha	16.51	1.18	0.54	0.04
VIT_14s0108g01020	VvEXPA16	331.53	17.80	78.37	13.22
VIT_14s0171g00210	Ankyrin	11.63	1.44	0.19	0.09
VIT_14s0219g00090	Unknown protein	1.90	0.25	0.11	0.07
VIT_15s0021g02300	VvSBP14_Squamosa promoter-binding protein	13.99	2.74	0.03	0.01
VIT_15s0046g02030	Protein kinase	40.52	16.26	5.16	1.22
VIT_15s0046g02410	Aquaporin TMP-C	43.47	11.88	5.48	3.54
VIT_15s0048g01210	Subtilisin serine endopeptidase (XSP1)	4.67	0.38	0.11	0.01
VIT_16s0039g02460	Chromomethylase CMT2	2.46	0.63	0.37	0.08
VIT_16s0039g02470	Chromomethylase CMT2	1.81	0.52	0.29	0.07
VIT_16s0050g00100	Myosin-related	9.63	1.87	1.12	0.14
VIT_16s0098g01070	Calmodulin-binding protein family	1.66	0.14	0.08	0.01
VIT_16s0098g01740	Unknown protein	3.35	0.68	0.40	0.30
VIT_17s0000g02010	Atypical receptor kinase MARK	78.34	15.27	8.07	1.88
VIT_17s0000g02600	Unknown protein	19.79	4.53	2.40	1.35
VIT_17s0000g03260	Reticulon (RTNLB9)	10.01	1.85	1.11	1.00
VIT_17s0000g05710	Unknown	2.99	0.40	0.22	0.02
VIT_17s0000g05720	No hit	7.90	0.87	0.46	0.06
VIT_17s0000g05730	No hit	8.59	1.00	0.56	0.06
VIT_17s0000g05830	Unknown protein	7.01	0.75	0.24	0.05
VIT_17s0000g05840	Calmodulin binding IQD31 (IQ-domain 31)	64.61	16.37	4.42	1.08
VIT_17s0000g06730	No hit	10.31	1.06	0.32	0.20
VIT_17s0000g08920	Ribitol dehydrogenase	18.07	3.99	0.13	0.10
VIT_17s0053g00700	Sucrose synthase 2 (SUS3)	10.09	0.87	0.59	0.08
VIT_17s0053g00740	VvNAC57_NAC domain-containing protein	1.32	0.14	0.10	0.03
VIT_18s0001g02000, VIT_18s0001g02010	Zinc finger (C2H2 type) family	9.31	2.87	0.29	0.08
VIT_18s0001g03570	Thaumatin ATLP-1	8.59	0.82	0.42	0.09

(Table continues on following page)

Supplementary table 1 (Continued from previous page)

Gene_ID	Functional annotation	P	PV	EV	H
VIT_18s0001g05160	Glycosyl hydrolase family 3 protein	3.18	0.62	0.19	0.17
VIT_18s0001g05180	Beta-D-xylosidase	112.83	44.81	4.57	0.37
VIT_18s0001g05480	Unknown protein	34.83	15.48	13.99	7.73
VIT_18s0001g07220	Cyclin delta-2	19.43	5.70	4.14	2.45
VIT_18s0001g07550	Kinesin family member 4/7/21/27	1.70	0.39	0.39	0.49
VIT_18s0001g09040	LIM domain protein WLIM1	117.14	35.69	7.72	1.05
VIT_18s0001g09920	Cyclin delta-3 (CYCD3_1)	12.07	1.03	0.41	0.02
VIT_18s0001g10040	LRX1 (leucine-rich repeat/extensin 1)	13.92	1.77	0.46	0.06
VIT_18s0001g10070	Receptor kinase	11.82	2.45	0.07	0.01
VIT_18s0001g10420	Remorin	10.44	2.34	0.56	0.26
VIT_18s0001g11300	LYS/HIS transporter 7 LHT7	3.21	0.34	0.06	0.01
VIT_18s0001g12530	Retrotransposon protein, Unclassified	22.03	1.82	0.24	0.38
VIT_18s0001g15000	ACT domain containing protein (ACR4)	246.61	89.45	12.59	4.48
VIT_18s0001g15510	Unknown	2.10	0.31	0.05	0.00
VIT_18s0072g00510	No hit	2.57	0.32	0.27	0.05
VIT_18s0076g00360	EMB1075 (embryo defective 1075) carboxy-lyase	214.90	68.42	42.39	27.35
VIT_18s0122g00550	Cyclin-dependent kinase B2;1	3.97	0.58	0.43	0.94
VIT_18s0122g00790	Cysteine endopeptidase, papain-type (XCP1)	8.30	0.83	0.10	0.02
VIT_18s0122g00980	Glucan endo-1,3-beta-glucosidase 7 precursor	165.04	30.98	2.86	0.14
VIT_19s0014g00090	Glucan endo-1,3-beta-glucosidase 4 precursor	56.24	12.21	0.21	0.04
VIT_19s0014g01470	Thioredoxin TTL1 (Tetratricopetide-repeat thioredoxin-like 1)	11.62	4.24	1.69	0.44
VIT_19s0015g00530	fasciclin arabinogalactan-protein (FLA1)	74.53	11.06	1.78	0.30
VIT_19s0015g01230	Unknown protein	3.84	0.54	0.29	0.24
VIT_19s0085g00810	Glucan endo-1,3-beta-glucosidase precursor	2.47	0.24	0.10	0.02
VIT_19s0177g00200	Plastocyanin domain	13.71	1.73	1.07	0.13

Supplementary table 2: List of the 100 putative green phase-specific positive biomarker transcripts having a similar expression profile during development. For each transcript, a mean FPKM value of the 10 variety samples was calculated for each growth stage. Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

Gene_ID	Functional annotation	P	PV	EV	H
VIT_00s0194g00080	Phototropic-responsive NPH3	19.6	23.0	2.4	0.9
VIT_00s0389g00020	Acyl-CoA oxidase (ACX1)	58.7	58.3	28.5	24.5
VIT_00s0616g00020	Ribosomal protein L19	85.0	90.6	38.7	18.8
VIT_00s1667g00010	LEM3 (ligand-effect modulator 3) family protein	53.9	47.2	24.3	11.2
VIT_01s0010g00590	Short-chain dehydrogenase/reductase SDR	11.2	11.0	3.5	2.3
VIT_01s0010g00750	Unknown protein	12.0	10.6	3.6	2.2
VIT_01s0010g02100	Indeterminate(ID)-domain 5	20.4	20.2	7.3	5.1
VIT_01s0010g03570	NLI interacting factor (NIF)	14.6	13.8	6.1	5.4
VIT_01s0011g02740	Phosphoenolpyruvate carboxylase	149.4	125.3	24.4	11.2
VIT_01s0137g00650	Unknown protein	89.5	72.9	36.2	15.6
VIT_02s0012g01420	Pigment defective 322	30.2	30.3	12.4	7.5
VIT_02s0025g01740	Auxin response factor 9	46.7	48.6	8.2	3.4
VIT_02s0025g03590	Phospholipid hydroperoxide glutathione peroxidase	96.1	113.5	8.9	2.6
VIT_02s0025g04710	Unknown protein	84.7	81.3	27.2	12.9
VIT_02s0087g00800	Unknown protein	68.3	67.4	35.3	32.5
VIT_03s0063g00100	Unknown protein	31.9	34.7	13.6	13.9
VIT_03s0063g01110	N-acetyl-gamma-glutamyl-phosphate reductase, chloroplast precursor	38.8	40.2	15.2	10.9
VIT_03s0091g00430	SWIB complex BAF60b domain-containing protein	82.0	85.6	27.9	17.3
VIT_04s0008g00200	Tryptophan-tRNA ligase	9.4	9.0	3.8	3.0
VIT_04s0008g01160	Indeterminate(ID)-domain 7	52.6	45.7	16.0	8.8
VIT_04s0008g05960	Unknown protein	59.4	54.3	20.9	26.0
VIT_04s0008g07340	Constans-like 4	136.5	134.8	36.2	36.2
VIT_04s0023g02620	Ubiquitin-conjugating enzyme E2 W	90.9	96.9	46.7	54.4
VIT_04s0043g00380	Unknown protein	7.9	8.3	2.8	2.2
VIT_04s0043g00860	Unknown protein	25.3	27.1	6.4	6.2
VIT_04s0044g01850	Auxin efflux carrier	166.2	144.7	56.1	19.8
VIT_04s0044g01870	Auxin efflux carrier	12.6	12.4	0.9	0.3
VIT_04s0079g00680	VvPSY1_phytoene synthase (PSY)	34.3	38.9	14.0	5.4
VIT_05s0020g00400	Galactosyltransferase family protein	16.4	13.4	4.7	2.2
VIT_05s0020g02690	Copper-binding family protein	240.8	284.9	23.9	7.5
VIT_05s0049g01480	Ribosomal protein L28, chloroplast (CL28) 50S	93.0	88.2	33.2	15.7
VIT_05s0062g00630	UDP-glucose transferase (UGT75B2)	8.1	7.7	0.8	0.1
VIT_05s0077g00560	Serine carboxypeptidase S10	3.5	4.1	0.3	0.2
VIT_05s0077g01240	Calmodulin-binding transcription activator 6	105.4	111.9	33.5	22.3
VIT_05s0094g00820	Unknown protein	34.1	30.7	9.3	5.3
VIT_06s0004g00220	Protein kinase APK1B	8.2	8.0	0.9	0.2
VIT_06s0004g03130	Auxin response factor 4	63.2	60.8	4.0	0.5
VIT_06s0004g03690	Glutathione S-transferase GSTO1	107.3	101.2	39.8	24.2
VIT_06s0004g04130	Protein tyrosine phosphatase protein (PAS2)	35.7	31.5	13.1	2.6
VIT_06s0004g04320	Alcohol dehydrogenase 3	33.7	27.8	6.6	1.4
VIT_06s0004g04430	Ubiquitin-conjugating enzyme E2 G1	104.9	107.9	45.9	37.6
VIT_06s0004g08380	Unknown protein	62.3	63.8	22.6	10.8
VIT_07s0005g00100	LPA1 (LOW PSII accumulation1)	14.0	13.7	5.0	3.1
VIT_07s0005g06440	Drug/metabolite transporter DMT family transporter	17.4	18.6	9.4	8.1
VIT_07s0031g00160	Ubiquitin-conjugating enzyme E2 I	154.5	151.5	62.8	67.5
VIT_07s0031g00780	Transcriptional factor B3	22.0	21.7	10.4	7.4
VIT_07s0031g02030	Polypyrimidine tract-binding protein 3	41.7	47.1	22.4	24.3
VIT_07s0031g02390	Phytoalexin-deficient 4 protein (PAD4)	91.8	91.9	19.2	17.6
VIT_07s0104g00630	Zinc finger (C3HC4-type ring finger)	43.1	43.7	22.6	23.5
VIT_07s0104g01290	TUBBY like protein 3 TLP3	72.2	69.6	34.2	30.3

(Table continues on following page)

Supplementary table 2 (Continued from previous page)

Gene_ID	Functional annotation	P	PV	EV	H
VIT_07s0141g00590	Magnesium transporter CorA-like family protein (MRS2-2)	35.7	32.6	14.4	4.3
VIT_08s0007g01570	fructose 1,6-bisphosphatase	92.5	87.8	21.1	9.7
VIT_08s0007g03930	Delta14-sterol reductase FK (FACKEL)	34.6	31.9	13.5	9.6
VIT_08s0007g06360	Unknown protein	35.3	36.0	10.2	13.1
VIT_08s0007g07470	Galactose mutarotase	6.5	6.5	1.3	0.4
VIT_08s0056g01120	MATE efflux family protein	18.9	18.1	4.2	1.4
VIT_10s0092g00640	Unknown protein	57.6	58.7	11.1	10.7
VIT_11s0016g00100	Adapter protein SPIKE1 (SPK1)	24.0	21.9	9.6	5.0
VIT_11s0016g02530	Protein arginine N-methyltransferase	50.6	48.4	20.7	16.8
VIT_11s0016g03960	Kinesin family member 2/24	57.6	58.4	24.1	14.0
VIT_11s0016g05510	Unknown protein	9.4	9.2	3.1	3.4
VIT_12s0028g01080	Photosystem II oxygen-evolving complex precursor, 23kda PSBP	481.2	433.8	152.7	60.7
VIT_12s0035g00970	Evolutionarily conserved C-terminal region 11 ECT11	77.1	77.2	19.2	6.7
VIT_12s0035g01800	Auxin response factor 1	56.4	53.9	28.6	33.4
VIT_12s0057g00600	Unknown protein	133.7	146.7	49.7	34.5
VIT_12s0059g01180	Bromo-adjacenty (BAH) domain-containing protein	47.4	44.2	21.3	11.5
VIT_13s0067g01950	Chromatin remodeling 42	13.0	12.5	4.5	2.0
VIT_13s0067g03420	RabGAP/TBC domain-containing protein	53.4	47.3	25.4	12.8
VIT_13s0084g00110	ABC transporter I member 8	47.9	43.9	21.0	17.3
VIT_14s0006g01980	Zinc finger (C3HC4-type ring finger)	49.2	44.3	12.4	18.0
VIT_14s0066g00870	ATPase-like domain-containing	8.9	10.2	3.0	2.5
VIT_15s0024g00040	LHCA3 (Photosystem I light harvesting complex gene 3)	480.1	430.0	81.8	19.7
VIT_15s0046g02460	Growth-on protein GRO11	105.3	113.7	48.8	52.0
VIT_15s0048g02090	Thioesterase family	11.4	11.6	3.1	1.4
VIT_16s0039g02550	Seed specific protein Bn15D1B	103.9	116.9	6.8	4.4
VIT_16s0050g02240	Unknown protein	344.6	333.1	95.4	45.4
VIT_17s0000g00930	Unknown protein	12.2	11.9	4.6	2.3
VIT_17s0000g01790	4-coumarate-CoA ligase 2	10.5	12.3	2.1	3.3
VIT_17s0000g02440	Integral membrane protein	17.7	20.3	1.7	0.5
VIT_17s0000g05610	Isopiperitenol dehydrogenase	5.2	5.2	0.7	0.4
VIT_17s0000g05650	Unknown protein	68.6	71.2	17.5	10.4
VIT_17s0000g09620	Rhodanese domain-containing protein	14.0	14.0	4.4	2.2
VIT_17s0000g10260	AarF domain containing kinase	35.2	34.4	13.7	8.5
VIT_18s0001g00280, VIT_18s0001g00290	OPCL1 (OPC-8:0 COA ligase1)	64.6	75.9	17.1	13.5
VIT_18s0001g00760	Plastocyanin, chloroplast precursor	278.8	274.0	93.9	28.3
VIT_18s0001g03850	Ribosomal protein S10 30S	76.8	72.9	32.2	19.5
VIT_18s0001g04790	Aspartyl protease	21.0	21.1	5.2	0.5
VIT_18s0001g05040	Pollen Ole e 1 allergen and extensin	593.5	479.9	136.9	18.5
VIT_18s0001g06720	Rieske [2Fe-2S] domain	37.0	33.5	10.4	4.1
VIT_18s0001g08090	IAA9	292.4	296.4	86.7	58.4
VIT_18s0001g10550	LHCA5 (Photosystem I light harvesting complex gene 5)	27.7	30.6	2.1	0.2
VIT_18s0041g00840	UDP-glucose: anthocyanidin 5,3-O-glucosyltransferase	5.4	6.1	0.3	0.0
VIT_18s0072g01160	Unknown	9.0	8.5	3.8	3.6
VIT_18s0089g00910	Auxin response factor 1	48.7	45.7	15.2	17.4
VIT_18s0122g00750	HSP20-like chaperone	2.2	2.1	0.4	0.1
VIT_18s0122g01140	Wuschel-related homeobox 13	38.9	42.5	11.8	4.9
VIT_18s0157g00180	S-adenosylmethionine carrier 1	19.4	18.8	2.8	0.5
VIT_19s0014g02970	Choline transporter	14.1	14.2	4.4	2.1
VIT_19s0015g01550	Phenylalanine-tRNA ligase	8.4	7.6	1.9	0.5
VIT_19s0015g01920	Protein phosphatase 2C	11.6	11.3	4.0	4.1

Supplementary table 3: List of the 80 transcripts having a coefficient of variation below than 0.1 among the 40-sample data set. The average of the 40 FPKM values was done and the coefficient of variation was calculated by dividing the standard deviation by the average of the 40 FPKM values.

Gene_ID	Functional annotation	Coefficient of variation	Average FPKM value
VIT_13s0073g00610	Ubiquitin-specific protease 12	0.0650	58.4
VIT_04s0044g00580	Actin 7 (ACT7) / actin 2	0.0655	1,295.7
VIT_02s0025g00100	Adaptor-related protein complex 2, beta 1 subunit	0.0673	86.0
VIT_04s0044g01210	Phosphatidylinositol 4-kinase	0.0712	14.3
VIT_08s0007g07570	CCR4-NOT transcription complex subunit 3	0.0761	52.6
VIT_01s0026g00310	Adaptor protein complex AP-1, gamma 1 subunit	0.0763	57.3
VIT_08s0007g01790	Defective in exine formation protein (DEX1)	0.0766	31.2
VIT_15s0021g02760	Unknown	0.0774	9.8
VIT_05s0077g00470	UPL6 (ubiquitin protein LIGASE 6)	0.0791	34.9
VIT_05s0077g00140	Transducin protein	0.0804	18.9
VIT_07s0005g05160	Proteasome 26S AAA-ATPase subunit (RPT2)	0.0806	154.7
VIT_00s0174g00150	Ubiquitin-specific protease 5 (UBP5)	0.0811	14.9
VIT_00s0125g00260	Unknown	0.0813	9.3
VIT_07s0031g01280	Zinc finger (C3HC4-type ring finger)	0.0817	26.4
VIT_00s0225g00220	Adaptor-related protein complex 1, mu 1 subunit	0.0828	89.0
VIT_01s0010g03700	Zinc finger (Ran-binding)	0.0831	62.3
VIT_11s0016g04420	DnaJ homolog, subfamily C, member 13	0.0834	33.5
VIT_01s0011g03160	PFT1 (phytochrome and flowering time 1) MED25	0.0843	34.0
VIT_18s0001g14860	SKP1	0.0847	323.2
VIT_14s0030g01350	WD-repeat protein 48	0.0851	23.3
VIT_08s0007g06240	CDK5RAP3	0.0854	17.0
VIT_14s0036g01460	SNF2 domain-containing protein helicase ATRX/CHR20	0.0854	13.8
VIT_04s0008g02930	Oligouridylate binding protein 1B UBP1B	0.0856	108.7
VIT_13s0158g00140	SEC31 homolog B	0.0857	84.6
VIT_10s0003g03860	KH domain-containing protein	0.0861	48.7
VIT_01s0137g00300	SDA1	0.0873	17.4
VIT_17s0000g00580	Calmodulin-7 (CAM7)	0.0876	551.6
VIT_13s0019g04730	Glycerol-3-phosphate dehydrogenase	0.0878	9.8
VIT_01s0026g01700	Unknown protein	0.0882	84.3
VIT_05s0124g00360	Protein kinase Pti1	0.0882	173.1
VIT_10s0003g03710	Cullin-4	0.0882	37.3
VIT_00s1373g00010	Transducin protein	0.0884	32.1
VIT_06s0004g07730	Inner envelope protein (IEP110) precursor	0.0885	25.0
VIT_18s0122g01080	Calcium ion binding protein	0.0888	37.5
VIT_15s0045g00540	Vesicle coat complex COPII, subunit SEC23	0.0891	18.4
VIT_00s1238g00010	BSD domain-containing protein	0.0901	64.0
VIT_07s0104g00060	XPD/UVH6 (ultraviolet hypersensitive 6)	0.0901	9.5
VIT_11s0206g00130	VAMP-like protein YKT61	0.0902	175.9
VIT_08s0040g01430	Unknown protein	0.0904	74.6
VIT_06s0004g00140	Protein kinase C substrate 80K-H	0.0905	40.3
VIT_18s0001g09310	Transcription initiation factor TFIID subunit 1-A	0.0906	20.1
VIT_04s0008g06080	Unknown protein	0.0909	20.8
VIT_18s0122g00040	DnaJ homolog, subfamily A, member 2	0.0911	80.2
VIT_02s0012g00910	Adaptor-related protein complex 2, mu 2 subunit	0.0913	86.5
VIT_05s0020g00790	HAT dimerisation domain-containing protein	0.0914	8.4
VIT_11s0065g01010	Methyl-CpG-binding domain 2 MBD02	0.0915	46.6
VIT_03s0091g01030	Telomere repeat binding factor 1	0.0927	19.9
VIT_06s0004g03190	Proteasome 20S beta subunit G1 (PBG1) (PRCH)	0.0929	164.6
VIT_16s0098g00340	Translation initiation factor eIF-3 subunit 10	0.0929	12.7
VIT_19s0014g05310	Sec23/sec24 transport	0.0934	58.9

(Table continues on following page)

Supplementary table 3 (Continued from previous page)

Gene_ID	Functional annotation	Coefficient of variation	Average FPKM value
VIT_14s0083g00360	Unknown protein	0.0938	19.1
VIT_13s0067g03320	Tonneau 1b TON1B	0.0939	89.3
VIT_14s0066g02380	Protein transport SEC13	0.0942	124.4
VIT_04s0023g00800	LUC7 N_terminus domain-containing protein	0.0943	48.9
VIT_09s0002g02140	Endomembrane protein 70, TM4 ;	0.0949	131.7
VIT_03s0063g02260	Phosphatidylinositol 4-kinase	0.0951	49.4
VIT_06s0004g03970	Unknown protein	0.0951	19.1
VIT_11s0052g00980	Yip1 domain , member 6	0.0951	48.3
VIT_18s0122g00070	DNAJ heat shock N-terminal domain-containing protein	0.0953	79.1
VIT_11s0052g00170	Armadillo-like helical domain-containing	0.0953	12.6
VIT_14s0068g01280	V-type H+-transporting ATPase subunit C	0.0954	130.9
VIT_12s0035g00280	Endomembrane protein 70	0.0959	81.9
VIT_09s0002g05770	Histone H2A variant 3 HTA9	0.0959	99.7
VIT_10s0071g00950	Serine/threonine protein phosphatase 2A	0.0968	74.8
VIT_01s0011g03410	DNA repair protein RAD23	0.0968	100.5
VIT_07s0095g00730	Sterile alpha motif (SAM) domain-containing protein	0.0970	44.8
VIT_16s0100g00320	Zfwd2 protein (ZFWD2)	0.0974	34.0
VIT_05s0124g00270	Unknown protein	0.0974	11.6
VIT_00s0189g00020	Wound-responsive protein	0.0979	304.6
VIT_06s0004g08370	Unknown protein	0.0982	16.0
VIT_01s0011g04600	Quiescin-sulfhydryl oxidase 2 ATQSOX2	0.0985	18.3
VIT_14s0068g00220	Proteasome	0.0985	68.1
VIT_13s0064g01230	U6 snRNA-associated Sm-like protein LSm7	0.0986	52.7
VIT_05s0077g00820	Unknown protein	0.0988	35.6
VIT_04s0023g02750	FG-GAP repeat-containing protein	0.0990	41.3
VIT_09s0002g06550	Signal peptidase, endoplasmic reticulum-type	0.0990	94.3
VIT_08s0040g00410	Proteasome 20S alpha subunit G (PAG1) (PRC8)	0.0992	161.0
VIT_14s0006g02960	KH domain containing protein	0.0992	26.2
VIT_05s0020g01580	CBS	0.0992	13.1
VIT_11s0016g04290	3-hydroxyisobutyryl-CoA hydrolase	0.0997	48.9

REFERENCES

1. Koenig D., et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc. Natl. Acad. Sci. USA. 110:2655-2662 (2013)
2. Martínez-López L.A, Ochoa-Alejo N., Martínez O. Dynamics of the chili pepper transcriptome during fruit development. BMC Genomics 15:143 (2014)
3. Feng C., Chen M., Xu C.J., Bai L., Yin X.R., Li X., Allan A.C., Ferguson I.B., Chen K.S. Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. BMC Genomics 13:19 (2012)
4. Huang G., et al. Comparative Transcriptome analysis of climacteric fruit of Chinese pear (*Pyrus ussuriensis*) reveals new insights into fruit ripening. Plos One vol.9 (2014)
5. Lin Q., Wang C., Dong W., et al. Transcriptome and metabolome analyses of sugar and organic acid metabolism in Ponkan (*Citrus reticulata*) fruit during fruit maturation. Gene 554(1):64-74 (2014)
6. Zenoni S., Ferrarini A., Giacomelli E., et al. Characterization of transcriptional complexity during berry development of *Vitis vinifera* using RNA-seq. Plant Physiol. 152(4):1787-1795 (2010)
7. Sweetman C., Wong D.C.J., Ford C.M., Drew D.P. Transcriptome analysis at four developmental stages of grape berry (*Vitis vinifera* cv. Shiraz) provides insights into regulated and coordinated gene expression. BMC Genomics 13:691 (2012)
8. Da Silva C., Zamperin G., Ferrarini A., et al. 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(12):4777-4788 (2013)
9. Degu A., Hochberg U, Sikron N., et al. Metabolite and transcript profiling of berry skin during fruit development elucidates differential regulation between Cabernet Sauvignon and Shiraz cultivars at branching points in the polyphenol pathway. BMC Plant Biology 14:188 (2014)
10. Tavazoie S., Hughes J.D., Campbell M.J., Cho R.J., Church G.M. Systematic determination of genetic network architecture. Nat. Genet. 22:281-285 (1999)
11. Jaskowiak P.A., Campello R.J., Costa I.G. On the selection of appropriate distances for gene expression data clustering. BMC Bioinformatics 15(Suppl 2):S2 (2014)
12. Palumbo M.C., Zenoni S., Fasoli M., Massonnet M., Farina L., Castiglione F., Pezzotti M., Paci P. Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development. Plant Cell 26(12):4617-4635 (2014)
13. Maere S., Heymans K., Kuiper M. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21:3448-3449 (2005)
14. Goddard R.H., Wick S.M., Silflow C.D., Snustad D.P. Microtubule component of the plant-cell cytoskeleton. Plant Physiol. 104:1-6 (1994)
15. Dhugga K. S., et al. Guar seed β-mannan synthase is a member of the cellulose synthase super gene family. Science 303:363-366 (2004)
16. Seymour G.B & Gross K.C. Cell wall disassembly and fruit softening. Postharvest News and Information 7:45N52N (1996)

17. Desprez T., Juraniec M., Crowell E.F., Jouy H., Pochylova Z., Parcy F., Höfte H., Gonneau M., Vernhettes S. Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci. USA. 104:15572-15577 (2007)
18. Taylor N.G., Scheible W.R., Cutler S., Somerville C.R., Turner S.R. The irregular xylem3 locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell 11(5):769-780 (1999)
19. Taylor N.G., Laurie S., Turner SR. Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. Plant Cell 12(12):2529-2540 (2000)
20. Rose J. K., Braam J., Fry S.C., Nishitani K. The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: Current perspectives and a new unifying nomenclature. Plant Cell Physiol. 43:1421-1435 (2002)
21. Fry S.C., Smith R.C., Renwick K.F., Martin D.J., Hodge S.K., Matthews K.J. Xyloglucan endotransglycosylase, a new wall-loosening enzyme activity from plants. Biochem J. 15:282 (1992)
22. Antosiewicz D.M., Purugganan M.M., Polisensky D.H., Braam, J. Cellular localization of Arabidopsis xyloglucan endotransglycosylase-related proteins during development and after wind stimulation. Plant Physiol. 115:1319-1328 (1997)
23. Thompson J.E. & Fry S.C. Restructuring of wall-bound xyloglucan by transglycosylation in living plant cells. Plant J. 26:23-34 (2001)
24. Farkas V., Sulova Z., Stratilova E., Hanna R., Maclachlan G. Cleavage of xyloglucan by nasturtium seed xyloglucanase and transglycosylation to xyloglucan subnit oligosaccharides. Arch. Biochem. Biophys. 298: 365-370 (1992)
25. Fanutti C., Gidley M.J., Reid J.S.G. Action of a pure xyloglucan endo-transglycosylase (formerly called xyloglucan-specific endo-1,4-β-D-glucanase) from the cotyledons of germinated nasturtium seeds. Plant J. 3:691-700 (1993)
26. Bourquin V., et al. Xyloglucan Endotransglycosylases have a function during the formation of secondary cell walls of vascular tissues. Plant Cell 14:3073-3088 (2002)
27. Matsui A., et al. AtXTH27 plays an essential role in cell wall modification during the development of tracheary elements. Plant J. 42:525-534 (2005)
28. Jürgens G. Cytokinesis: The Art of Partitioning. The Plant Cell. 12:827 (2000)
29. Zhang X.Y., Wang X.L., Wang X.F., Xia G.H., Pan Q.H., Fan R.C., Wu F.Q., Yu X.C., Zhang D.P. A shift of Phloem unloading from symplasmic to apoplasmic pathway is involved in developmental onset of ripening in grape berry. Plant Physiol. 142(1):220-232 (2006)
30. Dal Santo S., Vannozzi A., Tornielli G.B., Fasoli M., Venturini L., Pezzotti M., Zenoni S. Genome-wide analysis of the expansin gene superfamily reveals grapevine-specific structural and functional characteristics. PLoS One 8(4):e62206 (2013)
31. De Veylder L., Beeckman T., Inzé D. The ins and outs of the plant cell cycle. Nature Reviews Molecular Cell Biology 8:655-665 (2007)
32. Vanholme R., Demedts B., Morreel K., Ralph J., Wout Boerjan W. Lignin Biosynthesis and Structure. Plant Physiol. 153:895-890 (2010)

33. Zhao Q., Nakashima J., Chen F., Yin Y., Fu C., Yun J., Shao H., Wang X., Wang Z.Y., Dixon R.A. Laccase is necessary and nonredundant with peroxidase for lignin polymerization during vascular development in Arabidopsis. Plant Cell 25(10):3976-3987 (2013)
34. Xin H., Zhang J., Zhu W., Wang N., Fang P., Han Y., Ming R., Li S. The effects of artificial selection on sugar metabolism and transporter genes in grape. Tree Genetics and Genomes 9:1343-1349 (2013)
35. Wormit A., Trentmann O., Feifer I., et al. Molecular identification and physiological characterization of a novel monosaccharide transporter from Arabidopsis involved in vacuolar sugar transport. Plant Cell 18:3476-3490 (2006)
36. Hawker J.S. Changes in activities of enzymes concerned with sugar metabolism during development of grape berries. Phytochemistry 8:9-17 (1969)
37. D'Aoust M.A., Yelle S., Nguyen-Quoc B. Antisense inhibition of tomato fruit sucrose synthase decreases fruit setting and the sucrose unloading capacity of young fruit. Plant Cell 11(12):2407-2418 (1999)
38. Giovane A., Servillo L., Balestrieri C., Raiola A., D'Avino R., Tamburrini M., Ciardiello M.A., Camardella L. Pectin methylesterase inhibitor. Biochim Biophys Acta. 1696:245-252 (2004)
39. Ejiri S. Moonlighting functions of polypeptide elongation factor 1: from actin bundling to zinc finger protein R1-associated nuclear localization. Biosci Biotechnol Biochem. 66:1-21 (2002)
40. Famiani F., Farinelli D., Palliotti A., Moscatello S., Battistelli A., Walker RP. Is stored malate the quantitatively most important substrate utilised by respiration and ethanolic fermentation in grape berry pericarp during ripening? Plant Physiol Biochem. 76:52-57 (2014)
41. Wang S., Hagen G., Guilfoyle T.J. ARF-Aux/IAA interactions through domain III/IV are not strictly required for auxin-responsive gene expression. Plant Signal Behav. 8:6 (2013)
42. Inaba A., Ishida M., Sobajima Y. Changes in endogenous hormone concentrations during berry development in relation to ripening of Delaware grapes. J Japan Soc Hort Sci. 45:245-252 (1976)
43. Böttcher C., Keyzers R.A., Boss P.K., Davies C. Sequestration of auxin by the indole-3-acetic acid-amido synthetase GH3-1 in grape berry (*Vitis vinifera* L.) and the proposed role of auxin conjugation during ripening. J Exp Bot. 61:3615-3625 (2010)
44. Terrier N., Glissant D., Grimplet J., et al. Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (*Vitis vinifera* L.) development. Planta 222(5): 832-847 (2005)
45. Deluc L.G., Grimplet J., Wheatley M.D., et al. Transcriptomic and metabololite analyses of Cabernet Sauvignon grape berry development. BMC Genomics 8:429 (2007)
46. Lecourieux F., Kappel C., Lecourieux D., Serrano A., Torres E., Arce-Johnson P., Delrot S. An update on sugar transport and signalling in grapevine. J Exp Bot. 65(3):821-832 (2014)
47. Chen L.Q., Qu X.Q., Hou B.H., Sosso D., Osorio S., Fernie A.R., Frommer W.B. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science 335:207-211 (2012)
48. Maurel C., Verdoucq L., Luu D.T., Santoni V. Plant Aquaporins: Membrane Channels with Multiple Integrated Functions. Annual Review of Plant Biology 59:595-624 (2008)
49. Cosgrove D.J. Loosening of plant cell walls by expansins. Nature 407: 321-326 (2000)

50. Martin D.M., Aubourg S., Schouwey M.B., Daviet L., Schalk M., Toub O., Lund S.T., Bohlmann J. 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 (2010)
51. Martin D.M., Chiang A., Lund S.T., Bohlmann J. 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(3):919-929 (2012)
52. Rosenquist J.K. & Morrison J.C. The development of the cuticle and epicuticular wax of the grape berry. Vitis 27:63-70 (1988)
53. Holloway P.J. The chemical constitution of plant cuticle. In Cutler DF, Alvin KL, Price CE (eds) The Plant Cuticle. Academic Press, London, pp. 45-86 (1982)
54. Comménil P., Brunet L., Audran J.C. The development of the grape berry cuticle in relation to susceptibility to bunch rot disease. J. Exp. Bot. 48(313):1599-1607 (1997)
55. Casado C.G. & Heredia A. Ultrastructure of the cuticle during growth of the grape berry (*Vitis vinifera*). Physiol. Plant. 111:220-224 (2001)
56. Wang N., Zheng Y., Xin H., Fang L., Li S. Comprehensive analysis of NAC domain transcription factor gene family in *Vitis vinifera*. Plant Cell Rep. 32(1):61-75 (2013)
57. Zhang Y., Feng J.C. Identification and characterization of the grape WRKY family. Biomed Res Int. 2014:787680 (2014)
58. Robatzek S., Somssich I.E. A new member of the Arabidopsis WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defense-related processes. Plant J. 28:123-133 (2001)
59. Krasnow M., Matthews M., Shackel K. Evidence for substantial maintenance of membrane integrity and cell viability in normally developing grape (*Vitis vinifera* L.) berries throughout development. J Exp Bot. 59(4):849-859 (2008)
60. Marín I. Animal HECT ubiquitin ligases: evolution and functional implications. BMC Evol Biol. 10:56 (2010)
61. Marín I. Evolution of plant HECT ubiquitin ligases. PLoS One. 8(7):e68536 (2013)
62. Reid K.E., Olsson N., Schlosser J., Peng F., Lund ST. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol. 6:27-37 (2006)
63. Dal Santo S., Tornielli G.B., Zenoni S., Fasoli M., Farina L., Anesi A., Guzzo F., Delledonne M., Pezzotti M. The plasticity of the grapevine berry transcriptome. Genome Biol. 14:r54 (2013)
64. Wilkinson K.D. Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. FASEB J. 11:1245-1256 (1997)
65. Yan N., Doelling J.H., Falbel T.G., Durski A.M., Vierstra R.D. The ubiquitin-specific protease family from Arabidopsis. AtUBP1 and 2 are required for the resistance to the amino acid analog canavanine. Plant Physiol. 124:1828-1843 (2000)

66. Reyes-Turcu F.E., Ventii K.H., Wilkinson K.D. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. Annu Rev Biochem. 78:363-397 (2009)
67. Smalle J., Vierstra R.D. The ubiquitin 26S proteasome proteolytic pathway. Annu Rev Plant Biol. 55:555-590 (2004)
68. Lambermon M.H., Simpson G.G., Wieczorek Kirk D.A., Hemmings-Mieszczak M., Klahre U., Filipowicz W. UBP1, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. EMBO J. 19:1638-1649 (2000)
69. Weber C., Nover L., Fauth M. Plant stress granules and mRNA processing bodies are distinct from heat stress granules. Plant J. 56(4):517-530 (2008)
70. Goodall G.J. & Filipowicz W. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell 58(3):473-483 (1989)
71. Ko C.H., Brendel V., Taylor R.D., Walbot V. U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize. Plant Mol Biol. 36(4):573-583 (1998)
72. Fasoli M., Dal Santo S., Zenoni S., Tornielli G.B., Farina L., Zamboni A., Porceddu A., Venturini L., Bicego M., Murino V., Ferrarini A., Delledonne M., Pezzotti M. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. Plant Cell 24:3489-3505 (2012)
73. Holstein S.E. Clathrin and plant endocytosis. Traffic 3:614-620 (2002)
74. Chen X., Irani N.G., Friml J. Clathrin-mediated endocytosis: the gateway into plant cells. Curr Opin Plant Biol. 14:674-682 (2011)
75. Samaj J., Read N.D., Volkmann D., Menzel D., Baluska F. The endocytic network in plants. Trends Cell Biol. 15:425-433 (2005)
76. Dhonukshe P., Aniento F., Hwang I., Robinson D.G., Mravec J., Stierhof Y.D., Friml J. Clathrin-mediated constitutive endocytosis of PIN auxin efflux carriers in Arabidopsis. Curr Biol. 17:520-527 (2007)
77. Beck M., Zhou J., Faulkner C., MacLean D., Robatzek S. Spatio-temporal cellular dynamics of the Arabidopsis flagellin receptor reveal activation status-dependent endosomal sorting. Plant Cell 24:4205-4219 (2012)
78. Wang C., Yan X., Chen Q., Jiang N., Fu W., Ma B., Liu J., Li C., Bednarek S.Y., Pan J. Clathrin light chains regulate clathrin-mediated trafficking, auxin signaling, and development in Arabidopsis. Plant Cell 25:499-516 (2013)
79. Müllner D. fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. Journal of Statistical Software 53(9):1-18 (2013)
80. Forcato C. Gene prediction and functional annotation in the *Vitis vinifera* genome. PhD Thesis, Universita' Degli Studi Di Padova (2010)
81. Klipper-Aurbach Y., Wasserman M., Braunspiegel-Weintrob N., Borstein D., Peleg S., Assa S., Karp M., Benjamini Y., Hochberg Y., Laron Z. Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. Med. Hypotheses 45:486-490 (1995)

Chapter 8

Differences between red- and white-skinned berry transcriptomes during maturation phase

INTRODUCTION

Berry maturation phase is crucial to guarantee the dispersal of the seeds by herbivores. The only goal of this phase is to drive the berry attractiveness. Consequently, the process includes the softening of the berry, the increase of sugars in the pulp, the decline of organic acids and tannins, and the synthesis of volatile aromas[1]. Nevertheless, full-ripe berries from diverse wine cultivars differ by their shape, weight, their final composition, but the most evident phenotypic difference is the berry skin colour between red- and white-skinned berries. Red skin colour is due to the biosynthesis of anthocyanins in red grapes, which is regulated by the additive activity of two MybA-type transcription factors: VvMybA1 and VvMybA2. In white cultivars, the allelic variation caused by the insertion of the *Gret1* Gypsy-type retrotransposon into the promoter of *VvMybA1* prevents *VvMybA1* transcription[2] and thus prevents UDP-glucose:flavonoid 3-O-glucosyltransferase (*UFGT*) gene expression, which is responsible for anthocyanin biosynthesis[3,4]. In addition, the mutated allele of *VvMybA2* encodes a protein with a non-conservative amino acid change, rendering it unable to trigger *UFGT* transcription[5,6].

In this work, several Principal Component Analyses (PCAs) were performed to understand differences of transcriptome patterns and behaviours between red and white grapes throughout maturation phase. T-test analyses were done in order to identify differentially expressed genes between red- and white-skinned berries at the two ripening stages. In addition, another approach, consisting in the extraction of transcript loadings from PCA outcome and the detection of the loading threshold-breaking point, was used to select genes involved in berry transcriptome differences at maturity. A deep analysis of these genes showed that phenylpropanoid/flavonoid biosynthetic pathway-related gene expression is not sufficient to explain the differences between red- and white-grape transcriptomes, and that many other biological processes are implicated in this phenomenon. Therefore, this study permitted to gain insights into these biological processes and to indicate those that were specific to each skin colour group during maturation phase. Finally, total anthocyanin content of red grapes was measured by spectrophotometry allowing the statement of a hypothesis.

RESULTS AND DISCUSSION

Red- and white-skinned berry transcriptome relationships during maturation phase

In **chapter 5**, performing a principal component analysis (PCA) of the whole data set (as shown again in the plot in **Figure 1**) allowed observing that the main difference between white and red berry samples was localized during the maturation phase, especially at harvest stage when red-variety samples separate from the white ones.

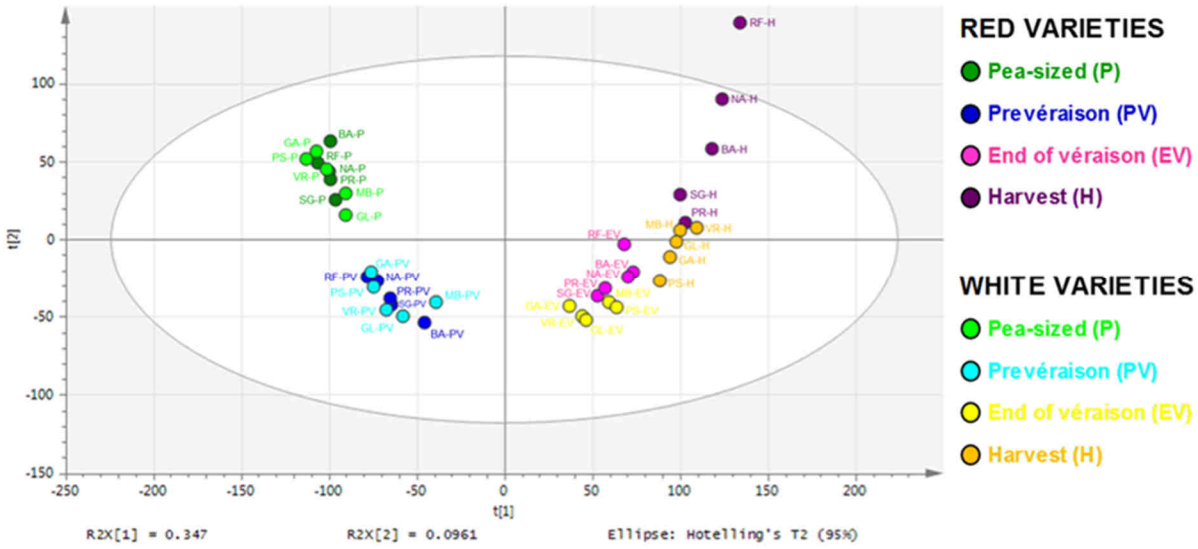


Figure 1: Plot of 40-sample data set PCA showing the first two principal components t[1]=34.7% and t[2]=9.61%.

Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest) and for each variety: SG (Sangiovese), BA (Barbera), NA (Negro amaro), RF (Refosco), PR (Primitivo) for the red grape varieties and VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina) for the white grape varieties.

In order to unravel the biological reasons of this result, two distinct PCA analyses were performed with either the either the 20 red-coloured samples (**Fig. 2 a**) – corresponding to the five red grape varieties at the four developmental stages (average of the three biological replicates) – or the 20 white-coloured samples (**Fig. 2 b**) – namely the five white grape varieties at the four growth stages (mean of the replicates). In **Figure 2**, distribution of red and white-berry samples is almost identical during the pre-véraison phase, with pea-sized berry stage and prevéraison samples separating mainly by PC2. Pre-véraison and post-véraison samples then separate along PC1: red-variety samples get separated by PC2 as if harvest samples slide in an oblique way with the

contribution of both PC1 and PC2, whereas white-variety samples cluster into two groups right next to each other by PC1.

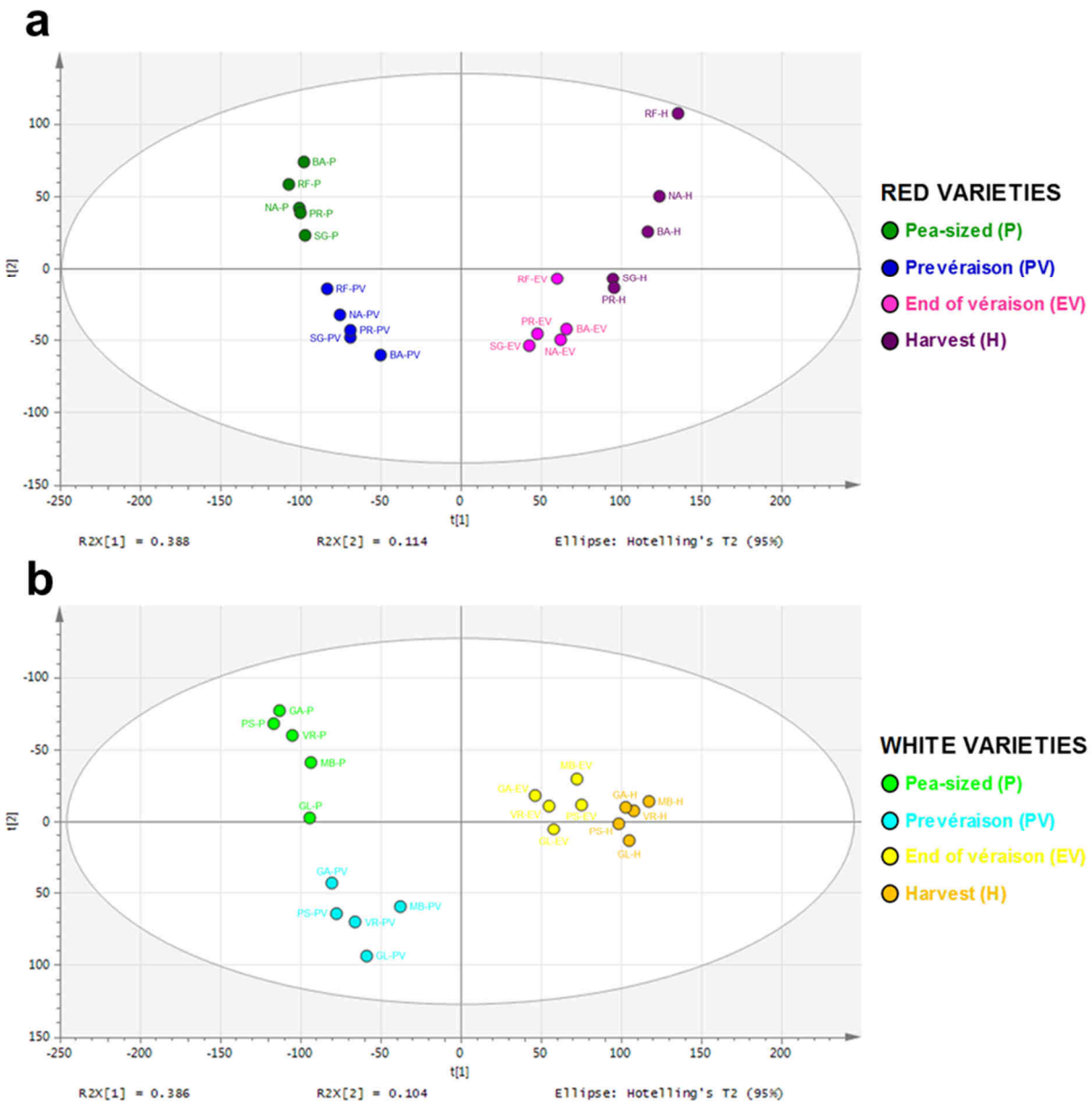


Figure 2: (a) Plot of 20-red-skinned-sample data set PCA showing the first two principal components t[1]=38.8% and t[2]=11.4%.

(b) Plot of 20-white-skinned-sample data set PCA showing the first two principal components t[1]=38.6% and t[2]=10.4%.

Abbreviations for the four developmental stages: P (Pea-sized), PV (Prevéraison), EV (End of véraison), H (Harvest) and for each variety: SG (Sangiovese), BA (Barbera), NA (Negro amaro), RF (Refosco), PR (Primitivo) for the red-skinned varieties and VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina) for the white-skinned varieties.

Differentially expressed genes between red- and white-skinned berries during maturation phase

In order to identify the genes responsible for the difference between red- and white-skinned berry transcriptome behaviours during the maturation phase, the five red- and the five white-skinned berry transcriptomes were compared at the end of véraison and harvest stages using between-subjects t-test with an overall threshold p-value of 0.01 (TMeV 4.3; http://www.tm4.org/mev). Thus, 443 DEGs were identified at the end of véraison, representing 2% of the expressed genes, corresponding to 382 and 61 genes with a higher expression level in red- and white-skinned samples respectively. Concerning the harvest stage, 837 were found as differentially expressed between red- and white-skinned samples, i.e. 4% of the data set, including 536 and 289 genes that were highly expressed in red- and white-skinned samples respectively.

Genes exhibiting a higher expression in red-skinned berries during maturation phase

Genes having a higher expression in red grapes at the end of véraison

Among the 382 genes with higher expression in red grapes, as expected, genes having the greatest log_2-fold change (Log_2-FC) values (**Table 1**) were involved in the general phenylpropanoid pathway[7], as two phenylalanine ammonia-lyase genes, a trans-cinnamate 4-monooxygenase gene and a 4-coumarate-CoA ligase gene; in flavonoid pathway[7], as two chalcone synthase genes, a chalcone isomerase gene, a flavonone-3-hydroxylase gene and ten flavonoid 3',5'-hydroxylase genes; in anthocyanin synthesis[3,8] with the expression of the UDP-glucose:flavonoid 3-O-glucosyltransferase gene *VvUFGT*, in anthocyanin modification[9,10] with two anthocyanin *O*-methylesterase (AOMT) genes and an anthocyanin acyl-transferase gene, and in anthocyanin transport[11,12] as the glutathione S-transferase gene *VvGST4* and the Multidrug And Toxic Extrusion (MATE) efflux family protein gene *VvAM3*. Also the transcription factor gene *VvMYBA1*, responsible for anthocyanin synthesis[13,14], was comprised among these genes, however it was found as expressed in white-skinned berries at both ripening stages, even though at a lower expression intensity than in red-skinned berries. That was surely due to a mismapping issue (**chapter 9**).

Table 1: Genes with a greater expression in red grapes than white grapes at the end of véraison (>2 log_2-fold change). Adjusted p-value (Adj p-value) indicates the significance of the differential expression between red and white grapes. Expression intensity average is given for each colour group in FPKM and log_2-fold change (Log_2-FC) was calculated for each transcript.

Gene_ID	Adj p-value	Functional annotation	Red grapes	White grapes	Log_2-FC
VIT_05s0051g00510	2.0E-03	Outer membrane protein	6.97	0.00	**16.1**
VIT_04s0079g00690	3.7E-05	Glutathione S-transferase (GST4)	1,428.47	0.15	**13.3**
VIT_16s0039g02230	7.0E-06	UDP-glucose:flavonoid 3-O-glucosyltransferase	123.04	0.03	**12.0**
VIT_03s0017g00870	1.5E-03	Anthocyanin Acyl-transferase	43.83	0.02	**11.2**
VIT_01s0010g03510	3.0E-05	Caffeoyl-CoA O-methyltransferase (AOMT)	199.92	0.09	**11.1**
VIT_06s0009g02880	4.0E-05	Flavonoid 3',5'-hydroxylase	38.17	0.02	**11.0**
VIT_01s0010g03490	8.8E-06	Caffeoyl-CoA O-methyltransferase (AOMT)	24.30	0.02	**10.1**
VIT_06s0009g02970	2.1E-06	Flavonoid 3',5'-hydroxylase	41.54	0.05	**9.7**
VIT_01s0010g03460, VIT_01s0010g03470, VIT_01s0010g03480	1.6E-05	Caffeoyl-CoA O-methyltransferase	27.18	0.04	**9.3**
VIT_06s0009g02840	2.2E-06	Flavonoid 3',5'-hydroxylase	45.95	0.09	**8.9**
VIT_06s0009g02810	3.0E-07	Flavonoid 3',5'-hydroxylase	43.92	0.13	**8.5**
VIT_08s0007g03560	1.7E-03	Anthocyanin membrane protein 1 (Anm1)	15.85	0.05	**8.4**
VIT_06s0009g02830	1.1E-05	Flavonoid 3',5'-hydroxylase	91.81	0.43	**7.8**
VIT_06s0009g02920	2.9E-05	Flavonoid 3',5'-hydroxylase	25.10	0.17	**7.2**
VIT_18s0001g09400	1.7E-06	Cytochrome b5 DIF-F (VvCytoB5)	295.75	3.24	**6.5**
VIT_16s0050g00910	1.5E-06	MATE efflux family protein AM3	250.96	2.96	**6.4**
VIT_06s0009g02860	3.7E-04	Flavonoid 3',5'-hydroxylase (F3'5'Hf)	8.81	0.11	**6.3**
VIT_06s0009g03000	3.1E-04	Flavonoid 3',5'-hydroxylase	68.93	0.94	**6.2**
VIT_03s0088g00260	4.4E-05	Serine carboxypeptidase S10	17.06	0.28	**5.9**
VIT_06s0009g02910	1.9E-03	Flavonoid 3',5'-hydroxylase	4.03	0.07	**5.9**
VIT_06s0009g03010	3.0E-04	Flavonoid 3',5'-hydroxylase	74.40	1.28	**5.9**
VIT_18s0001g06750	3.6E-03	Steroleosin-B	2.23	0.07	**4.9**
VIT_13s0019g04460	3.0E-04	Phenylalanine ammonia-lyase 2 (PAL2)	63.59	2.31	**4.8**
VIT_05s0136g00260	7.0E-04	Chalcone synthase (CHS3)	981.35	37.13	**4.7**
VIT_06s0004g02620	8.7E-04	Phenylalanine ammonia-lyase	60.07	3.20	**4.2**
VIT_00s0485g00010	3.9E-03	Receptor kinase homolog LRK14	1.86	0.10	**4.2**
VIT_04s0023g03370	4.2E-03	Flavonone- 3-hydroxylase (F3H1)	113.82	6.75	**4.1**
VIT_11s0016g04330	3.1E-03	Unknown	4.14	0.32	**3.7**
VIT_08s0040g00440	1.5E-03	Unknown protein	34.04	2.67	**3.7**
VIT_06s0004g08150	2.5E-04	Trans-cinnamate 4-monooxygenase	90.20	8.76	**3.4**
VIT_00s2679g00010	7.0E-03	No hit	13.57	1.35	**3.3**
VIT_10s0003g01180, VIT_10s0003g01190	4.5E-04	Ketol-acid reductoisomerase	3.71	0.42	**3.1**
VIT_05s0020g01900	7.7E-03	No hit	2.34	0.28	**3.0**
VIT_02s0033g00410	2.4E-03	VvMYBA1	71.21	9.86	**2.9**
VIT_08s0007g05580	1.4E-04	Embryo-abundant protein	15.75	2.44	**2.7**
VIT_14s0068g00920	8.0E-03	Chalcone synthase (CHS2)	157.28	25.26	**2.6**
VIT_05s0020g04290	6.1E-04	No hit	1.34	0.22	**2.6**
VIT_17s0000g04360	4.8E-03	Wall-associated kinase 2 (WAK2)	0.45	0.08	**2.6**
VIT_07s0005g05040	5.9E-03	ARR24 Type A	0.82	0.15	**2.4**
VIT_04s0023g02480	5.3E-03	Dehydrin 1b	23.01	4.67	**2.3**
VIT_16s0039g02040	4.8E-03	4-coumarate-CoA ligase 3	12.95	2.65	**2.3**
VIT_11s0052g00570	1.3E-03	Nodulin MtN21	2.98	0.64	**2.2**
VIT_18s0075g00080	8.9E-03	Anaphase-promoting complex component APC1	0.75	0.16	**2.2**
VIT_13s0067g03820	7.2E-03	Chalcone--flavonone isomerase (CHI1)	183.13	41.07	**2.2**
VIT_13s0067g02290	3.7E-03	BTB/POZ domain-containing protein	1.66	0.38	**2.1**
VIT_02s0234g00110	2.3E-03	Oleosin OLE-4	36.83	8.40	**2.1**
VIT_18s0117g00370	2.3E-03	R protein L6	5.21	1.29	**2.0**

Furthermore, among these 382 genes, the nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (GO:0006139) was identifies as a significantly

overrepresented Gene Ontology (GO) category (p-value = 1.2079E-5), using the enrichment tool BiNGO[15] and the GoSlim Plants annotation. This overrepresented GO category included genes related to DNA/RNA metabolic process as a DNA replication protein gene (VIT_05s0077g00800), a helicase gene (VIT_14s0068g00810) and RNA polymerase II mediator complex (VIT_02s0025g02580), and also twenty-four transcription factor genes (**Supplementary table 1**).

Genes with higher expression in red full-ripe grapes

At harvest, 536 genes were identified as having a higher expression intensity in red-skinned berries. In order to gain insights into biological processes, the genes were analysed for overrepresented functions using the GO enrichment tool BiNGO[15] and the GoSlim Plants annotation (**Fig. 3**).

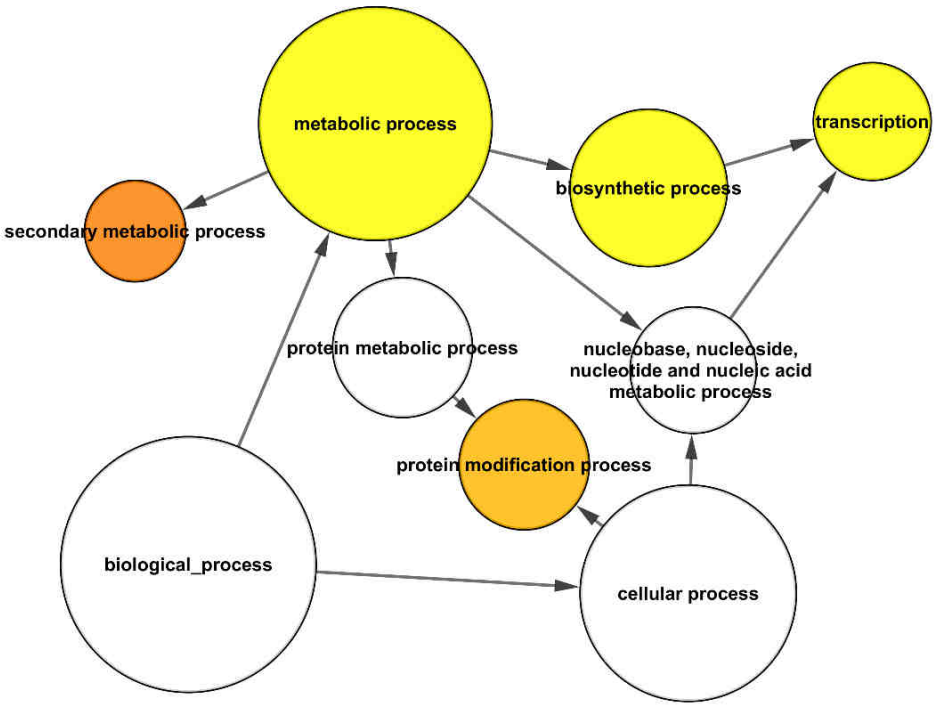


Figure 3: Enriched GO terms in genes exhibiting a higher expression in red grapes at harvest stage. The network graphs show BiNGO visualizations of the overrepresented GO terms for the 536 transcripts with a higher expression in red grapes than white grapes at harvest stage. Node size is positively correlated with the number of genes belonging to the category. Non-coloured nodes are not overrepresented, but they may be the parents of overrepresented terms. Coloured nodes represent GO terms that are significantly overrepresented (Benjamini and Hochberg corrected p-value < 0.01), with the shade indicating significance as shown in the colour bar.

Five GO categories were found significantly overrepresented: secondary metabolic process (GO:0019748), protein modification process (GO:0006464), metabolic process (GO:0008152), transcription (GO:0006350) and biosynthetic process (GO:0009058).

Among the genes belonging to the overrepresented secondary metabolic process GO category, we identified forty-two transcripts (**Table 2**) mostly involved in phenylpropanoid, flavonoid, anthocyanin and stilbene synthesis. Some stilbenes, particularly resveratrol, have long been known for their benefits to human health[16]. We also found an ascorbate oxidase gene involved in defensive mechanism against the deleterious actions of Reactive Oxygen Species (ROS)[17-19] that are known to be generated at the onset of berry ripening[20]. Melino and co-workers[21] described that ripening berries showed a low ascorbate:dehydroascorbate ratio suggesting a higher ascorbate oxydation than regeneration. The higher expression level of this ascorbate oxidase gene in red grapes suggests that red-skinned berries underwent an important oxidative stress at harvest. In addition, two genes related to lignin synthesis were identified: a sinapyl alcohol dehydrogenase gene which encoded protein catalyses the last step in syringyl monolignol biosynthesis[22] and a laccase gene necessary for lignin polymerisation[23], suggesting the formation of secondary cell wall in the red-skinned berries at harvest.

Regarding genes belonging to the protein modification category, most part of them (56/82) encode kinases involved in signal transduction, as four S-receptor kinase genes (VIT_12s0028g01650; VIT_13s0156g00590; VIT_04s0008g05500; VIT_12s0028g01640) and five wall-associated kinase genes (VIT_18s0001g11620; VIT_17s0000g04370; VIT_01s0011g03980; VIT_17s0000g04380; VIT_17s0000g04360), showing that processes specific of red grapes maturation involved signalling cascades.

With regard to the metabolic process category, presence of several genes involved in carbohydrate metabolic process was noticed, as glycolysis with three pyruvate kinase genes (VIT_06s0004g00130; VIT_08s0007g05430; VIT_08s0007g05490) and sugar transport with the hexose transporter gene *VvHT5* (VIT_05s0020g03140) and cell-wall apoplastic invertase gene *VvcwINV4* (VIT_09s0002g02320). Hayes and colleagues[24] observed that *VvHT5* mRNA accumulation, although weak, was mostly associated with late ripening days in Cabernet Sauvignon berries whereas Afoufa-Bastien and co-workers[25] described that *VvHT5* transcripts were hardly detectable in Chardonnay berries. In addition, Hayes and colleagues[24] discovered that *VvcwINV4* transcript

levels increased progressively up to véraison and declined around 10 weeks after flowering, and subsequently increased in the final stages of berry ripening. Our results suggest that red grapes expressed preferentially *VvHT5* and *VvcwINV4* genes in comparison to white grapes.

Table 2: Genes of the overrepresented secondary metabolic process GO category that exhibited a higher expression in red grapes than white grapes at the end of véraison. Adjusted p-value (Adj p-value) indicates the significance of the differential expression between red and white grapes. Expression intensity average is given for each colour group in FPKM and log_2-fold change (Log_2-FC) was calculated for each transcript.

Gene_ID	Adj p-value	Functional annotation	Red grapes	White grapes	Log_2-FC
VIT_04s0079g00690	7.8E-06	Glutathione S-transferase (GST4)	621.73	0.18	**11.7**
VIT_01s0010g03510	2.3E-04	Caffeoyl-CoA O-methyltransferase (AOMT)	29.37	0.02	**10.9**
VIT_16s0100g01170	1.0E-02	VvSTS46_Stilbene synthase	85.70	1.38	**6.0**
VIT_10s0003g04910	5.8E-03	Sinapyl alcohol dehydrogenase	1.75	0.03	**5.6**
VIT_16s0100g00910	7.3E-03	VvSTS21_Stilbene synthase	266.02	5.73	**5.5**
VIT_16s0100g00830	9.3E-03	VvSTS15_Stilbene synthase	273.93	6.38	**5.4**
VIT_16s0100g00780	5.8E-03	VvSTS10_Stilbene synthase	142.57	3.73	**5.3**
VIT_16s0100g00900	9.5E-03	VvSTS20_Stilbene synthase	233.59	6.57	**5.2**
VIT_16s0039g01130	9.3E-03	Phenylalanin ammonia-lyase	87.93	2.55	**5.1**
VIT_16s0100g01190	7.3E-03	VvSTS47_Stilbene synthase	140.09	4.47	**5.0**
VIT_16s0100g00960	7.5E-03	VvSTS26_Stilbene synthase	158.28	5.06	**5.0**
VIT_16s0100g01020	1.0E-02	VvSTS30_Stilbene synthase	107.05	3.54	**4.9**
VIT_16s0100g01000	8.2E-03	VvSTS28_Stilbene synthase	323.93	11.07	**4.9**
VIT_07s0031g01070	5.1E-03	Ascorbate oxidase	5.85	0.21	**4.8**
VIT_16s0100g00850	5.0E-03	VvSTS17_Stilbene synthase	58.88	2.30	**4.7**
VIT_16s0100g01070	4.1E-03	VvSTS35_Stilbene synthase	125.64	5.18	**4.6**
VIT_16s0039g01100	8.2E-03	Phenylalanin ammonia-lyase	138.68	5.72	**4.6**
VIT_16s0039g01170	5.5E-03	Phenylalanine ammonium lyase	86.09	3.93	**4.5**
VIT_16s0039g01110	6.1E-03	Phenylalanin ammonia-lyase	45.13	2.15	**4.4**
VIT_16s0100g01200	4.1E-03	VvSTS48_Stilbene synthase	270.00	13.31	**4.3**
VIT_16s0039g01360	7.0E-03	Phenylalanin ammonia-lyase	42.92	2.13	**4.3**
VIT_00s2849g00010	7.8E-03	Phenylalanine ammonia-lyase	107.35	5.76	**4.2**
VIT_16s0039g01300	7.2E-03	Phenylalanine ammonia-lyase	70.85	3.81	**4.2**
VIT_16s0100g00750	2.5E-03	VvSTS7_Stilbene synthase	91.62	5.31	**4.1**
VIT_05s0136g00260	3.2E-04	Chalcone synthase (CHS3)	488.23	28.33	**4.1**
VIT_16s0039g01240	7.0E-03	Phenylalanin ammonia-lyase	39.46	2.43	**4.0**
VIT_00s2508g00010	1.0E-02	Phenylalanine ammonia-lyase	35.02	2.18	**4.0**
VIT_16s0039g01280	6.4E-03	Phenylalanin ammonia-lyase	36.00	2.35	**3.9**
VIT_16s0100g01100	4.2E-03	VvSTS36_Stilbene synthase	43.30	2.97	**3.9**
VIT_06s0004g02620	4.0E-04	Phenylalanine ammonia-lyase	24.59	1.83	**3.8**
VIT_18s0117g00550	6.9E-03	Laccase	12.65	0.95	**3.7**
VIT_06s0004g05700	1.2E-03	Glutathione S-transferase 8	1.46	0.11	**3.7**
VIT_11s0052g01090	8.8E-03	4-coumarate-CoA ligase 2	74.72	6.60	**3.5**
VIT_08s0040g01710	4.3E-03	Phenylalanine ammonia-lyase (PAL1)	102.35	9.53	**3.4**
VIT_16s0039g02040	2.5E-03	4-coumarate-CoA ligase 3	20.56	2.14	**3.3**
VIT_14s0068g00920	5.0E-04	Chalcone synthase (CHS2)	86.29	14.50	**2.6**
VIT_00s0615g00020	3.4E-03	Cinnamyl alcohol dehydrogenase	2.64	0.44	**2.6**
VIT_03s0063g00140	8.2E-03	Caffeoyl-CoA O-methyltransferase	13.15	2.27	**2.5**
VIT_02s0025g02920	4.9E-03	Quercetin 3-O-methyltransferase 1	65.60	14.23	**2.2**
VIT_14s0068g00930	2.6E-03	Chalcone synthase (CHS1)	19.13	4.23	**2.2**
VIT_13s0067g03820	5.2E-03	Chalcone--flavonone isomerase (CHI1)	266.39	59.16	**2.2**
VIT_18s0001g12800	2.3E-03	Dihydroflavonol 4-reductase	97.40	23.06	**2.1**
VIT_00s0218g00010	3.2E-03	Cinnamyl alcohol dehydrogenase	1.02	0.28	**1.9**

Finally, forty-seven transcription factor genes were identified in the enriched transcription GO category (**Supplementary table 2**). Among these transcription factor genes, ten encode WRKY proteins. WRKY name is derived from the conserved DNA binding domain sequence WRKYGQK; the conservative domain is approximately 60 residues, followed by a Cys2His2 or Cys2HisCys zinc-binding motif[26,27]. WRKY proteins were divided into Group I-III based on the number of WRKY domains and the structure of zinc fingers[28,29]. WRKY proteins have been found to play essential roles in pathogen defence in response to bacteria[30,31], fungi[32,33], and viruses[34,35]. Evidence also supported that WRKY transcription factors were involved in modulating gene expression in plants during abiotic stresses such as cold[36,37], salt[38,39] and drought[40,41]. Besides roles in response to biotic and abiotic stress, WRKY proteins were also implicated in processes that modulate plant developmental processes such as morphogenesis of trichomes and embryos, senescence, dormancy, and metabolic pathways[29]. Concerning grapevine, Liu and colleagues[42] showed that the expression of *VvWRKY11* (VIT_04s0069g00920), isolated from "Beifeng" – an interspecific cultivar of *Vitis thunbergii* × *Vitis vinifera*, increased dehydration tolerance in transgenic Arabidopsis plants. Consequently, the higher expression of *VvWRKY11* in red grapes suggests that red-skinned berries might have expressed this transcription factor to cope with a dehydration stress at harvest that was not felt by white grapes. However, as red and white variety vines were cultivated in the same vineyard, it is very unlike that only the red ones underwent water-deficit stress. It is then possible that *VvWRKY11* could be related to the response to other related environmental stimuli or be involved in other processes typical of red varieties that have not been characterized yet. Indeed, although eighty individual WRKY genes have been identified in grapevine[43], the WRKY gene family in grapevine remains totally uncharacterized.

Genes with a greater expression in red grapes during the whole maturation phase

In total, 58 transcripts were found to have a higher accumulation level at both end of véraison and harvest stages (**Table 3**). Thus, these genes, expressed the whole ripening time until harvest, might be involved in the onset of ripening in red grapes. The major part of these transcripts (19/58) was involved in phenylpropanoid pathway, and more particularly in anthocyanin biosynthesis as *VvUFGT*, and transport like *VvGST4* and *VvAM3*. Genes belonging to this secondary metabolic pathway were

expected to be involved as explaining the skin colour of these berries, which is the most evident phenotypic difference.

Table 3: Genes having a higher expression in red grapes than white grapes during the whole maturation phase. Expression intensity average is given for each colour group in FPKM at the end of véraison (EV) and harvest (H).

		Red grapes		White grapes	
Gene_ID	**Functional annotation**	**EV**	**H**	**EV**	**H**
VIT_00s0226g00100	Regulator of chromosome condensation (RCC1)	1.17	5.54	0.40	0.56
VIT_00s0287g00070, VIT_00s0287g00080	Vesicle-associated membrane protein	23.63	27.23	16.83	19.30
VIT_00s0577g00020	Phosphoglycerate mutase	38.92	48.67	29.58	32.61
VIT_00s0868g00010	DNA-binding storekeeper protein	71.50	92.70	54.56	69.76
VIT_01s0010g03510	Caffeoyl-CoA O-methyltransferase (AOMT)	199.92	29.37	0.09	0.02
VIT_01s0011g02760	No hit	2.66	7.95	1.00	3.45
VIT_01s0011g02770	Zinc finger CCHC domain-containing protein 8	5.23	17.54	2.12	7.91
VIT_01s0026g01870	Avr9/Cf-9 induced kinase 1	1.15	4.96	0.51	1.53
VIT_03s0017g01020	Methyl-CpG-binding domain 8 MBD08	17.09	17.94	12.56	12.85
VIT_03s0088g00260	Serine carboxypeptidase S10	17.06	3.30	0.28	0.21
VIT_03s0097g00470	ATHVA22A (Arabidopsis thaliana HVA22 homologue A)	7.10	8.65	2.56	2.26
VIT_04s0079g00690	Glutathione S-transferase (GST4)	1,428.47	621.73	0.15	0.18
VIT_05s0020g00030	WD-40 repeat	12.56	15.02	7.93	9.44
VIT_05s0051g00300	Disease resistance protein (TIR-NBS-LRR class)	3.59	4.08	1.70	1.62
VIT_05s0077g00800	DNA replication protein	5.34	7.36	3.43	3.79
VIT_05s0136g00260	Chalcone synthase (CHS3)	981.35	488.23	37.13	28.33
VIT_06s0004g02620	Phenylalanine ammonia-lyase	60.07	24.59	3.20	1.83
VIT_06s0004g08150	Trans-cinnamate 4-monooxygenase	90.20	266.53	8.76	15.01
VIT_06s0009g01610	Unknown protein	26.05	34.71	22.75	28.20
VIT_06s0009g02810	Flavonoid 3',5'-hydroxylase	43.92	15.86	0.13	0.02
VIT_06s0009g02830	Flavonoid 3',5'-hydroxylase	91.81	26.56	0.43	0.07
VIT_06s0009g02840	Flavonoid 3',5'-hydroxylase	45.95	14.66	0.09	0.03
VIT_06s0009g02860	Flavonoid 3',5'-hydroxylase (F3'5'Hf)	8.81	2.70	0.11	0.04
VIT_06s0009g02880	Flavonoid 3',5'-hydroxylase	38.17	9.03	0.02	0.02
VIT_06s0009g02920	Flavonoid 3',5'-hydroxylase	25.10	7.62	0.17	0.04
VIT_06s0009g02970	Flavonoid 3',5'-hydroxylase	41.54	11.15	0.05	0.02
VIT_06s0009g03000	Flavonoid 3',5'-hydroxylase	68.93	16.94	0.94	0.24
VIT_06s0009g03010	Flavonoid 3',5'-hydroxylase	74.40	20.70	1.28	0.25
VIT_06s0061g01410	TRNA isopentenyltransferase	9.77	13.07	6.22	9.15
VIT_07s0005g05040	ARR24 Type A	0.82	1.49	0.15	0.52
VIT_07s0031g01240	Unknown protein	4.83	6.33	3.72	4.42
VIT_08s0007g05430	Pyruvate kinase	2.77	4.20	1.17	1.30
VIT_08s0007g05490	Pyruvate kinase	9.87	13.39	3.41	3.75
VIT_08s0007g05880	Dehydration-induced protein (ERD15)	207.81	223.13	101.69	99.02
VIT_08s0007g06370	Photoperiod independent early flowering1 (PIE1)	6.15	7.40	5.11	5.80
VIT_08s0040g00990	Transcription elongation factor S-II	32.12	43.92	23.45	29.38
VIT_08s0040g02800	Lipin	8.74	14.42	6.17	7.47
VIT_08s0058g00880	Class E vacuolar protein-sorting machinery protein HSE1	51.13	84.86	38.49	53.32
VIT_09s0018g01660	Biopterin transport-related protein BT1	7.04	11.74	4.96	7.20
VIT_10s0003g00140	VvERF064_ERF/AP2 Gene Family	26.31	41.79	12.50	17.95
VIT_10s0003g01150	Autophagy 7 (APG7)	12.59	15.29	8.87	11.12
VIT_10s0003g01230	Serine esterase	17.06	19.22	12.70	14.80

(Table continues on following page)

Table 3 (Continued from previous page)

		Red grapes		White grapes	
Gene_ID	**Functional annotation**	**EV**	**H**	**EV**	**H**
VIT_10s0003g01390	Cupin, RmlC-type	24.13	51.08	16.38	28.98
VIT_10s0116g00860	Cardiolipin synthase	14.64	18.15	11.33	12.21
VIT_13s0067g00910	R protein MLA10	11.22	7.50	5.51	4.11
VIT_13s0067g03030	Unknown protein	20.53	28.23	16.44	20.70
VIT_13s0067g03820	Chalcone isomerase (CHI1)	183.13	266.39	41.07	59.16
VIT_13s0156g00030	No hit	1.45	3.47	0.63	1.41
VIT_14s0068g00920	Chalcone synthase (CHS2)	157.28	86.29	25.26	14.50
VIT_14s0068g02150	Kelch repeat-containing F-box family protein	21.19	121.94	6.93	18.41
VIT_16s0039g02040	4-coumarate-CoA ligase 3	12.95	20.56	2.65	2.14
VIT_16s0039g02230	UDP-glucose:flavonoid 3-O-glucosyltransferase	123.04	35.91	0.03	0.06
VIT_16s0050g00910	MATE efflux family protein AM3	250.96	115.21	2.96	2.43
VIT_16s0098g00190	Receptor kinase homolog LRK10	2.46	4.01	1.31	1.06
VIT_16s0098g01020	Splicing factor, arginine/serine-rich 1/9	29.17	45.67	21.97	31.08
VIT_17s0000g04360	Wall-associated kinase 2 (WAK2)	0.45	3.87	0.08	0.35
VIT_18s0001g05400	Unknown	6.79	14.50	5.53	8.18
VIT_18s0001g09400	Cytochrome b5 DIF-F (VvCytoB5)	295.75	76.84	3.24	3.71
VIT_19s0014g01550	DAG protein	50.63	71.58	36.46	54.36

Furthermore, the early responsive dehydration protein 15 gene *ERD15* was higher expressed in ripening red grapes. This gene is known to be a rapidly drought-responsive gene in *Arabidopsis thaliana*[44]. In their work, Kariola and colleagues[45] showed that alteration of *ERD15* expression modulated abscisic acid (ABA) responsiveness in Arabidopsis. ABA is believed to play an important role for the regulation of berry ripening, as the endogenous ABA concentrations in the berries increase dramatically at the onset of ripening[46,47]. Moreover, exogenous ABA application on pre-véraison berries hastened berry ripening[48]. ABA also plays a key role in plant adaptation to adverse environmental conditions, including drought. ABA accumulation controls the various downstream responses for the plant to acquire tolerance to these stresses[49-52]. In addition, endogenous ABA concentrations in the grape berries during ripening were influenced by stress treatments as water-deficit irrigation[53,54]. The increase of the ABA levels by the stress treatments was correlated to the progress of ripening, particularly, to the colouring in berry skins [54-57]. Indeed, exogenous ABA applications on the berries were reported to improve the berry colouring of both table grapes and wine grapes[58-63]. Finally, Deluc and colleagues[54] demonstrated that ABA concentrations were increased significantly by water deficit at véraison and one week after véraison in Cabernet Sauvignon (red grape), but not in Chardonnay berries (white grape). The rise in ABA concentrations by water deficit appeared to precede and enhance anthocyanin, hexose and proline accumulation in

Cabernet Sauvignon berries. Consequently, the greater expression of *ERD15* in red grapes suggests its involvement in the onset of ripening in red grapes including anthocyanin biosynthesis linked to a high sugar concentration (**chapter 3**). A vesicle-associated membrane protein (VAMP) gene was also identified. VAMP proteins belong to soluble N-ethylmaleimide sensitive factor attachment protein receptor (SNARE) protein family which are core factors in driving vesicle fusion with target membranes, which is critical in eukaryotes having distinct subcellular organelles[64]. The higher expression of a VAMP gene in red grapes suggests that vesicular transport seemed to be more active in these berries during maturation phase maybe to support flavonoid and anthocyanin transport.

Genes exhibiting a greater expression in white-skinned berries during maturation phase

Genes with a higher expression in white grapes at the end of véraison stage

At the end of véraison, 61 transcripts were classified as having a higher expression in white-skinned berries (**Table 4**). Among these genes, two cellulose synthase-like genes (VIT_02s0025g01750; VIT_09s0018g01950) and two polygalacturonase genes (VIT_02s0025g01330; VIT_05s0077g01740) were identified. Cellulose synthase-like (CSL) proteins contain sequence motifs that are characteristic of β-glycosyltransferases, but lack the N-terminal region containing the zinc-finger domains – which is found in cellulose synthase A (CESA) – that function in protein dimerization. Because of their similarity to, and divergence from CESA, CSL proteins are considered good candidates for the synthases that are localized in the Golgi and that form the β-D-glycan backbone of hemicelluloses such as xyloglucan, xylan, mannan and other β-D-glycans in the cell wall[65]. Consequently, the higher expression of these two cellulose synthase-like genes suggests a potential higher hemicellulose contents in white grape cell walls. In addition, polygalacturonases catalyse the hydrolysis of α-(1,4) glycosidic bonds between two unesterified galacturonic acid units. In fact, a higher polygalacturonans solubility was noticed by Nunan and co-workers[66] in cell wall of Muscat Gordo Blanco berry mesocarp (white grape) after véraison. That being said, we can hypothesize that white grape cell walls go through modifications that are not that remarkable in red grapes.

Table 4: Genes with a greater expression in white grapes than red grapes at the end of véraison. Adjusted p-value (Adj p-value) indicates the significance of the differential expression between red and white grapes. Expression intensity average is given for each colour group in FPKM and log_2-fold change (Log_2-FC) was calculated for each transcript.

Gene_ID	Adj p-value	Functional annotation	White grapes	Red grapes	Log_2-FC
VIT_02s0025g01750	7.8E-03	Cellulose synthase CSLG3	0.74	0.06	**12.0**
VIT_03s0017g02260	9.8E-04	Protease inhibitor/seed storage/lipid transfer protein	64.03	7.06	**9.1**
VIT_04s0008g01740	4.5E-03	Spermine synthase ACAULIS5	1.16	0.15	**7.8**
VIT_13s0067g02910	2.0E-03	Non-specific lipid-transfer protein	2.63	0.42	**6.3**
VIT_02s0025g00180	1.8E-03	Phosphoglycerate mutase	2.05	0.42	**4.9**
VIT_16s0098g00690	5.3E-03	Major Facilitator Superfamily	118.14	25.39	**4.7**
VIT_00s0214g00020	8.2E-05	TORMOZembryo defective UTP13	0.71	0.18	**3.9**
VIT_02s0025g03590	9.8E-03	Phospholipid hydroperoxide glutathione peroxidase	13.81	3.96	**3.5**
VIT_02s0154g00260	1.2E-03	Nitrate transporter	5.12	1.52	**3.4**
VIT_09s0018g01950	1.9E-03	Cellulose synthase CSLC05	3.61	1.11	**3.3**
VIT_01s0026g01910	4.0E-03	Myb domain protein 88	3.01	1.01	**3.0**
VIT_18s0001g14980	9.1E-03	3-methyl-2-oxobutanoate dehydrogenase	1.41	0.47	**3.0**
VIT_05s0077g01740	4.0E-03	Polygalacturonase GH28	8.08	2.83	**2.9**
VIT_02s0025g01260	7.3E-03	NADPH HC toxin reductase	6.54	2.42	**2.7**
VIT_02s0025g03010	2.4E-03	Copper chaperone (CCH)	5.66	2.13	**2.7**
VIT_11s0016g01850	3.0E-03	UDP-glucuronosyl/UDP-glucosyltransferase	1.66	0.63	**2.6**
VIT_06s0004g07330	6.8E-03	No hit	0.36	0.16	**2.3**
VIT_07s0129g00470	3.9E-03	Unknown protein	10.30	5.03	**2.0**
VIT_15s0046g01340	9.1E-03	Unknown protein	5.75	2.90	**2.0**
VIT_04s0008g05010	6.8E-03	Unknown protein	3,092.65	1,577.58	**2.0**
VIT_12s0028g00710	8.8E-03	4-hydroxyphenylpyruvate dioxygenase	147.40	75.54	**2.0**
VIT_06s0004g01160	2.5E-03	10-formyltetrahydrofolate synthetase	69.47	35.69	**1.9**
VIT_02s0025g01330	3.1E-03	Polygalacturonase GH28	104.65	53.96	**1.9**
VIT_08s0040g03230	7.0E-03	Unknown protein	20.87	10.80	**1.9**
VIT_08s0007g08540	5.8E-03	Mg-chelatase subunit XANTHA-F	37.71	19.60	**1.9**
VIT_03s0063g02020	3.2E-03	Tic Complex Tic62 Subunit	14.23	7.58	**1.9**
VIT_06s0004g02130, VIT_06s0004g02140	1.7E-03	MATE efflux family protein	197.36	106.94	**1.8**
VIT_07s0005g04400	5.9E-03	Photosystem II reaction centre W (PsbW)	28.36	15.50	**1.8**
VIT_16s0098g01280	8.4E-03	Rhodanese domain-containing protein	9.37	5.19	**1.8**
VIT_08s0007g00710	2.9E-03	Unknown	4.42	2.47	**1.8**
VIT_01s0011g04160	9.8E-03	Phosphatidic acid phosphatase alpha	29.72	16.68	**1.8**
VIT_09s0054g00790	2.0E-03	Brittle-1, chloroplast precursor	7.88	4.45	**1.8**
VIT_17s0000g03010	7.5E-03	Phosphatidyl serine synthase	2.97	1.70	**1.7**
VIT_04s0008g07020	2.8E-03	Translocon at the inner envelope membrane of chloroplasts 55	73.10	42.94	**1.7**
VIT_09s0002g07500	5.8E-03	Hydrolase, alpha/beta fold	3.82	2.27	**1.7**
VIT_14s0030g01340	4.2E-03	Beta-ketoacyl-CoA synthase	8.24	4.94	**1.7**
VIT_07s0129g00480	5.1E-03	Molecular chaperone DnaJ	7.48	4.52	**1.7**
VIT_12s0134g00510	1.4E-03	Cytochrome B6-F complex iron-sulfur subunit	52.63	31.90	**1.7**
VIT_09s0002g05180	2.4E-03	No hit	1,022.74	624.64	**1.6**
VIT_07s0005g03630	4.7E-03	Ankyrin	14.97	9.25	**1.6**
VIT_05s0029g00680	2.6E-03	Beta-D-galactosidase	8.57	5.31	**1.6**
VIT_07s0031g02780	3.4E-03	Cyclase/dehydrase	6.78	4.29	**1.6**
VIT_18s0122g00830	5.0E-03	PMI1 (plastid movement impaired1)	15.50	9.88	**1.6**
VIT_09s0054g01770	6.2E-03	Oxidoreductase/ transcriptional repressor	22.12	14.26	**1.6**
VIT_05s0020g02800	2.2E-03	Carbohydrate kinase	36.90	23.96	**1.5**
VIT_14s0060g01030	5.4E-03	Haloacid dehalogenase hydrolase	3.17	2.08	**1.5**
VIT_10s0042g00560	7.4E-03	Unknown protein	1.35	0.91	**1.5**
VIT_13s0019g01620	4.3E-03	No hit	5.85	3.97	**1.5**
VIT_12s0057g00020	6.3E-03	Protein thylakoid formation1	100.83	68.54	**1.5**
VIT_01s0127g00480	5.6E-03	Unknown protein	1.41	1.02	**1.4**
VIT_07s0031g02460	8.1E-03	Unknown protein	16.54	12.00	**1.4**
VIT_07s0005g04600, VIT_07s0005g04610	1.2E-03	Alcohol dehydrogenase class III	259.20	188.25	**1.4**

(Table continues on following page)

Table 4 (Continued from previous page)

Gene_ID	Adj p-value	Functional annotation	White grapes	Red grapes	Log_2-FC
VIT_03s0038g04080	7.0E-03	Glutamyl-tRNA synthetase	11.91	8.67	**1.4**
VIT_15s0048g00770	7.5E-03	Kinesin motor KHC	8.04	6.04	**1.3**
VIT_02s0025g02120	3.6E-03	1,4-alpha-D-glucan maltohydrolase	28.83	22.06	**1.3**
VIT_05s0077g00060	7.7E-03	D1 protease	8.10	6.28	**1.3**
VIT_11s0016g00760	9.5E-03	Ubiquinone biosynthesis protein COQ9	38.20	30.26	**1.3**
VIT_07s0005g00450	4.7E-03	Unknown protein	12.26	9.83	**1.2**
VIT_19s0135g00100	7.2E-03	Ribosomal protein L18A (RPL18aB) 60S	53.97	44.58	**1.2**
VIT_04s0044g00550	7.9E-03	Carbamoyl phosphate synthetase small subunit	48.56	40.19	**1.2**
VIT_10s0003g03720	4.4E-03	Unknown protein	52.87	44.14	**1.2**

Furthermore, a nitrate transporter gene exhibiting a higher expression in white grapes at the end of véraison was identified. According the current knowledge, there are two phases of intense nitrogen incorporation corresponding to the peaks of grape dry mass formation. The first takes place during the two weeks before the pea-sized berry stage, and the second starts one month later, at the onset of ripening, and lasts for another two weeks. Berries are able to accumulate about 50% of the final nitrogen mass during this period[67]. The higher RNA accumulation of this nitrate transporter transcript seemed to indicate a higher nitrogen uptake by white grapes rather than red grapes. This nitrogen supply might be used to produce volatile thiol precursors – S-cysteine conjugate non volatile precursors – mainly found in white grapes[68-70]. Nevertheless, this hypothesis needs to be confirmed by measurement of total nitrogen contents and S-cysteine conjugates at this growth stage.

<u>Genes with a greater expression in white grapes at harvest stage</u>

Among the 289 transcripts having a higher RNA accumulation in white grapes at harvest, only the photosynthesis GO category (GO:0015979) was found significantly overrepresented (p-value = 4.7996E-6). Within the category, the photosystem II stability/assembly factor gene *HCF136* (VIT_01s0011g02150), a photosystem II reaction centre W gene (VIT_07s0005g04400) and the protein thylakoid formation 1 gene *Thf1* (VIT_12s0057g00020) were identified. Photosystem II stability/assembly factor HCF136 was found to be essential for photosystem II (PSII) biogenesis and also required for assembly of an early intermediate in PSII assembly that includes D2 (psbD) and cytochrome b559 in *Arabidopsis thaliana*[71]. In addition, photosystem II reaction center W protein stabilizes dimeric PSII[72] and Thf1 gene product plays a crucial role in a dynamic process of vesicle-mediated thylakoid membrane biogenesis[73]. Consequently, the higher RNA accumulation of these three transcripts in white grapes

suggests that at full maturity some genes related to photosynthesis were not completely switched off yet compared to red-skinned berries.

Furthermore, the *PsbS* gene encoding the photosystem II 22 kDa protein was classified as having a higher expression in white grapes. Photosystem II 22 kDa protein, an intrinsic chlorophyll-binding protein of photosystem II, is necessary for nonphotochemical quenching but not for efficient light harvesting and photosynthesis in *Arabidopsis thaliana*[74]. Non-photochemical quenching is a process maintaining the balance between dissipation and utilization of light energy to minimize generation of oxidizing molecules, thereby protecting the plant against photo-oxidative damage. Hence, we can think that the *PsbS* gene, which was three times more expressed in white grapes, might be involved in the photo-protection of the white-skinned berries during ripening and then might offset the anthocyanin lack.

Regarding sugar transport, the hexose transporter *VvHT2* transcript was about three times more accumulated in white grapes at harvest in comparison to red grapes. Nonetheless, *VvHT2* expression is mainly associated with the increase of sugar levels occurring at véraison[75,76]. Consequently, this result suggests that *VvHT2* transcript accumulation stays high in white grapes during whole ripening phase compared to red grapes. Moreover, two terpene synthase genes *VvTPS55* and *VvTPS59* exhibited a greater expression in white-skinned berries at harvest stage. Terpenoids are aromatic metabolites responsible for wine flavours. *VvTPS55* and *VvTPS59* in particular belong to the VvTPS-g subfamily, as experimentally functional annotated by Martin *et al.*[77], whose members are thought to be mainly involved in the biosynthesis of aromatic compounds, such as linalool, nerolidol and geraniol. The higher expression of terpene synthases gene expression in white-skinned berries could explain some of the differences between red and white wine aromas.

Genes with a higher expression in white grapes during the whole maturation phase

In total, twenty-one transcripts were found to have a higher accumulation level at both end of véraison and harvest stages (**Table 5**).

Table 5: Genes exhibiting a higher expression in white grapes than red grapes during the whole maturation phase. Expression intensity average is given for each colour group in FPKM at the end of véraison (EV) and harvest (H).

		White-skinned berries		Red-skinned berries	
Gene_ID	**Functional annotation**	**EV**	**H**	**EV**	**H**
VIT_01s0026g01910	Myb domain protein 88	3.01	2.07	1.01	0.48
VIT_01s0127g00480	Unknown protein	1.41	1.78	1.02	1.16
VIT_02s0025g01260	NADPH HC toxin reductase	6.54	1.65	2.42	0.14
VIT_02s0025g02120	1,4-alpha-D-glucan maltohydrolase	28.83	24.03	22.06	16.67
VIT_03s0063g02020	Tic Complex Tic62 Subunit	14.23	5.97	7.58	1.65
VIT_05s0020g02800	Carbohydrate kinase	36.90	22.85	23.96	12.21
VIT_05s0077g00060	D1 protease	8.10	4.71	6.28	3.21
VIT_06s0004g02130, VIT_06s0004g02140	MATE efflux family protein	197.36	234.28	106.94	97.72
VIT_07s0005g04400	Photosystem II reaction centre W (PsbW)	28.36	12.92	15.50	2.97
VIT_07s0031g02460	Unknown protein	16.54	14.32	12.00	8.96
VIT_07s0031g02780	Cyclase/dehydrase	6.78	3.73	4.29	1.45
VIT_07s0129g00470	Unknown protein	10.30	4.68	5.03	1.37
VIT_07s0129g00480	Molecular chaperone DnaJ	7.48	4.17	4.52	2.06
VIT_08s0007g08540	Mg-chelatase subunit XANTHA-F	37.71	20.97	19.60	6.85
VIT_09s0002g05180	No hit	1,022.74	361.61	624.64	151.16
VIT_09s0002g07500	Hydrolase, alpha/beta fold	3.82	2.73	2.27	1.45
VIT_09s0054g01770	Oxidoreductase/ transcriptional repressor	22.12	14.73	14.26	7.37
VIT_12s0057g00020	Protein thylakoid formation1	100.83	69.47	68.54	34.16
VIT_13s0067g02910	Non-specific lipid-transfer protein	2.63	1.19	0.42	0.15
VIT_14s0030g01340	Beta-ketoacyl-CoA synthase	8.24	6.80	4.94	2.48
VIT_16s0098g00690	Major Facilitator Superfamily	118.14	49.84	25.39	7.79

Among these genes, genes related to the photosynthesis were identified, as a photosystem II reaction centre W gene (VIT_07s0005g04400), the protein thylakoid formation 1 gene *Thf1* (VIT_12s0057g00020) and the Mg-chelatase subunit f gene orthologous to the barley *Xantha-f* gene (VIT_08s0007g08540). Magnesium-protoporphyrin IX chelatase (Mg-chelatase) is located at the branchpoint of tetrapyrrole biosynthesis, at which point protoporphyrin IX is distributed for the synthesis of chlorophyll and heme. In barley (*Hordeum vulgare*), magnesium chelatase is a complex enzyme consisting of six 40-kD subunits, six 70-kD subunits[78], and an unknown number of 150-kD subunits encoded by the genes Xantha-h, -g, and -f respectively[79]. The XANTHA-F 150-kD subunit has been suggested to be the catalytic subunit as it binds the protoporphyrin IX substrate and most likely also the Mg^{2+} substrate[80-83]. The higher expression of these photosynthesis-related genes in white

grapes shows that this process seemed to be less slowed down in white- rather in red-skinned berries during ripening.

Furthermore, two transcripts involved in transport belonged to the higher-expressed genes in white grapes during the whole maturation phase. These two genes encode a MATE efflux family protein (VIT_06s0004g02130,VIT_06s0004g02140) and a major facilitator superfamily protein (VIT_16s0098g00690). Multidrug and toxic compound extrusion (MATE) family of transporters[84] and major facilitator superfamily (MFS) belong to the multidrug transporters with the ATP binding cassette (ABC) superfamily, the small multidrug resistance family and the resistance-nodulation-cell division family. The MATE family is characterized by the presence of twelve putative transmembrane segments and by the absence of “signature sequences” specific to the other multidrug transporter superfamilies. In higher plants, studies demonstrated that MATE proteins are mainly involved in the transport and trafficking of xenobiotic and small organic molecules[85], as secondary metabolites in vacuoles. In particular, MATEs from Medicago and barley have been characterized for mobilization of flavonoids in cell organelles[86,87]. Therefore, considering that this MATE efflux family protein gene was also expressed in red grapes, we can hypothesise that this gene could be involved in flavonoid transport, which are yellow pigments that contribute to the colour of white wines. On the other hand, the major facilitator superfamily (MFS) transporters are single-polypeptide secondary carriers capable only of transporting small solutes in response to chemiosmotic ion gradients. According to the current annotation, the Arabidopsis genome contains more than 120 genes predicted to encode members of the MFS transporter family[88,89], which have been traditionally classified according to the extensive sequence similarities they share with MFS transporters characterized in other organisms[90]. To date, the few plant MFS transporters characterized have been essentially implicated in sugar, oligopeptide, and nitrate transport[91,92]. In addition, an Arabidopsis MFS member, Zinc-Induced Facilitator 1 (ZIF1), has been described as a tonoplast-localized transporter involved in zinc tolerance[93], indicating that MFS transporters might also influence ion homeostasis in plants. Thus, this MFS gene could be related to many biological functions in grape berry. In particular, the higher expression of this gene in white grapes during the whole ripening phase suggests its involvement in either sugar or nitrate transport, which could be used for S-cysteine conjugate non volatile thiol precursors, as described above.

DEGs between red- and white-skinned berries during maturation phase are not entirely responsible for transcriptome differences

In order to observe the influence of the 1,249 DEGs between red- and white-skinned berries during maturation phase on the post-véraison transcriptome behaviours, a novel PCA was performed on the 40-sample data set after removing the 1,249 DEGs (**Fig. 4**).

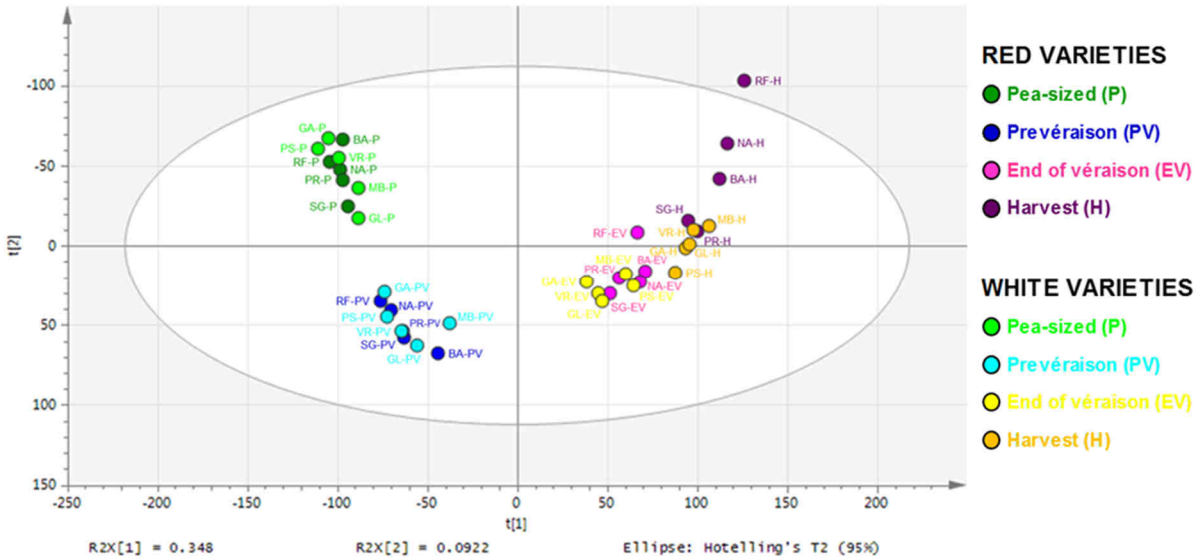


Figure 4: Plot of 40-sample data set PCA after removing the 1,249 DEGs between red- and white-skinned samples throughout maturation phase. The plot shows the first two principal components t[1]=34.8% and t[2]=9.2%.

Abbreviations for the four developmental stages: P (Pea-sized), PV (Prevéraison), EV (End of véraison), H (Harvest) and for each variety: SG (Sangiovese), BA (Barbera), NA (Negro amaro), RF (Refosco), PR (Primitivo) for the red-skinned varieties and VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina) for the white-skinned varieties.

In **Figure 4**, the maturation phase transcriptomes seem to get slightly closer to each others compared to the previous principal component analysis (**Fig. 1**). This result shows that the difference between red- and white-skinned-berry transcriptome behaviours was only marginally due to the differential expression of the 1,249 genes identified in the previous section. Hence, those genes were not exclusively responsible for the separation between red-skinned berry transcriptomes at harvest stage and further investigations were necessary.

Anthocyanin content seems to influence transcriptome behaviours of red grape varieties at harvest stage

Transcripts involved in the sample separation at harvest

In order to delve into the relationship between samples at harvest stage, a PCA was performed only on the ten full-ripe berry samples (**Fig. 5**). A unique Principal Component (PC) separated the three red-coloured Barbera, Negro amaro and Refosco berry transcriptomes from the others.

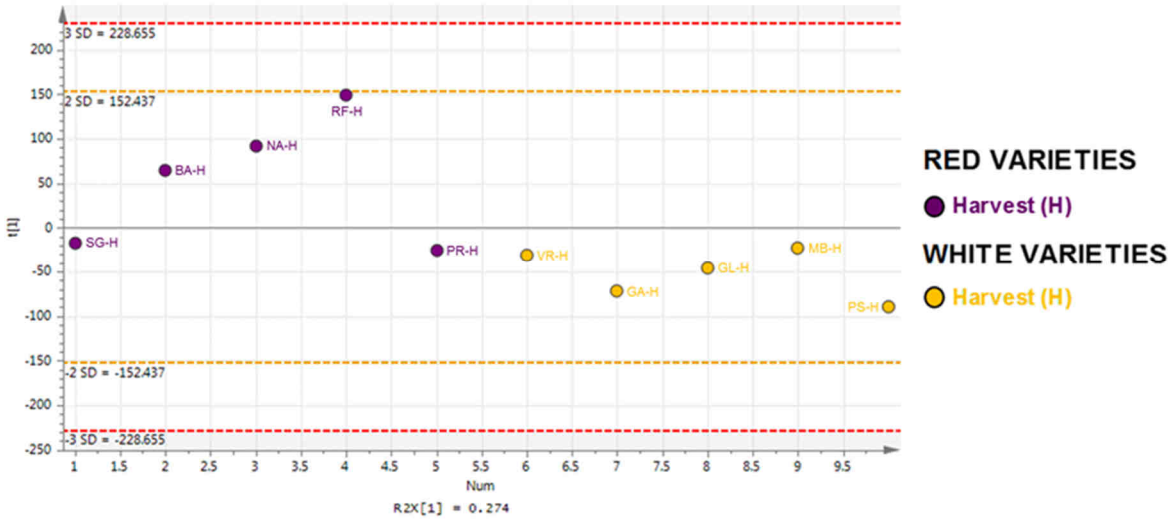


Figure 5: Plot of 10-harvest-sample data set PCA showing one principal component t[1]=27.4%. The dataset used for this analysis was composed of 21,697 genes resulting from the 40-sample dataset PCA.

Abbreviations for each variety: SG (Sangiovese), BA (Barbera), NA (Negro amaro), RF (Refosco), PR (Primitivo) for the red-skinned varieties and VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina) for the white-skinned varieties.

In order to identify the genes involved in the 27.4% of variability described by the unique principal component, a threshold for the PCA loadings was evaluated by removing step by step genes with 0.5-decreasing loading values until reaching the breaking point at which the PCA observed in **Figure 5** was no longer reproducible. After several tests, the desired threshold seemed to be 0.65/-0.65 corresponding to 6,003 removed genes: 3,450 and 2,553 genes with positive and negative loadings respectively, i.e. positive and negative correlation to the principal component in analysis. These 6,003 genes were then removed from the complete data set and the resulting PCA is shown in **Figure 6b**. The exclusion of the genes generated a tight

clusterization of post-véraison transcriptomes, whereas pre-véraison samples seemed to be only slightly influenced, confirming that these 6,003 genes were mostly involved in the maturation process and not important in pre-véraison stages for all varieties.

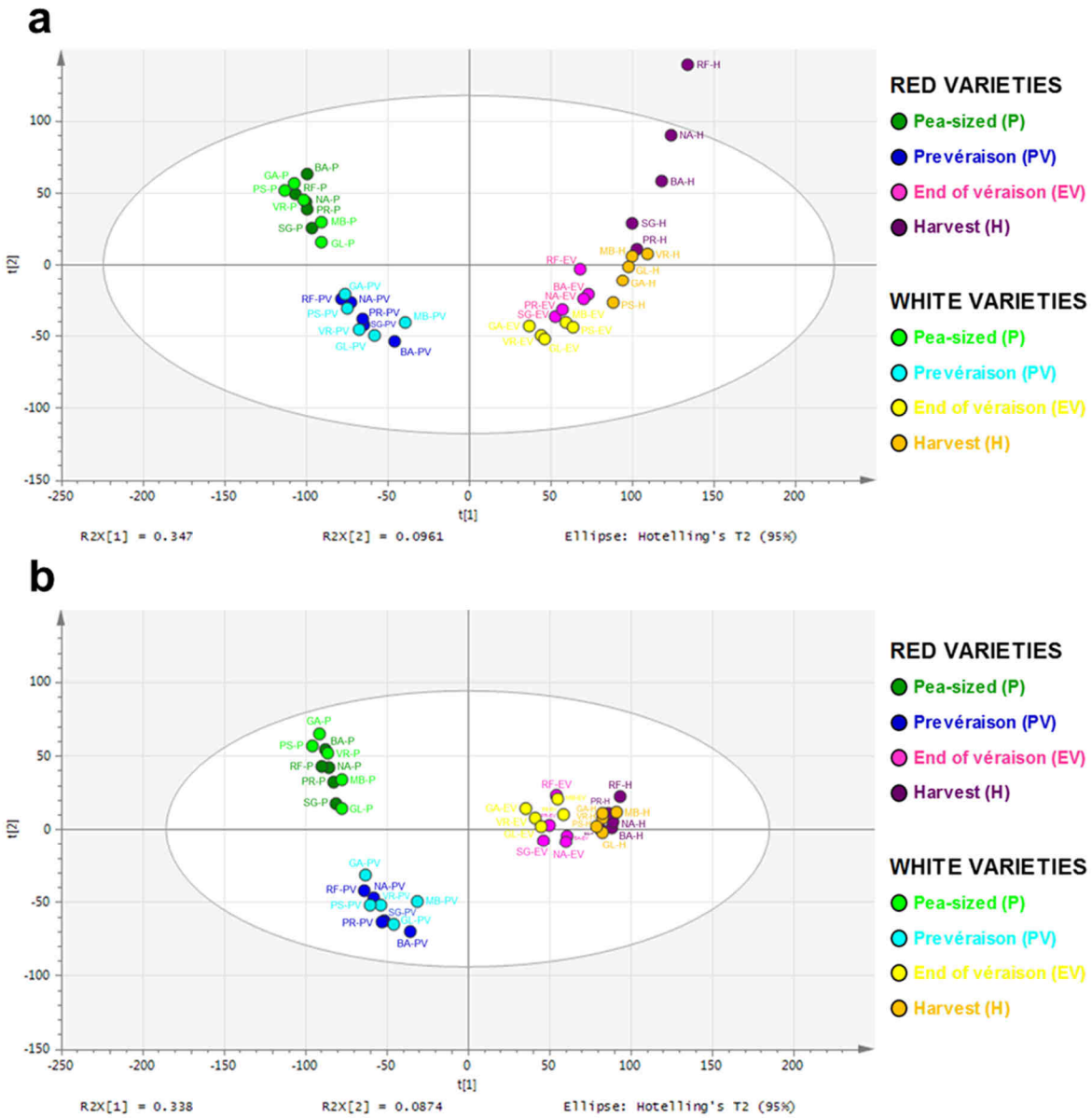


Figure 6: (a) Plot of 40-sample dataset PCA showing the first two principal components t[1]=34.7% and t[2]=9.61%.

(b) Plot of 40-sample dataset PCA after removing genes with harvest loadings higher than 0.65 and lower than -0.65. The plot shows the first two principal component t[1]=33.8% and t[2]=8.74%. The dataset used for this analysis was composed of 15,228 genes.

Abbreviations for the four developmental stages: P (Pea-sized), PV (Prevéraison), EV (End of véraison), H (Harvest) and for each variety: SG (Sangiovese), BA (Barbera), NA (Negro amaro), RF (Refosco), PR (Primitivo) for the red-skinned varieties and VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina) for the white-skinned varieties.

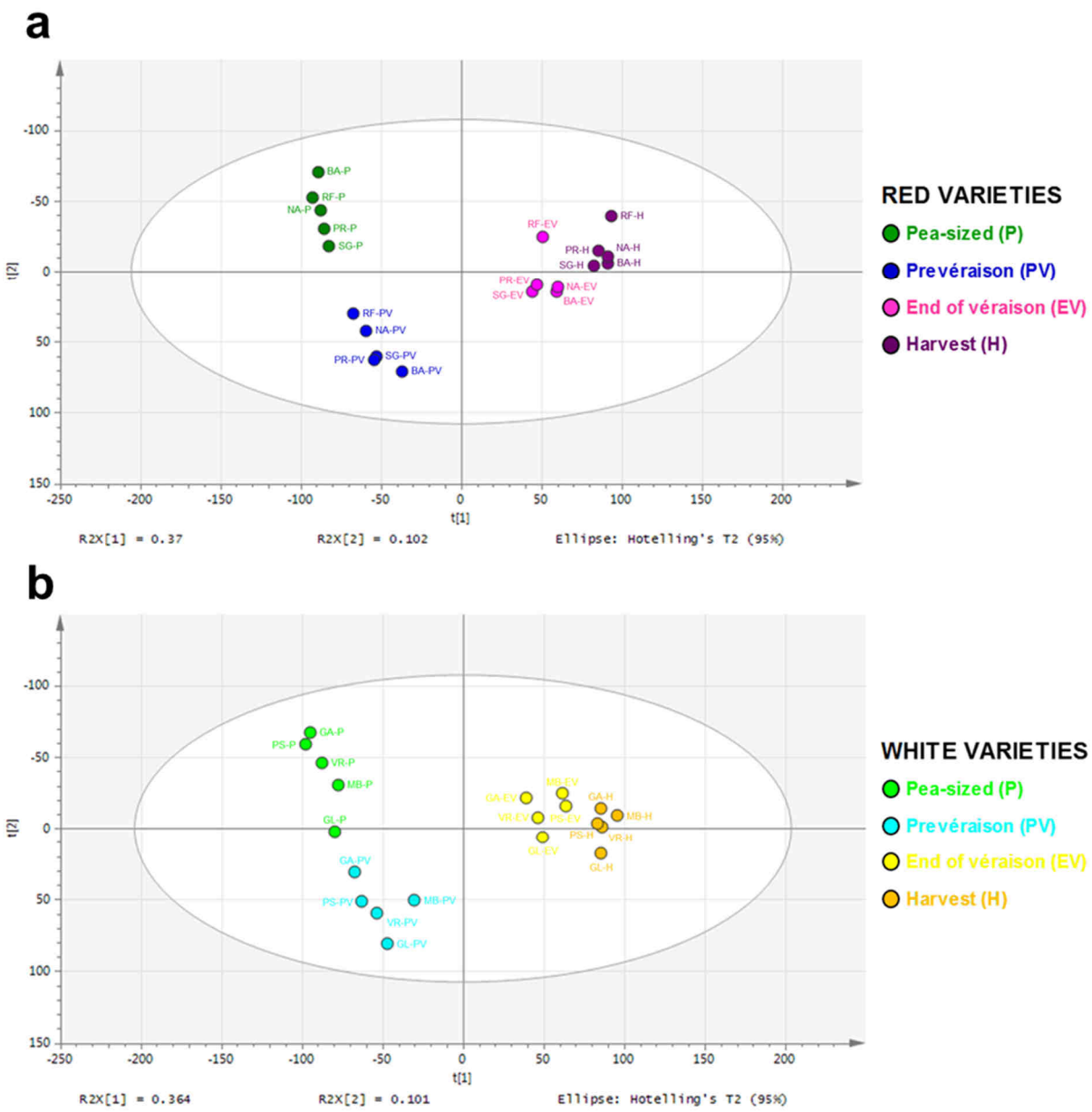


Figure 7: (a) Plot of 20-red-skinned-sample data set PCA after removing genes with harvest loading values higher than 0.65 and lower than -0.65. The plot shows the first two principal components t[1]=37.0% and t[2]=10.2%.

(b) Plot of 20-WHITE-sample dataset PCA after removing genes with harvest loading values higher than 0.65 and lower than -0.65. The plot shows the first two principal components t[1]=36.4% and t[2]=10.1%. The datasets used for (a) and (b) analyses were composed of 15,228 genes corresponding to the remaining genes after removing genes with harvest loadings higher than 0.65 and lower than -0.65.

Abbreviations for the four developmental stages: P (Pea-sized), PV (Prevéraison), EV (End of véraison), H (Harvest) and for each variety: SG (Sangiovese), BA (Barbera), NA (Negro amaro), RF (Refosco), PR (Primitivo) for the red-skinned varieties and VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina) for the white-skinned varieties.

In order to determine whether the grouping observed in **Figure 6b** was confirmed also for the red-coloured transcriptomes during maturation phase alone, the 6,003 genes were removed from both the 20-red- and the 20-white-sample data sets. The resulting data sets were analysed in two separated PCAs (**Fig. 7**). In **Figure 7a**, the removing of the 6,003 genes allows clustering the post-véraison transcriptomes of red-skinned berries (**Fig. 7a**). Moreover, the gene exclusion do not affect the behaviour of white-coloured berry transcriptomes for which the plot in **Figure 7b** seems almost identical to the previous white-grape PCA (see **Fig. 2b**). These observations suggest that those 6,003 genes were specifically involved in the maturation process of red-coloured berries. Furthermore, the extracted genes seemed to explain also the differences within the red grapes group – Barbera, Negro amaro and Refosco were the most separated from the other harvest stage samples – suggesting that they might also be related to specific varietal traits of those grapevine cultivars

Genes exhibiting a greater expression in Barbera, Negro amaro and Refosco full-ripe berries

Our analysis allowed the identification of 3,450 transcripts having a loading value greater than 0.65, corresponding to genes more expressed in Barbera, Negro amaro and Refosco than the other varieties at harvest. In order to gain insights into biological processes involved in this difference, the genes were analysed for overrepresented functions using the GO enrichment tool BiNGO[15] and the GoSlim Plants annotation (**Fig. 8**). Five GO categories were found significantly overrepresented: secondary metabolic process (GO:0019748), transcription (GO:0006350), cellular process (GO:0009987), catabolic process (GO:0009056) and protein modification process (GO:0006464).

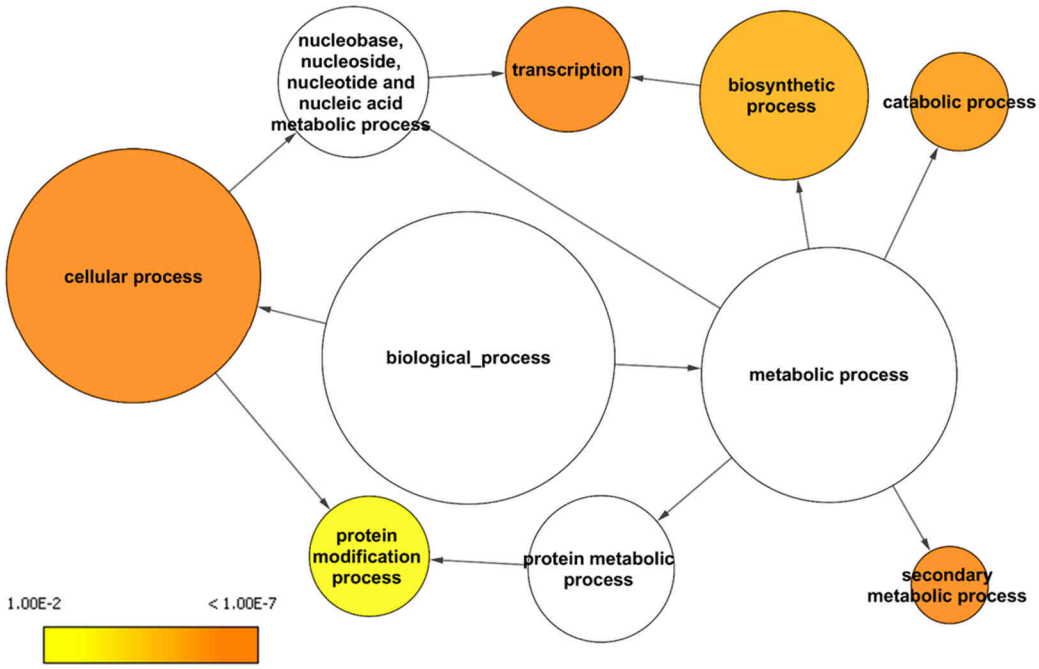


Figure 8: Enriched GO terms in removed positive-loading genes at harvest.
The network graphs show BiNGO visualizations of the overrepresented GO terms for the 3,450 transcripts having loadings higher than 0.65. Node size is positively correlated with the number of genes belonging to the category. Non-coloured nodes are not overrepresented, but they may be the parents of overrepresented terms. Coloured nodes represent GO terms that are significantly overrepresented (Benjamini and Hochberg corrected p-value < 0.01), with the shade indicating significance as shown in the colour bar.

Regarding the overrepresented secondary metabolic process GO category, most of the genes were related to phenylpropanoid and stilbene biosynthesis, as fifteen phenylalanine ammonia-lyase genes, four 4-coumarate-CoA ligase genes, three chalcone synthase genes, the flavonone-3-hydroxylase gene *VvF3H1*, and thirty-seven stilbene synthase genes (**Supplementary table 3**). Flavonoids differ both in quantity and in composition among grape varieties; in particular, total anthocyanin content has been found to be much higher in Barbera, Negro amaro and Refosco compared to Primitivo and especially Sangiovese[94] – about 4-time reduced in anthocyanin content with respect to Negro amaro – and the pattern of expression of the genes involved in flavonoid biosynthesis very likely mirror that result. It seemed indeed that the entire phenylpropanoid pathway – especially the earliest steps – was more transcriptionally active in the three red varieties. Stilbenes are constitutively synthesised in healthy grape berries from véraison to ripening[95]. Nevertheless,

significant differences in final content exist among varieties, in particular resveratrol content was found to be smaller in Refosco and Primitivo with respect to Barbera and Negro amaro[96]. Moreover, stilbene synthesis also increases upon pathogen infection and in response to abiotic stresses[95,97]. We also identified two cinnamyl alcohol dehydrogenase genes, two sinapyl alcohol dehydrogenase genes and twenty-two laccase genes involved in lignin biosynthesis and cell wall lignification[98], suggesting the formation of secondary cell wall in the three red-skinned pericarps at maturity maybe increasing pedicel-rachis attachment site thickness or causing thicker skins in these three varieties. Furthermore, two ascorbate oxidase genes were found having a higher expression in Barbera, Negroamaro and Refosco full-ripe berries. Our results suggest either that the three red grapes underwent more abiotic stresses (light, temperature…) than the other varieties at harvest, or that ascorbate oxidation was preferentially used to avoid ROS damage in these three red grapes.

Among the significantly overrepresented transcription GO category, 311 transcription factor activity-related genes were identified. Among them, twenty ethylene response factor (ERF) genes were found. Although ERFs are assumed to play a role in fruit development during the ripening process in several species[99-101], they also function as important plant-specific transcription factors in regulating biotic and abiotic stress response through interaction with various stress pathways[102-107]. Regarding the growth stage, these ERF genes might be involved in response to abiotic or biotic stimulus, suggesting that Barbera, Negro amaro and Refosco full-ripe berries underwent more environmental stresses than the other varieties. However, the expression of these twenty ERF genes also indicates that ethylene may play a more crucial role in grape, which is a nonclimacteric fruit, than previously thought, particularly in the late stages of ripening. Furthermore, seven jasmonate ZIM-domain (JAZ/TIFY) genes were also identified. Jasmonate-dependent gene activation involves hormone-induced degradation of a transcriptional repressor, the jasmonate ZIM/tify-domain (JAZ/TIFY) proteins[108]. The products of these JAZ/TIFY genes share two conserved domains, a ZIM/tify (zinc-finger protein expressed in inflorescence meristem) and a Jas (jasmonate-associated) domain. The ZIM/tify domain mediates homo- and heteromeric hormone-independent interactions between individual JAZ/TIFY proteins[109]. In contrast, the Jas domain is required for hormone-dependent interactions of JAZ/TIFY with MYC2 and COI1[110,111] and for nuclear localization[112]. In response to environmental or developmental signals that stimulate the biosynthesis of jasmonic

acids, the elevated levels of jasmonate-Ile promote the interaction of JAZ/TIFY repressors with SCF CoI1-mediated ubiquitination and subsequently degrade JAZ/TIFY proteins via the 26S proteasome, which in turn releases MYC2 and probably other transcription factors. Activation of MYC2 induces transcription of early JA-responsive genes including JAZ/TIFY genes themselves. JAs are involved in plant responses to several abiotic and biotic stresses such as wounding (mechanical stress), drought stress, salt stress, ozone, pathogen infection, and insect attack[110,111,113,114]. Consequently, the higher expression of seven JAZ/TIFY genes in Barbera, Negro amaro and Refosco full-ripe berries indicates that these berries underwent more environmental stresses. In addition, Vezzulli and co-workers[115] sprayed with methyl-jasmonate solution at fruit set, at véraison, and at ripening clusters of *Vitis vinifera* cv Barbera. Cumulative methyl-jasmonate treatments significantly increased berry resveratrol and viniferin at ripening. Therefore, the seven JAZ/TIFY genes might be involved in the higher expression of the 37 stilbene synthase genes above-described.

Concerning the cellular process GO category, genes involved in the regulation of the cytoskeleton were identified, as genes encoding dynein and kinesin proteins which are motor proteins. This result suggests a more intensive transport of cellular cargo inside the cells of Barbera, Negro amaro and Refosco berries at harvest.

Finally, among the protein modification process GO category, we noticed the higher RNA accumulation of many kinases mostly involved in signal transduction in the three red grapes.

<u>Genes exhibiting a lower expression in Barbera, Negroamaro and Refosco full-ripe berries</u>

Regarding the 2,553 genes having loading values below the threshold -0.65 and exhibiting a lower expression in the three red-coloured varieties Barbera, Negro amaro and Refosco, the photosynthesis (GO:0015979), generation of precursor metabolites and energy (GO:0006091), translation (GO:0006412) and cellular process (GO:0009987) were identified as significantly over-represented categories using GO enrichment tool BiNGO[15] and the GoSlim Plants annotation (**Fig. 9**).

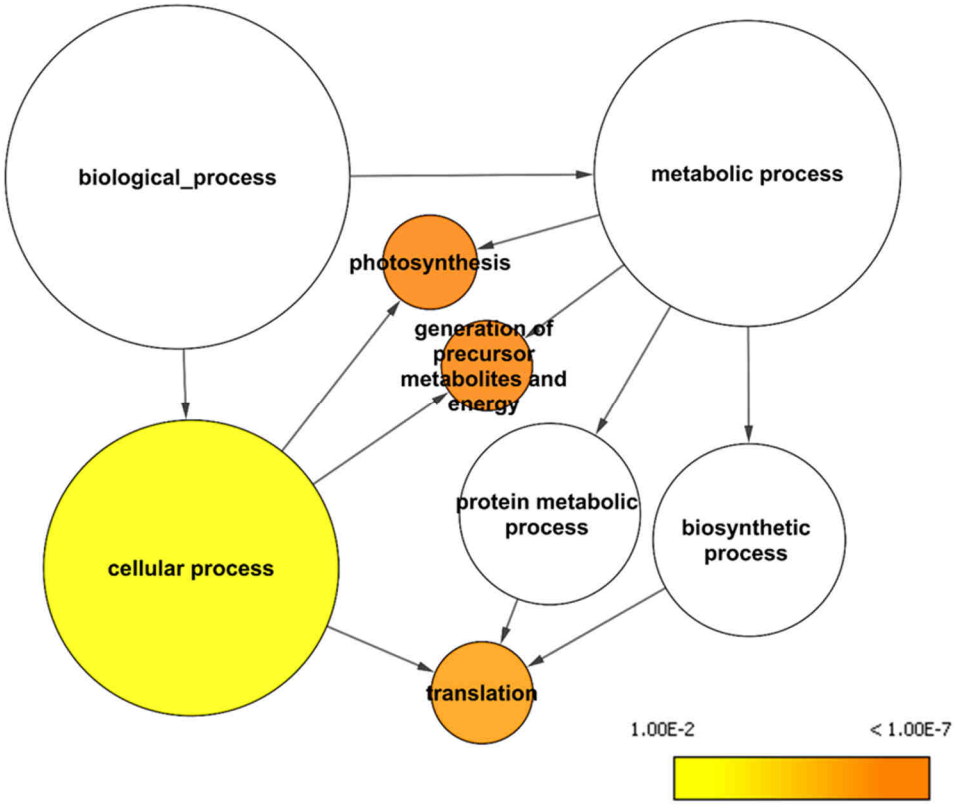


Figure 9: Enriched GO terms in removed negative-loading genes at harvest.
The network graphs show BiNGO visualizations of the overrepresented GO terms for the 2,553 transcripts having loadings lower than -0.65. Node size is positively correlated with the number of genes belonging to the category. Non-coloured nodes are not overrepresented, but they may be the parents of overrepresented terms. Coloured nodes represent GO terms that are significantly overrepresented (Benjamini and Hochberg corrected p-value < 0.01), with the shade indicating significance as shown in the colour bar.

The overrepresentation of the photosynthesis GO category suggests that photosynthesis shutdown in Barbera, Negro amaro and Refosco full-ripe berries was more significant compared to the other varieties. The higher level of the photosynthesis-related genes in white-skinned berries has been already revealed by the t-test analysis between full-ripe red and white grape trancriptomes described above. This new result suggests that photosynthesis was more active in white grapes but also in Sangiovese and Primitivo berries in comparison to the three other red grapes. Nevertheless, we can also wonder if all the collected full-ripe berries were at the same maturity state.

Among the overrepresented translation GO category, fifty-seven genes encoding ribosomal proteins were found. Ribosomal proteins are the proteins that make up the ribosomal subunits in conjunction with ribosomal RNA. Snoep and co-

workers[116] presented a core model for protein production in *Escherichia coli* to know whether there was an optimal ribosomal concentration for non-ribosome protein production. Analysing the steady-state solution of the model over a range of mRNA concentrations indicated that such an optimum ribosomal content existed and that the optimum shifts to higher ribosomal contents at higher specific growth rates. Unfortunately, such model was not proposed for plants. Therefore, we cannot know if protein production was less active in Barbera, Negro amaro and Refosco full-ripe berries in comparison to the other varieties or inversely.

Finally, three hexose transporter genes were identified among the overrepresented cellular process GO category: *VvHT2* (VIT_18s0001g05570), *VvHT3/HT7* (VIT_11s0149g00050) and *VvHT6/TMT2* (VIT_18s0122g00850) encoding a tonoplast monosaccharide transporter. All these hexose transporters are localized on the tonoplast[117]. *VvHT3/HT7* is homologous to *AtSTP7*, which has not been functionally characterized. VvHT3 is not able to transport any of the tested radiolabelled sugars in the deficient yeast model[24], and its transport function is also unknown. *VvHT6/TMT2* transcripts are highly accumulated at véraison[76,118], suggesting that this transporter may be responsible for vacuolar accumulation of hexose at the inception of ripening. *VvHT2*, whose expression is mainly associated with véraison[24,76], is homologous to *AtSTP5*, but its transport activity has not yet been characterized. The higher expression of these hexose transporter genes suggests faster sugar storage into the vacuole in white-skinned, Sangiovese and Primitivo full-ripe berries. Nonetheless, Barbera, Negro amaro and Refosco full-ripe berries showed higher Brix degree values than the other varieties, suggesting an upstream regulation as the generation of hexoses by acid invertases and/or sucrose synthases.

Anthocyanin content seems to influence transcriptome behaviours of red grape varieties at harvest stage

The expression of flavonoid-related genes was found to play a role in the difference between Barbera, Negro amaro and Refosco, and the other red varieties Sangiovese and Primitivo. In order to confirm this evidence, total anthocyanin concentration in the red berries was measured using spectrophotometry (520 nm) (**Fig. 10**).

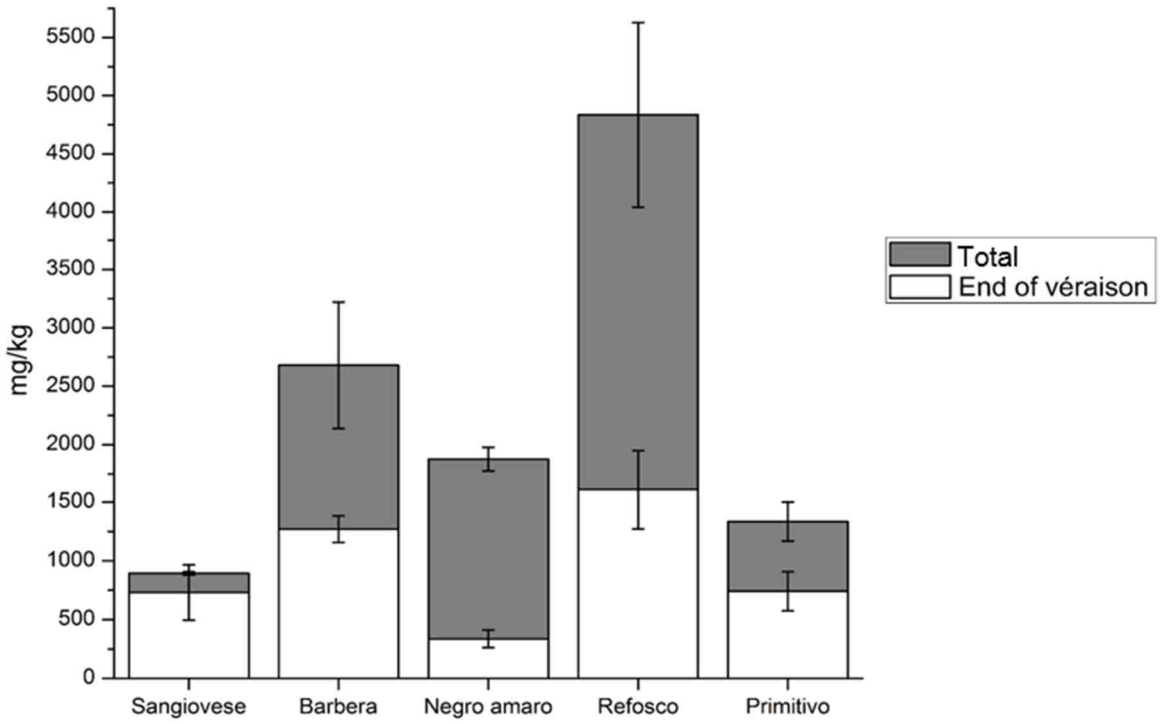


Figure 10: Anthocyanin contents of the red-coloured varieties at the end of véraison and harvest (Total). Values represent the mean ± standard deviation of three biological replicates.

As expected Refosco, Barbera and Negroamaro berries at harvest contained more anthocyanins than Sangiovese and Primitivo, supporting the hypothesis that the distribution observed in the 40-sample PCA plot (**Fig. 1**) was influenced by the difference transcriptional activity related to a different total anthocyanin content. Nevertheless, Negro amaro has usually a higher content in anthocyanins – 3.15 g/kg according Mattivi *et al.*[94] – suggesting that the environment of Conegliano Veneto (North East of Italy), where the sample were collected, might not be perfect for this variety, which is mainly cultivated in Southern Italy (**chapter 2**). Moreover, Negro amaro berries had the smallest amount of anthocyanin at the end of véraison (~336 mg/kg), indicating that anthocyanin biosynthesis in Negro amaro berries mostly occurred later right before harvest. Therefore, we can hypothesize that ripening berry transcriptome behaviours could be affected either by berry anthocyanin content at harvest or by the level of anthocyanin production between the end of véraison and harvest stages. However, further metabolomic investigations should be necessary to confirm this hypothesis but also to understand which of anthocyanin content or composition predominantly influenced transcriptome behaviours throughout ripening phase.

CONCLUSION

Performing two distinct PCA analysis with either the 20 red-coloured samples or the 20 white-coloured samples indicated that red- and white-skinned berry transcriptomes described different patterns during maturation phase. In fact, a post-véraison transcriptome grouping was observed for the white-coloured berry transcriptomes whereas red-skinned berry transcriptomes were split like on the plot of the 40-sample data set PCA. In order to find out genes involved in the difference between red- and white-skinned berry transcriptome behaviours during the maturation phase, t-test analyses were performed between the five red- and the five white-skinned berry transcriptomes at the two ripening stages. Nevertheless, removing of these differentially expressed genes was not enough to explain red grape transcriptome patterns. Then, a novel analysis, consisting in the extraction of transcript loadings from PCA outcome and the detection of the loading threshold-breaking point, allowed identifying the 6,003 transcripts specifically involved in the maturation process of Barbera, Negro amaro and Refosco red-coloured berries. After doing anthocyanin content measurements, these three red grapes were found to have a high anthocyanin concentration at full-maturity as Mattivi and colleagues[94]. Consequently, we hypothesized that red grape transcriptome could be influenced by anthocyanin accumulation in the skin, forming a dark screen on berry surface.

Along the analysis of the transcripts specifically involved in the maturation process of Barbera, Negro amaro and Refosco berries, a higher expression level of photosynthesis-related genes was observed in white grapes suggesting that photosynthesis was still slightly active in these grapes. In addition, we noticed that a higher accumulation of photosynthesis-related transcripts in full-ripe Sangiovese and Primitivo berries in comparison to the other red grapes (Barbera, Negro amaro and Refosco), suggesting that anthocyanin content could influence the shutdown of this process. In fact, the anthocyanin accumulation in the skin, forming a dark screen, could diminish the photosynthesis process in the pulp and this effect could be correlated to the anthocyanin content.

Furthermore, ripening red grapes expressed preferentially the hexose transporter gene *VvHT5* and the cell wall invertase gene *VvINV4* whereas *VvHT2* exhibited a higher expression in white grapes. In addition, Barbera, Negro amaro and Refosco berries accumulated less *VvHT2*, *VvHT3/HT7* and *VvHT6/TMT2* transcripts

than the other varieties. Cell wall invertases are of particular importance when considering the physiological roles of hexose transporters because cell wall invertase activity generates apoplastic hexoses, the substrate of hexose transporters, which may be transported into heterotrophic cells via hexose transporter activity. Therefore, although Barbera, Negro amaro and Refosco full-ripe berries expressed slightly less hexose transporter genes, their high sugar content could be explained by a greater expression of *VvINV4*. Moreover, influence of sugar supply on anthocyanin accumulation has been demonstrated by several studies[119-122] but little is known about the effect of anthocyanin production on sugar accumulation. In this work, anthocyanin production and sugar accumulation genes seemed to be co-regulated at harvest in Barbera, Negro amaro and Refosco berries, and thus influencing their maturity.

Moreover, many abiotic and biotic stress-related genes had a greater expression in red grapes and more particularly in Barbera, Negro amaro and Refosco berries compared to white grapes. Considering that Barbera, Negro amaro and Refosco berries did not reach full-maturity at the same time - Barbera full-ripe berries were collected early September (**chapter 3**) as the white-skinned berries, and Negro amaro and Refosco berries reached full-maturity two weeks later as Sangiovese and Primitivo – they did not undergo the same environmental stresses at the end of ripening. Therefore, we can wonder whether this phenomenon was related to anthocyanin accumulation. In addition, even if stilbene synthesis is genotype-dependent, it could be also induced by biotic[123-127] and abiotic stresses[128-130]. Thus, we can wonder whether biotic and abiotic stresses have induced expression level of stilbene synthase genes in Barbera, Negro amaro and Refosco berries.

Finally, this deep analysis showed that phenylpropanoid/flavonoid biosynthetic pathway-related gene expression is not sufficient to explain the differences between red- and white-grape transcriptomes during maturation phase, however products of this biosynthetic pathway (anthocyanins) could influence other biological processes. Moreover, further metabolomic investigations should be necessary to confirm our hypothesis and may also provide indications at which extent anthocyanin content and composition influence transcriptome throughout grape berry maturation phase.

METHODS

PCA and Loading normalization

Principal component analyses (PCAs) were performed using SIMCA P+ 13 software (Umetrics, USA).

In the case of the Harvest stage, the PCA was performed on 21,746-gene data set. Loadings, which represent new values of the genes in the model plane, were extracted and normalized by applying the following formula:

$$loading\ value \times \sqrt{PC\ variance \times N}$$

with N as the total number of genes involved in the PCA, i.e. 21,231 genes in this case. In fact, 515 genes were excluded from the analysis when less than two samples have a FPKM value different from the median 0.

Loading threshold-breaking point, i.e. the loading threshold avoiding the PCA to be processed after the removing of the genes with a higher or lower normalized loading value, was obtained by performing successive 0.5-loading-threshold decreases.

T-test analysis

At end of véraison and harvest stages, red- and white-skinned berry trancriptomes were compared using between-subjects t-test using 0.01 level to calculate the significance (TMeV 4.3; http://www.tm4.org/mev). In order to perform this analysis correctly, FPKM values were log-transformed previously to obtain normal-distributed data.

GO Enrichment Analysis

GO enrichment analysis was performed by using BiNGO 2.4 plug-in tool in Cytoscape version 3.1.0 with PlantGOslim categories, as described by Maere *et al.*[15]. Overrepresented PlantGOslim categories were identified using a hypergeometric test with a significance threshold of 0.01 after a Benjamini & Hochberg False-Discovery Rate (FDR) correction[131].

Anthocyanin quantification

The same post-véraison powdered samples used for RNA extraction (**chapter 3**) were extracted in four volumes (w/v) of methanol acidified with 1 % of chloride hydroxide (HCL) (v/v) in an ultrasonic bath at room temperature and 40 kHz for 20 minutes. After

a centrifugation (13500 rpm, 4°C), absorbance at 520 nm of the supernatant was measured by the spectrophotometer GeneQuant 100 (GE Healthcare Life-Sciences, Buckimghamshire, UK). Anthocyanin concentration was obtained by dividing the absorbance measurement by the coefficient of regression 0.0283 acquired by standard scale measurements.

SUPPLEMENTARY MATERIAL

Supplementary table 1: 24 transcription factor genes exhibiting a higher expression in red grapes than white grapes at the end of véraison. Adjusted p-value (Adj p-value) indicates the significance of the differential expression between red and white grapes. Expression intensity average is given for each colour group in FPKM and log_2-fold change (Log_2-FC) was calculated for each transcript.

Gene_ID	Adj p-value	Functional annotation	Red grapes	White grapes	Log_2-FC
VIT_02s0033g00410	2.4E-03	Myb VvMYBA1	71.21	9.86	**2.85**
VIT_02s0025g00420	5.6E-04	WRKY DNA-binding protein 27	2.58	1.13	**1.19**
VIT_01s0026g01550	6.4E-03	Homeodomain leucine zipper protein HB-1	14.36	7.44	**0.95**
VIT_19s0014g03290	4.5E-03	VvNAC17_NAC domain-containing protein	146.52	97.85	**0.58**
VIT_11s0037g00150	5.7E-03	WRKY DNA-binding protein 32	33.61	23.17	**0.54**
VIT_10s0003g00290	8.5E-03	Homeobox-1	17.76	12.60	**0.50**
VIT_10s0003g00270	3.8E-03	Homeobox-1	24.24	17.30	**0.49**
VIT_08s0040g00990	1.1E-03	Transcription elongation factor S-II	32.12	23.45	**0.45**
VIT_06s0009g03530	1.5E-04	Homeodomain protein 14	10.24	7.61	**0.43**
VIT_12s0055g00420	4.3E-04	BZIP transcription factor	22.62	17.00	**0.41**
VIT_08s0040g00610	9.0E-03	GTB1 (global transcription factor group B1)	56.87	42.93	**0.41**
VIT_00s0475g00040	8.0E-03	Myb family	20.02	15.18	**0.40**
VIT_12s0028g03940	7.4E-03	Transcription initiation factor TFIID subunit 1-A	15.90	12.59	**0.34**
VIT_03s0038g00610	2.5E-03	PHD finger transcription factor	29.58	23.47	**0.33**
VIT_18s0076g00330	3.1E-03	VirE2-interacting protein (VIP1)	55.02	43.67	**0.33**
VIT_14s0108g00040	1.9E-03	Structure-specific recognition protein 1 SSRP1	91.02	72.49	**0.33**
VIT_07s0104g01050	1.6E-03	Homeobox protein	9.26	7.40	**0.32**
VIT_14s0036g00940	8.2E-05	Transcription initiation factor TFIID subunit 10	30.14	24.48	**0.30**
VIT_18s0122g00380	9.4E-03	VvSBP17_Squamosa promoter-binding protein	43.11	35.15	**0.29**
VIT_17s0000g04350	5.4E-03	TATA-binding protein-associated phosphoprotein Dr1 protein	54.02	45.31	**0.25**
VIT_17s0000g01500	3.2E-04	SWIB complex BAF60b domain-containing protein	10.76	9.27	**0.22**
VIT_07s0005g00200	5.4E-03	AAA-type ATPase	49.00	42.20	**0.22**
VIT_17s0000g01860	7.7E-03	BPC2/BBR/BPC2/BPC2 (basic pentacysteine2)	25.24	22.31	**0.18**
VIT_18s0001g07730	2.3E-03	VvZML4_TIFY gene family	24.88	22.37	**0.15**

Supplementary table 2: 47 transcription factor genes having a higher expression in red grapes than white grapes at harvest. Adjusted p-value (Adj p-value) indicates the significance of the differential expression between red and white grapes. Expression intensity average is given for each colour group in FPKM and log_2-fold change (Log_2-FC) was calculated for each transcript.

Gene_ID	Adj p-value	Functional annotation	Red grapes	White grapes	Log_2-FC
VIT_10s0597g00050	5.0E-03	Heat shock transcription factor B2A	3.55	0.13	4.8
VIT_14s0068g01770	5.9E-03	WRKY DNA-binding protein 75	1.56	0.06	4.7
VIT_08s0058g00690	9.2E-03	WRKY DNA-binding protein 33	68.11	2.92	4.5
VIT_19s0027g00870	2.1E-03	VvNAC30_NAC domain-containing protein	1.78	0.12	3.9
VIT_12s0055g00340	6.0E-04	WRKY DNA-binding protein 9	1.48	0.10	3.8
VIT_12s0028g00860	1.1E-03	VvNAC36_NAC domain-containing protein	15.74	1.19	3.7
VIT_00s0179g00170	2.0E-03	Indeterminate(ID)-domain 12	0.56	0.06	3.3
VIT_13s0019g00480	4.1E-04	Zinc finger (C2H2 type) family	2.71	0.29	3.2
VIT_11s0052g00450	5.6E-03	WRKY DNA-binding protein 11	15.68	1.74	3.2
VIT_08s0007g08750	1.2E-03	Heat shock transcription factor B3	3.49	0.41	3.1
VIT_04s0069g00920	3.5E-03	WRKY DNA-binding protein 11	26.91	3.27	3.0
VIT_04s0023g01330	4.0E-04	Homeobox protein 5	2.06	0.32	2.7
VIT_12s0028g00980	3.6E-03	Myb family	8.64	1.39	2.6
VIT_02s0025g01280	3.7E-03	WRKY DNA-binding protein 53	3.68	0.64	2.5
VIT_01s0010g03930	9.3E-03	WRKY DNA-binding protein 75	109.75	22.15	2.3
VIT_06s0004g04950	9.8E-03	Scarecrow-like transcription factor 14 SCL14	48.81	10.43	2.2
VIT_17s0000g06930	2.1E-03	Unfertilized embryo sac 10 UNE10	1.22	0.27	2.2
VIT_06s0004g04960	1.3E-03	Scarecrow-like transcription factor 14 SCL14	23.75	5.34	2.2
VIT_10s0003g02810	2.4E-03	WRKY DNA-binding protein 71	1.85	0.42	2.1
VIT_19s0085g00050	2.4E-03	Myb domain protein 58	3.00	0.71	2.1
VIT_11s0016g01250	5.7E-03	Growth-regulating factor 6	0.67	0.18	1.9
VIT_10s0003g01170	9.8E-04	Basic helix-loop-helix (bHLH) family	8.11	2.27	1.8
VIT_19s0014g02240	8.4E-03	Ethylene responsive element binding factor 4	6.51	1.84	1.8
VIT_00s0375g00040	4.4E-03	VvNAC03_NAC domain-containing protein	107.59	31.99	1.8
VIT_07s0005g01450	6.7E-04	BZIP DNA-binding protein BZIP53	35.58	10.77	1.7
VIT_06s0080g00360	1.6E-03	BZIP transcription factor	6.09	1.89	1.7
VIT_08s0007g02810	3.0E-03	VvRAV5_ERF/AP2 Gene Family	7.69	2.44	1.7
VIT_06s0004g06710	9.2E-03	Cys-3-His zinc finger protein	63.88	22.01	1.5
VIT_10s0003g01160	1.8E-03	Basic helix-loop-helix (bHLH) family	8.41	2.95	1.5
VIT_10s0003g01600	2.9E-04	WRKY DNA-binding protein 65	15.79	5.71	1.5
VIT_01s0011g03110	9.7E-03	Myb family	11.77	4.40	1.4
VIT_13s0019g01810	2.4E-03	Scarecrow transcription factor 14 (SCL14)	12.14	4.70	1.4
VIT_13s0074g00400	7.2E-03	PTL (PETAL LOSS)	3.25	1.38	1.2
VIT_13s0019g01220	8.7E-03	Scarecrow transcription factor 3 (SCL3)	5.06	2.30	1.1
VIT_09s0002g01190	5.5E-03	Scarecrow transcription factor 29 (SCL29)	4.85	2.26	1.1
VIT_12s0028g02350	8.0E-03	Basic helix-loop-helix (bHLH) family	1.59	0.88	0.8
VIT_10s0003g03910	2.3E-03	TCP family transcription factor 24	6.99	3.91	0.8
VIT_07s0031g00080	4.7E-03	WRKY DNA-binding protein 15	10.50	5.94	0.8
VIT_04s0044g00510	8.7E-03	GT2-like trihelix DNA-binding protein	16.57	9.61	0.8
VIT_14s0108g00980	6.4E-03	Cycling DOF factor 1	34.46	21.07	0.7
VIT_08s0058g01300	9.7E-04	High-level expression of sugar-inducible like 1	21.05	13.32	0.7
VIT_18s0001g01820	3.5E-03	VvNAC15_NAC domain-containing protein	62.39	41.11	0.6
VIT_08s0040g00990	9.0E-03	Transcription elongation factor S-II	43.92	29.38	0.6
VIT_04s0023g01430	8.9E-03	Indeterminate(ID)-domain 2	73.59	50.16	0.6
VIT_08s0007g00360	7.6E-03	Myb domain protein 3R-5	16.79	11.70	0.5
VIT_08s0058g01410	2.2E-03	Transcription factor S-II (TFIIS)	50.38	38.16	0.4
VIT_17s0000g06340	3.7E-03	MADS-box agamous-like 30	34.99	27.04	0.4

Supplementary table 3: Gene composition of the overrepresented secondary metabolic process GO category resulting from enrichment analysis of the 3,450 genes with a higher expression in Barbera, Negro amaro and Refosco full-ripe berries. Gene expression level is indicated for each variety in FPKM. Abbreviations for each variety: BA (Barbera), NA (Negro amaro), RF (Refosco), SG (Sangiovese), PR (Primitivo) for the red-skinned varieties and VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina) for the white-skinned varieties.

Gene_ID	Functional annotation	BA	NA	RF	SG	PR	VR	GA	GL	MB	PS
VIT_01s0010g02740	4-coumarate-CoA ligase	1.06	1.50	4.96	0.61	0.56	0.37	0.39	0.19	0.40	0.39
VIT_11s0052g01090	4-coumarate-CoA ligase 1	33.98	61.34	240.79	18.96	18.54	11.03	2.84	2.21	14.35	2.59
VIT_17s0000g01790	4-coumarate-CoA ligase 2	2.95	6.18	6.58	6.65	3.32	0.94	1.36	2.98	1.48	0.35
VIT_16s0039g02040	4-coumarate-CoA ligase 3	11.01	25.52	50.75	7.38	8.15	2.51	1.42	1.04	4.65	1.09
VIT_07s0031g01130	Ascorbate oxidase	3.23	5.05	36.30	1.28	1.56	1.95	0.13	0.31	0.50	0.29
VIT_07s0031g01070	Ascorbate oxidase	0.87	4.13	21.19	1.46	1.58	0.49	0.06	0.20	0.10	0.18
VIT_14s0068g00660	ATAN11 (ANTHOCYANIN11)	42.05	34.69	45.41	40.15	32.54	36.53	34.16	34.06	35.42	28.65
VIT_01s0010g03510	Caffeoyl-CoA O-methyltransferase (AOMT)	35.82	21.12	56.68	2.42	30.78	0.02	0.02	0.02	0.02	0.01
VIT_16s0098g01790	Calmodulin-binding region IQD21	5.06	3.91	11.03	5.56	3.42	3.43	0.68	0.33	3.03	0.48
VIT_03s0038g01460	Chalcone synthase	0.85	2.11	1.49	0.59	0.82	0.47	0.46	0.45	0.61	0.26
VIT_14s0068g00920	Chalcone synthase (CHS2)	132.08	93.24	87.18	76.63	42.31	12.13	24.43	6.83	9.14	19.98
VIT_05s0136g00260	Chalcone synthase (CHS3)	968.86	288.65	691.49	182.46	309.68	13.06	66.09	18.52	12.13	31.85
VIT_13s0067g03820	Chalcone--flavonone isomerase (CHI1)	320.63	249.30	471.36	205.52	85.12	70.70	75.49	37.42	37.63	74.54
VIT_07s0129g01030	Cinnamyl alcohol dehydrogenase	0.44	0.64	1.09	0.67	0.19	0.13	0.30	0.10	0.37	0.21
VIT_00s0615g00020	Cinnamyl alcohol dehydrogenase	0.94	5.06	3.51	2.34	1.36	0.63	0.39	0.47	0.52	0.21
VIT_18s0001g12800	Dihydroflavonol 4-reductase	118.49	129.95	92.43	109.34	36.80	28.48	36.87	20.43	10.98	18.54
VIT_18s0001g00340	Diphenol oxidase	0.14	0.28	4.70	0	0	0.02	0	0	0.01	0.05
VIT_00s0481g00020	Diphenol oxidase	0.07	0.17	3.11	0.03	0	0	0	0.01	0	0
VIT_18s0122g01150	Diphenol oxidase	0.11	0.69	1.87	0	0	0	0	0	0.01	0.06
VIT_04s0023g03370	Flavonone-3-hydroxylase (F3H1)	73.07	55.30	60.04	22.12	12.28	7.94	15.80	7.45	5.55	9.37
VIT_07s0104g01800	Glutathione S-transferase 13 GSTF13	13.76	1.54	14.86	0.80	2.72	2.32	0.55	0.26	0.42	2.70
VIT_19s0093g00320	Glutathione S-transferase 25 GSTU25	2.71	2.42	1.72	0.02	0.39	0.28	0.04	0.18	0.14	0.17
VIT_06s0004g05670	Glutathione S-transferase 25 GSTU7	2.02	10.43	8.09	0.78	0.80	0.53	0.09	0.33	0.62	0.33
VIT_08s0007g01410	Glutathione S-transferase 8 GSTU8	0.24	0.26	0.16	0.07	0.12	0.04	0.03	0.03	0.03	0.09
VIT_08s0007g01400	Glutathione S-transferase 8 GSTU8	1.63	0.63	1.84	0.56	0.27	0.22	0.10	0.19	0.22	0.39
VIT_06s0004g05700	Glutathione S-transferase 8 GSTU8	1.35	2.68	1.89	1.05	0.32	0.08	0.08	0.13	0.06	0.21
VIT_14s0060g02170	Glutathione S-transferase GSTU8	0.16	0.04	0.22	0	0.03	0	0	0	0	0.02
VIT_18s0117g00550	Laccase	11.19	11.90	21.71	6.10	12.36	3.27	0.10	0.72	0.48	0.20
VIT_18s0075g00590	Laccase	1.33	2.51	18.34	0	0	0	0	0	0.05	0.09
VIT_18s0122g00690	Laccase	0.77	4.47	3.72	0.09	2.06	1.12	0.30	0.11	0.57	0.13
VIT_18s0075g00620	Laccase	0.58	2.26	9.44	0	0	0	0.01	0.02	0	0.02
VIT_18s0001g00400	Laccase	0.22	0.30	6.64	0	0	0.02	0.00	0.00	0.03	0.04
VIT_18s0075g00810	Laccase	0.03	0.21	1.08	0	0.01	0	0.01	0	0	0
VIT_00s1212g00020	Laccase	0.24	0.25	9.97	0	0.01	0.04	0.01	0	0.02	0.03
VIT_18s0075g00700	Laccase	0.22	0.30	1.71	0	0	0.05	0	0.04	0.04	0.17
VIT_13s0067g01970	Laccase	0.81	1.29	2.20	0.37	0.62	1.03	0.84	0.44	0.32	0.25
VIT_15s0046g00190	Laccase	4.08	4.12	3.31	2.50	2.03	2.52	1.26	1.12	2.14	0.88
VIT_18s0075g00580	Laccase	0.66	1.25	6.68	0.09	0.10	0.06	0.03	0.13	0.08	0.11
VIT_00s0731g00010	Laccase	1.36	3.12	54.28	0.04	0.02	0.09	0	0.03	0.03	0.01
VIT_18s0001g01280	Laccase	0.13	0.17	3.53	0	0	0.03	0.02	0	0	0.03
VIT_18s0001g00310	Laccase	0.19	0.22	3.78	0	0	0.01	0	0.01	0	0.02
VIT_00s0444g00010	Laccase	2.01	6.24	50.49	0.03	0.03	0.10	0	0.05	0.03	0.05
VIT_18s0117g00590	Laccase	1.97	2.84	5.43	0.03	0.25	0.13	0	0.09	0.04	0.17
VIT_13s0019g01930	Laccase	0.02	0	0.05	0	0	0	0	0	0	0
VIT_18s0001g00850	Laccase	1.16	3.38	15.90	0.01	0.01	0.08	0	0.01	0.04	0.04
VIT_18s0001g00790	Laccase	0.95	3.21	25.86	0.01	0.04	0.05	0.01	0.03	0.02	0.01
VIT_18s0001g01010	Laccase	0.18	0.06	3.89	0	0	0	0	0	0	0.01
VIT_18s0001g00730	Laccase	0.80	0.08	31.56	0	0.02	0.12	0	0	0.05	0.02
VIT_08s0007g06460	Laccase	0.02	0.02	0.06	0.04	0.02	0.01	0.01	0.02	0.01	0

(Table continues on following page)

Supplementary table 3 (Continued from previous page)

Gene_ID	Functional annotation	BA	NA	RF	SG	PR	VR	GA	GL	MB	PS
VIT_16s0039g01110	Phenylalanine ammonia-lyase	17.67	67.57	122.29	9.29	8.83	3.45	0.22	0.99	5.62	0.45
VIT_16s0039g01240	Phenylalanine ammonia-lyase	12.26	63.08	100.00	8.32	13.65	3.28	0.29	0.98	7.32	0.30
VIT_16s0039g01100	Phenylalanine ammonia-lyase	51.08	123.41	474.13	23.29	21.51	9.02	0.44	2.67	15.46	1.03
VIT_16s0039g01360	Phenylalanine ammonia-lyase	10.67	59.59	124.52	9.69	10.14	3.17	0.27	0.92	6.02	0.27
VIT_16s0039g01280	Phenylalanine ammonia-lyase	10.59	50.46	94.57	9.93	14.43	3.02	0.25	0.96	7.25	0.25
VIT_16s0039g01120	Phenylalanine ammonia-lyase	13.41	83.24	134.30	10.88	7.60	4.41	0.39	1.11	7.68	0.33
VIT_16s0039g01130	Phenylalanine ammonia-lyase	28.40	52.48	339.82	8.06	10.89	5.24	0.26	0.97	5.88	0.39
VIT_06s0004g02620	Phenylalanine ammonia-lyase	26.86	36.69	39.72	9.52	10.16	1.27	1.27	0.80	1.99	3.80
VIT_11s0016g01520	Phenylalanine ammonia-lyase	0.73	0.45	0.75	0.39	0.20	0.30	0.12	0.18	0.50	0.21
VIT_00s2849g00010	Phenylalanine ammonia-lyase	35.07	160.46	293.92	24.19	23.12	9.54	0.45	1.83	16.15	0.81
VIT_00s2508g00010	Phenylalanine ammonia-lyase	11.47	47.57	96.27	10.38	9.41	3.21	0.18	0.87	6.39	0.22
VIT_08s0040g01710	Phenylalanine ammonia-lyase (PAL1)	34.28	90.24	292.12	57.35	37.77	16.82	3.37	2.66	19.99	4.82
VIT_16s0039g01300	Phenylalanine ammonia-lyase	24.52	87.73	208.04	16.65	17.32	6.28	0.30	1.41	10.56	0.48
VIT_13s0019g04460	Phenylalanine ammonia-lyase 2 (PAL2)	48.21	50.00	61.89	11.04	23.50	2.90	12.76	1.33	3.56	21.12
VIT_16s0039g01170	Phenylalanine ammonia-lyase	38.55	91.93	264.26	14.77	20.96	6.47	0.42	2.03	9.88	0.85
VIT_02s0025g02920	Quercetin 3-O-methyltransferase 1	54.38	64.09	84.74	65.66	59.10	14.62	6.39	8.58	33.44	8.14
VIT_10s0003g04910	Sinapyl alcohol dehydrogenase	1.42	1.47	5.78	0	0.06	0.08	0.01	0.01	0.02	0.05
VIT_18s0122g00450	Sinapyl alcohol dehydrogenase	0.49	0.16	0.62	0.03	0.02	0.05	0.06	0.10	0.04	0.09
VIT_10s0042g00840	VvSTS1_Stilbene Synthase	16.06	83.65	201.50	0.49	2.94	3.24	0.27	1.43	4.89	0.30
VIT_10s0042g00890	VvSTS3_Stilbene Synthase	51.66	105.39	280.71	3.15	1.48	2.12	0.27	1.65	5.87	0.54
VIT_10s0042g00910	VvSTS4_Stilbene synthase	1.86	10.79	24.23	1.32	1.86	0.59	0	0.18	0.41	0.01
VIT_10s0042g00920	VvSTS5_Stilbene synthase	54.84	145.92	391.68	8.61	7.10	5.55	0.34	1.97	7.07	0.65
VIT_10s0042g00930	VvSTS6_Stilbene synthase	13.63	49.26	139.78	2.31	0.89	1.29	0.14	0.66	2.52	0.26
VIT_16s0100g00750	VvSTS7_Stilbene synthase	37.29	144.61	209.69	33.33	33.20	12.53	0.78	2.45	9.04	1.73
VIT_16s0100g00770	VvSTS9_Stilbene synthase	61.67	161.51	411.07	4.28	8.63	14.66	0.43	1.65	7.41	0.70
VIT_16s0100g00780	VvSTS10_Stilbene synthase	58.14	166.99	457.95	13.84	15.94	9.67	0.55	2.49	4.53	1.39
VIT_16s0100g00800	VvSTS12_Stilbene synthase	37.94	142.62	586.75	4.30	18.81	9.18	0.30	2.03	5.18	0.51
VIT_16s0100g00810	VvSTS13_Stilbene synthase	34.43	121.15	422.73	11.80	9.24	9.26	0.41	1.52	9.71	0.75
VIT_16s0100g00830	VvSTS15_Stilbene synthase	121.26	246.24	959.51	16.58	26.05	16.06	0.62	3.73	9.66	1.84
VIT_16s0100g00840	VvSTS16_Stilbene synthase	19.48	101.11	404.51	3.98	10.89	7.40	0.28	1.10	4.21	0.56
VIT_16s0100g00850	VvSTS17_Stilbene synthase	25.73	43.78	200.37	11.59	12.93	3.79	0.18	0.91	6.14	0.49
VIT_16s0100g00860	VvSTS18_Stilbene synthase	129.89	419.31	1,746.22	16.48	17.59	13.70	0.78	5.12	9.22	0.95
VIT_16s0100g00880	VvSTS19_Stilbene synthase	47.94	197.72	605.37	11.82	23.30	13.45	0.48	3.05	12.67	1.12
VIT_16s0100g00900	VvSTS20_Stilbene synthase	45.13	392.45	676.89	18.51	34.98	14.26	0.74	2.80	13.84	1.23
VIT_16s0100g00910	VvSTS21_Stilbene synthase	87.68	243.86	953.37	18.80	26.40	13.48	0.76	2.84	9.87	1.68
VIT_16s0100g00920	VvSTS22_Stilbene synthase	24.16	98.40	386.79	3.70	8.38	5.02	0.18	1.50	3.01	0.30
VIT_16s0100g00930	VvSTS23_Stilbene synthase	21.02	53.88	224.50	17.50	18.12	8.59	0.35	1.76	15.34	0.78
VIT_16s0100g00940	VvSTS24_Stilbene synthase	50.51	107.81	450.05	9.28	6.63	6.76	0.28	1.78	6.06	0.45
VIT_16s0100g00960	VvSTS26_Stilbene synthase	39.68	221.63	475.36	21.09	33.64	8.24	0.19	1.79	14.30	0.79
VIT_16s0100g00990	VvSTS27_Stilbene synthase	46.11	166.78	314.80	5.58	25.13	8.59	0.35	1.16	8.40	0.75
VIT_16s0100g01000	VvSTS28_Stilbene synthase	62.75	495.58	960.99	44.28	56.05	17.99	0.92	4.37	30.28	1.79
VIT_16s0100g01010	VvSTS29_Stilbene synthase	55.98	142.69	307.51	5.18	20.72	10.78	0.34	1.02	9.43	0.73
VIT_16s0100g01020	VvSTS30_Stilbene synthase	22.16	140.67	345.98	11.05	15.42	6.36	0.40	1.44	8.97	0.56
VIT_16s0100g01030	VvSTS31_Stilbene synthase	80.68	249.06	650.85	8.53	24.84	15.84	1.07	3.10	8.19	1.35
VIT_16s0100g01040	VvSTS32_Stilbene synthase	15.39	57.47	161.02	0.84	4.17	2.09	0.12	0.46	1.10	0.05
VIT_16s0100g01070	VvSTS35_Stilbene synthase	47.95	195.73	336.64	20.33	27.55	14.06	0.95	2.51	6.70	1.67
VIT_16s0100g01100	VvSTS36_Stilbene synthase	16.38	55.10	119.54	11.38	14.09	5.68	1.18	1.50	2.98	3.49
VIT_16s0100g01120	VvSTS39_Stilbene synthase	64.80	42.27	385.79	2.59	5.55	2.63	0.39	1.68	0.84	0.39
VIT_16s0100g01130	VvSTS41_Stilbene synthase	34.26	49.85	156.11	4.77	13.18	10.91	0.34	2.18	3.32	1.01
VIT_16s0100g01140	VvSTS42_Stilbene synthase	28.73	62.23	192.09	2.50	9.11	6.09	0.21	0.96	2.16	0.59
VIT_16s0100g01150	VvSTS43_Stilbene synthase	20.31	7.64	119.70	0.19	2.32	0.76	0.07	0.71	0.22	0.13
VIT_16s0100g01160	VvSTS45_Stilbene synthase	43.39	59.12	231.72	5.25	14.75	12.67	0.58	2.22	4.54	1.23
VIT_16s0100g01170	VvSTS46_Stilbene synthase	29.82	110.43	277.93	4.74	5.57	2.58	0.23	1.06	2.56	0.45
VIT_16s0100g01190	VvSTS47_Stilbene synthase	37.56	262.01	368.43	17.18	15.28	12.29	0.83	1.79	5.83	1.62
VIT_16s0100g01200	VvSTS48_Stilbene synthase	104.83	398.44	724.18	64.84	57.71	39.53	2.04	5.66	15.40	3.91

REFERENCES

1. Conde C., Silva P., Fontes N., et al. Biochemical changes throughout grape berry development and fruit and wine quality. Food 1:1-22 (2007)
2. Kobayashi S., Goto-Yamamoto N., Hirochika H. Retrotransposon-induced mutations in grape skin color. Science 304(5673):982 (2004)
3. Boss P.K., Davies C., Robinson S.P. Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L cv Shiraz grape berries and the implications for pathway regulation. Plant Physiol. 111(4):1059-1066 (1996)
4. Boss P.K., Davies C., Robinson S.P. Expression of anthocyanin biosynthesis pathway genes in red and white grapes. Plant Mol Biol. 32(3):565-569 (1996)
5. Fournier-Level A., Le Cunff L., Gomez C., et al. Quantitative genetic bases of anthocyanin variation in grape (*Vitis vinifera* L. ssp sativa) berry: A Quantitative Trait Locus to Quantitative Trait Nucleotide integrated study. Genetics 183(3):1127-39 (2009)
6. Walker A.R., Lee E., Bogs J., McDavid D.A., Thomas M.R., Robinson S.P. White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant J. 49(5):772-785 (2007)
7. Braidot E., Zancani M., Petrussa E., Peresson C., Bertolini A., Patui S., Macrì F., Vianello A. Transport and accumulation of flavonoids in grapevine (*Vitis vinifera* L.). Plant Signal Behav. 3(9):626-632 (2008)
8. Larson R.L. & Coe E.H. Gene-dependent flavonoid glucosyltransferase in maize. Biochem Genet. 15:153-156 (1977)
9. Bailly C., Cormier F., Do C.H. Characterization and activities of S-adenosyl-Lmethionine:cyanidin 3-glucoside 3′-O-methyltransferase in relation to anthocyanin accumulation in *Vitis vinifera* cell suspension cultures. Plant Sci. 122:81-89 (1997)
10. Mazzuca P., Ferranti P., Picariello E., Chianese G.L., Addeo F. Mass spectrometry in the study of anthocyanins and their derivatives: differentiation of *Vitis vinifera* and hybrid grapes by liquid chromatography/electrospray ionization mass spectrometry and tandem mass spectrometry. J Mass Spect. 40:83-90 (2005)
11. Mueller L.A., Goodman C.D., Silady R.A., Walbot V. AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. Plant Physiol. 123:1561-1570 (2000)
12. Gomez C., Conejero G., Torregrosa L., Cheynier V., Terrier N., Ageorges A. In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST. Plant J. 67(6):960-970 (2011)
13. Kobayashi S., Ishimaru M., Hiraoka K., Honda C. Myb-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis. Planta 215:924-933 (2002)
14. Kobayashi S., Yamamoto N.G., Hirochika H. Association of VvmybA1 gene expression with anthocyanin production in grape (*Vitis vinifera*) skin - color mutants. J Japan Soc Hort Sci. 74:196-203 (2005)
15. Maere S., Heymans K., Kuiper M. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21:3448-3449 (2005)

16. Jang M., Cai L., Udeani G.O., et al. Cancer chemopreventive activity of resveratrol, a natural product derived from grapes. Science 275(5297):218-220 (1997)
17. Smirnoff N., Conklin P.L., Loewus F.A. Biosynthesis of Ascorbic Acid in Plants: A Renaissance. Annu Rev Plant Physiol Plant Mol Biol. 52:437-467 (2001)
18. Smirnoff N. Ascorbic Acid: Metabolism and Functions of a Multi-Facetted Molecule. Current Opinion in Plant Biology 3(3):229-235 (2000)
19. Hancock R. & Viola R. Biosynthesis and catabolism of L-ascorbic acid in plants. Critical Reviews in Plant Sciences 24:167-188 (2005)
20. Pilati S., Brazzale D., Guella G., Milli A., Ruberti C., Biasioli F., Zottini M., Moser C. The onset of grapevine berry ripening is characterized by ROS accumulation and lipoxygenase-mediated membrane peroxidation in the skin. BMC Plant Biol. 14:87 (2014)
21. Melino V.J., Soole K.L., Ford C.M. Ascorbate metabolism and the developmental demand for tartaric and oxalic acids in ripening grape berries. BMC Plant Biology 9(1):145 (2009)
22. Li L.G., Cheng X.F., Leshkevich J., Umezawa T., Harding S.A., Chiang V.L. The last step of syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. Plant Cell 13:1567-1586 (2001)
23. Zhao Q., Nakashima J., Chen F., Yin Y., Fu C., Yun J., Shao H., Wang X., Wang Z.Y., Dixon R.A. Laccase is necessary and nonredundant with peroxidase for lignin polymerization during vascular development in Arabidopsis. Plant Cell 25(10):3976-3987 (2013)
24. Hayes M.A., Davies C., Dry I.B. Isolation, functional characterization, and expression analysis of grapevine (*Vitis vinifera* L.) hexose transporters: differential roles in sink and source tissues. J Exp Bot. 58:1985-1997 (2007)
25. Afoufa-Bastien D., Medici A., Jeauffre J., Coutos-Thévenot P., Lemoine R., Atanassova R., Laloi M. The *Vitis vinifera* sugar transporter gene family: phylogenetic overview and macroarray expression profiling. Plant Biology 10:245 (2010)
26. Rushton P.J., Torres J.T., Parniske M., Wernert P., Hahlbrock K., Somssich I.E. Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. The EMBO Journal 15(20):5690-5700 (1996)
27. Eulgem T., Rushton P.J., Robatzek S., Somssich I.E. The WRKY superfamily of plant transcription factors. Trends in Plant Science 5(5):199-206 (2000)
28. Ali K., Maltese F., Zyprian E., Rex M., Choi Y.H, Verpoorte R. NMR metabolic fingerprinting based identification of grapevine metabolites associated with downy mildew resistance. J Agr Food Chem. 57(20):9599-9606 (2009)
29. Rushton P.J., Somssich I.E., Ringler P., Shen Q.J. WRKY transcription factors. Trends in Plant Science 15(5):247-258 (2010)
30. Mukhtar M.S., Deslandes L., Auriac M.C., Marco Y., Somssich I.E. The Arabidopsis transcription factor WRKY27 influences wilt disease symptom development caused by *Ralstonia solanacearum*. Plant J. 56(6):935-947 (2008)
31. Tao Z., Liu H.B., Qiu D.Y., Zhou Y., Li X.H., Xu C.G., Wang S.P. A Pair of Allelic *WRKY* Genes Play Opposite Roles in Rice-Bacteria Interactions. Plant Physiol. 151(2):936-948 (2009)

32. Marchive C., Mzid R., Deluc L., Barrieu F., Pirrello J., Gauthier A., Corio-Costet M.F., Regad F., Cailleteau B., Hamdi S., Lauvergeat V. Isolation and characterization of a *Vitis vinifera* transcription factor, *VvWRKY1*, and its effect on responses to fungal pathogens in transgenic tobacco plants. J Exp Bot. 58(8):1999-2010 (2007)
33. Gallou A., Declerck S., Cranenbrouck S. Transcriptional regulation of defence genes and involvement of the WRKY transcription factor in arbuscular mycorrhizal potato root colonization. Funct Integr Genomics 12(1):183-198 (2012)
34. Yang P.Z., Chen C.H., Wang Z.P., Fan B.F., Chen Z.X. A pathogen- and salicylic acid-induced WRKY DNA-binding activity recognizes the elicitor response element of the tobacco class I chitinase gene promoter. Plant J. 18(2):141-149 (1999)
35. Wang Y.N., Dang F.F., Liu Z.Q., Wang X., Eulgem T., Lai Y., Yu L., She J.J., Shi Y.L., Lin J.H., Chen C.C., Guan D.Y., Qiu A., He S.L. *CaWRKY58*, encoding a group I WRKY transcription factor of *Capsicum annuum*, negatively regulates resistance to *Ralstonia solanacearum* infection. Mol Plant Pathol. 14(2):131-144 (2013)
36. Zhou X.F., Wang G.D., Sutoh K., Zhu J.K., Zhang W.X. Identification of cold-inducible microRNAs in plants by transcriptome analysis. Bba-Gene Regul Mech. 1779(11):780-788 (2008)
37. Zou C.S., Jiang W.B., Yu D.Q. Male gametophyte-specific WRKY34 transcription factor mediates cold sensitivity of mature pollen in Arabidopsis. J Exp Bot. 61(14):3901-3914 (2010)
38. Jiang Y.Q., Deyholos M.K. Functional characterization of Arabidopsis NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses. Plant Mol Biol. 69(1-2):91-105 (2009)
39. Wu X.L., Shiroto Y., Kishitani S., Ito Y., Toriyama K. Enhanced heat and drought tolerance in transgenic rice seedlings overexpressing OsWRKY11 under the control of HSP101 promoter. Plant Cell Rep. 28(1):21-30 (2009)
40. Ren X.Z., Chen Z.Z., Liu Y., Zhang H.R., Zhang M., Liu Q.A., Hong X.H., Zhu J.K., Gong Z.Z. *ABO3*, a WRKY transcription factor, mediates plant responses to abscisic acid and drought tolerance in Arabidopsis. Plant J. 63(3):417-429 (2010)
41. Shen H.S., Liu C.T., Zhang Y., Meng X.P., Zhou X., Chu C.C., Wang X.P. *OsWRKY30* is activated by MAP kinases to confer drought tolerance in rice. Plant Mol Biol. 80(3):241-253 (2012)
42. Liu H.Y., Yang W.L., Liu D.C., Han Y.P., Zhang A.M., Li S.H. Ectopic expression of a grapevine transcription factor VvWRKY11 contributes to osmotic stress tolerance in Arabidopsis. Mol Biol Rep. 38(1):417-427 (2011)
43. Zhang Y. & Feng J.C. Identification and characterization of the grape WRKY family. Biomed Res Int. 2014:787680 (2014)
44. Kiyosue T., Yamaguchi-Shinozaki K., Shinozaki K. ERD15, a cDNA for a dehydration-induced gene from Arabidopsis thaliana. Plant Physiol. 106:1707 (1994)
45. Kariola T., Brader G., Helenius E., Li J., Heino P., Palva E.T. EARLY RESPONSIVE TO DEHYDRATION 15, a negative regulator of abscisic acid responses in Arabidopsis. Plant Physiol. 142(4):1559-1573 (2006)
46. Coombe B.G. & Hale C.R. The hormone content of ripening grape berries and the effects of growth substance treatments. Plant Physiol. 51(4):629-634 (1973)

47. Düring H., Alleweldt G., Koch R. Studies on hormonal control of ripening in berries and grape vines. Acta Horticulturae 80:397-405 (1978)
48. Inaba A., Ishida M., Sobajima Y. Changes in endogenous hormone concentrations during berry development in relation to ripening of delaware grapes. J Jap Soc Hort Sci. 45(3):245-252 (1976)
49. Seki M., Ishida J., Narusaka M., Fujita M., Nanjo T., Umezawa T., Kamiya A., Nakajima M., Enju A., Sakurai T., Satou M., Akiyama K., Yamaguchi-Shinozaki K., Carninci P., Kawai J., Hayashizaki Y., Shinozaki K. Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. Funct Integr Genomics. 2(6):282-291 (2002)
50. Rabbani M.A., Maruyama K., Abe H., Khan M.A., Katsura K., Ito Y., Yoshiwara K., Seki M., Shinozaki K., Yamaguchi-Shinozaki K. (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA get-blot analyses. Plant Physiol. 133:1755-1767
51. Shinozaki K. & Yamaguchi-Shinozaki K. Gene networks involved in drought stress response and tolerance. J Exp Bot. 58(2):221-227 (2007)
52. Huang D., Wu W., Abrams S.R., Cutler A.J. The relationship of drought-related gene expression in *Arabidopsis thaliana* to hormonal and environmental factors. J Exp Bot. 59(11):2991-3007 (2008)
53. Okamoto G., Kuwamura T., Hirano K. Effects of water deficit stress on leaf and berry ABA and berry ripening in Chardonnay grapevines (*Vitis vinifera*). Vitis 43:15-17 (2004)
54. Deluc L.G., Quilici D.R., Decendit A., Grimplet J., Wheatley M.D., Schlauch K.A., Merillon J.M., Cushman J.C., Cramer G.R. Water deficit alters differentially metabolic pathways affecting important flJ.M. and quality traits in grape berries of Cabernet Sauvignon and Chardonnay. BMC Genomics 10:212 (2009)
55. Castellarin S.D., Matthews M.A., Di Gaspero G., Gambetta G.A. Water deficits accelerate ripening and induce changes in gene expression regulating flavonoid biosynthesis in grape berries. Planta 227:101-112 (2007a)
56. Castellarin S.D., Pfeiffer A., Sivilotti P., Degan M., Peterlunger E., Di Gaspero G. Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. Plant Cell Environ. 30:1381-1399 (2007b)
57. Matthews M.A. & Anderson M.M. Fruit ripening in *Vitis vinifera* L—responses to seasonal water deficits. Am J Enol Vitic. 39:313-320 (1988)
58. Jeong S.T., Goto-Yamamoto N., Kobayshi S., Esaka M.G. Effects of plant hormones and shading on the accumulation of anthocyanins and the expression of anthocyanin biosynthetic genes in grape berry skin. Plant Sci. 167:247-252 (2004)
59. Mori K., Saito H., Goto-Yamamoto N., Kitayama M., Kobayashi S., Sugaya S., Gemma H., Hashizume K. Effects of abscisic acid treatment and night temperatures on anthocyanin composition in 'Pinot noir' grapes. Vitis 44:161-165 (2005)
60. Peppi M.C., Fidelibus M., Dokoozlian N., Walker M.A. Abscisic acid applications improve the color of crimson seedless table grapes. Am J Enol Vitic. 57:388A (2006)
61. Peppi M.C., Walker M.A., Fidelibus M.W. Application of abscisic acid rapidly upregulated UFGT gene expression and improved color of grape berries. Vitis 47:11-14 (2008)

62. Giribaldi M., Gény L., Delrot S., Schubert A. Proteomic analysis of the effects of ABA treatments on ripening *Vitis vinifera* berries. J. Exp. Bot. 61(9):2447-2458 (2010)
63. Wheeler S., Loveys B., Ford C., Davies C. The relationship between the expression of abscisic acid biosynthesis genes, accumulation of abscisic acid and the promotion of *Vitis vinifera* L. berry ripening by abscisic acid. Aust J Grape Wine Res. 15:195-204 (2009)
64. Söllner T., Bennett M.K., Whiteheart S.W., Scheller R.H., Rothman J.E. A protein assembly-disassembly pathway in vitro that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. Cell 75:409-418 (1993)
65. Cosgrove D.J. Growth of the plant cell wall. Nature 6:950-961 (2005)
66. Nunan K.J., Sims I.M., Bacic A., Robinson S.P., Fincher G.B. Changes in cell wall composition during ripening of grape berries. Plant Physiol. 118(3):783-792 (1998)
67. Wermelinger B. Nitrogen dynamics in grapevine: Physiology and modeling. In: Proceedings of the International Symposium on Nitrogen in Grapes and Wine. Seatle pp. 23-31 (1991)
68. Tominaga T., Furrer A., Henry R., Dubourdieu D. Identification of new volatile thiols in aroma of *Vitis vinifera* L. var. Sauvignon blanc wines. Flavour Fragr J. 13:159-162 (1998)
69. Tominaga T., Murat M., Dubourdieu D. Development of a method for analyzing the volatile thiols involved in the characteristic aroma of wines made from *Vitis vinifera* L. cv Sauvignon blanc. J Agric Food Chem. 46:1044-1048 (1998)
70. Tominaga T., Baltenweck-Guyot R., Peyrot des Gachons C., Dubourdieu D. Contribution of volatile thiols to the aromas of white wines made from several *Vitis vinifera* grape varieties. Am J Enol Vitic. 51:178-181(2000)
71. Plücken H., Müller B., Grohmann D., Westhoff P., Eichacker L.A. The HCF136 protein is essential for assembly of the photosystem II reaction center in Arabidopsis thaliana. FEBS Lett. 532(1-2):85-90 (2002)
72. García-Cerdán J.G., Kovács L., Tóth T., Kereïche S., Aseeva E., Boekema E.J., Mamedov F., Funk C., Schröder W.P. The PsbW protein stabilizes the supramolecular organization of photosystem II in higher plants. Plant J. 65(3):368-381 (2011)
73. Wang Q., Sullivan R.W., Kight A., Henry R.L., Huang J., Jones A.M., Korth K.L. Deletion of the chloroplast-localized Thylakoid formation1 gene product in Arabidopsis leads to deficient thylakoid formation and variegated leaves. Plant Physiol. 136(3):3594-3604 (2004)
74. Li X.P., Björkman O., Shih C., Grossman A.R., Rosenquist M., Jansson S., Niyogi K.K. A pigment-binding protein essential for regulation of photosynthetic light harvesting. Nature 403(6768):391-395 (2000)
75. Fillion L., Ageorges A., Picaud S., et al. Cloning and expression of a hexose transporter gene expressed during the ripening of grape berry. Plant Physiol. 120:1083-1093 (1999)
76. Terrier N., Glissant D., Grimplet J., et al. Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (*Vitis vinifera* L.) development. Planta 222(5): 832-847 (2005)
77. Martin D.M., Aubourg S., Schouwey M.B., Daviet L., Schalk M., Toub O., Lund S.T, Bohlmann J. 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 (2010)

78. Elmlund H., Lundqvist J., Al-Karadaghi S., Hansson M., Hebert H., Lindahl M. A new cryo-EM singe-particle ab initio reconstruction method visualizes secondary elements in an ATP-fueled AAA+ motor. J Mol Biol. 375:934-947 (2008)
79. von Wettstein D., Henningsen K.W., Boynton C., Kannangara G., Nielsen O.F. The genetic control of chloroplast development in barley. In NK Boardman, AW Linnane, RM Smillie, eds, Autonomy and Biogenesis of Mitochondria and Chloroplast. North-Holland Publishing Company, Amsterdam, pp. 205-223 (1971)
80. Willows R.D., Gibson L.C.D., Kanangara C.G., Hunter C.N., von Wettstein D. Three separate proteins constitute the magnesium chelatase of Rhodobacter sphaeroides. Eur J Biochem. 235:438-443 (1996)
81. Jensen P.E., Gibson L.C.D., Hunter C.N. Determinants of catalytic activity with the use of purified I, D and H subunits of the magnesium protoporphyrin IX chelatase from *Synechocystis* PCC6803. Biochem J. 334:335-344 (1998)
82. Willows R.D & Beale S.I. Heterologous expression of the *Rhodobacter capsulatus* BchI, -D, and -H genes that encode magnesium chelatase subunits and characterization of the reconstituted enzyme. J Biol Chem. 273:34206-34213 (1998)
83. Karger G.A., Reid J.D., Hunter C.N. Characterization of the binding of deuteroporphyrin IX to the magnesium chelatase H subunit and spectroscopic properties of the complex. Biochemistry. 40:9291-9299 (2001)
84. Brown M.H., Paulsen I.T., Skurray R.A. The multidrug efflux protein NorM is a prototype of a new family of transporters. Mol. Microbiol. 31:393-395 (1999)
85. Diener A.C., Gaxiola R.A., Fink G.R. Arabidopsis ALF5, a multidrug efflux transporter gene family member, confers resistance to toxins. Plant Cell 13:1625-1638 (2001)
86. Zhao J., et al. MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in Medicago truncatula. Plant Cell 23:1536-1555 (2011)
87. Marinova K., Kleinschmidt K., Weissenbock G., Klein M. Flavonoid biosynthesis in barley primary leaves requires the presence of the vacuole and controls the activity of vacuolar flavonoid transport. Plant Physiol. 144:432-444 (2007)
88. Ren Q., Kang K.H., Paulsen I.T. TransportDB: A relational database of cellular membrane transport systems. Nucleic Acids Res. 32 (2004)
89. Huala E., et al. The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res. 29:102-105 (2001)
90. Pao S.S., Paulsen I.T., Saier M.H. Jr. Major facilitator superfamily. Microbiol. Mol. Biol. Rev. 62:1-34 (1998)
91. Büttner M. The monosaccharide transporter(-like) gene family in Arabidopsis. FEBS Lett. 581:2318-2324 (2007)

92. Tsay Y.F., Chiu C.C., Tsai C.B., Ho C.H., Hsu P.K. Nitrate transporters and peptide transporters. FEBS Lett. 581:2290-2300 (2007)
93. Haydon M.J. & Cobbett C.S. A novel major facilitator superfamily protein at the tonoplast influences zinc tolerance and accumulation in Arabidopsis. Plant Physiol. 143:1705-1719 (2007)
94. Mattivi F., Guzzon R., Vrhovsek U., Stefanini M., Velasco R. Metabolite profiling of grape: Flavonols and anthocyanins. J Agric Food Chem. 54(20):7692-7702 (2006)
95. Gatto P., Vrhovsek U., Muth J., et al. Ripening and genotype control stilbene accumulation in healthy grapes. J Agr Food Chem. 56(24):11773:11785 (2008)
96. Vincenzi S., Tomasi D., Gaiotti F., Lovat L., Giacosa S., Torchio F., Río Segade S., Rolle L. Comparative study of the resveratrol content of twenty-one Italian red grape varieties. S. Afr. J. Enol. Vitic. 34(1):30-35 (2013)
97. Vannozzi A., Dry I.B., Fasoli M., Zenoni S., Lucchin M. 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 (2012)
98. Vanholme R., Demedts B., Morreel K., Ralph J., Boerjan W. Lignin Biosynthesis and Structure. Plant Physiol. 153:895-905 (2010)
99. Wang A., Tan D., Takahashi A., Li T.Z., Harada T. MdERFs, two ethylene-response factors involved in apple fruit ripening. J Exp Bot. 58:3743-3748 (2007)
100. Chen G., Hu Z., Grierson D. Differential regulation of tomato ethylene responsive factor LeERF3b, a putative repressor, and the activator Pti4 in ripening mutants and in response to environmental stresses. J Plant Phys. 6:662-670 (2007)
101. El-Sharkawy I., Kim W.S., El-Kereamy A., Jayasankar S., Svircev A.M.D., Brown C.W. Isolation and characterization of four ethylene signal transduction elements in plums (*Prunus salicina* L.). J Exp Bot. 58:3631-3643 (2007)
102. Liu Q., Kasuga M., Sakuma Y., et al. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. Plant Cell 10(8):1391-1406 (1998)
103. Park J.M., Park C.J., Lee S.B., Ham B.K., Shin R., Paek K.H. Overexpression of the tobacco Tsi1 gene encoding an EREBP/AP2-type transcription factor enhances resistance against pathogen attack and osmotic stress in Tobacco. Plant Cell 13(5):1035-1046 (2001)
104. Gu Y.Q., Wildermuth M.C., Chakravarthy S., et al. Tomato transcription factors *Pti4*, *Pti5*, and *Pti6* activate defense responses when expressed in Arabidopsis. Plant Cell 14(4):817-831 (2002)
105. Thara V.K., Tang X., Gu Y.Q., Martin G.B., Zhou J.M. *Pseudomonas syringae* pv tomato induces the expression of tomato EREBP-like genes *Pti4* and *Pti5* independent of ethylene, salicylate and jasmonate. Plant Journal 20(4):475-483 (1999)
106. Kizis D. & Pagès M. Maize DRE-binding proteins DBF1 and DBF2 are involved in rab17 regulation through the drought-responsive element in an ABA-dependent pathway. Plant Journal 30(6):679-689 (2002)

107. Lorenzo O., Piqueras R., Sánchez-Serrano J.J., Solano R. ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. Plant Cell 15(1):165-178 (2003)

108. Vanholme B., Grunewald W., Bateman A., Kohchi T., Gheysen G. The tify family previously known as ZIM. Trends Plant Sci. 12:239-244 (2007)

109. Chini A., Fonseca S., Chico J.M., Fernandez-Calvo P., Solano R. The ZIM domain mediates homo- and heteromeric interactions between Arabidopsis JAZ proteins. The Plant Journal 59:77-87 (2009)

110. Chini A., Fonseca S., Fernandez G, Adie B, Chico J, et al. The JAZ family of repressors is the missing link in jasmonate signalling. Nature 448:666-671 (2007)

111. Thines B., Katsir L., Melotto M., Niu Y., Mandaokar A., et al. JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling. Nature 448:661-665 (2007)

112. Grunewald W., Vanholme B., Pauwels L., Plovie E., Inze D., Gheysen G., Goossens A. Expression of the Arabidopsis jasmonate signaling repressor JAZ1/TIFY10A is stimulated by auxin. EMBO Reports 10:923-928 (2009)

113. Yan Y., Stolz S., Chételat A., Reymond P., Pagni M., et al. A downstream mediator in the growth repression limb of the jasmonate pathway. Science's STKE 19:2470 (2007)

114. Devoto A., Turner J.G. Jasmonate-regulated Arabidopsis stress signalling network. Physiol Plant. 123:161-172 (2005)

115. Vezzulli S., Civardi S., Ferrari F., Bavaresco L. Methyl jasmonate treatment as a trigger of resveratrol synthesis in cultivated grapevine (*Vitis vinifera* L). Am J Enol Vitic. 58:530-533 (2007)

116. Snoep J.L., Westerhoff H.V., Rohwer J.M., Hofmeyr J.H. Is there an optimal ribosome concentration for maximal protein production? Syst Biol (Stevenage) 153(5):398-400 (2006)

117. Lecourieux F., Kappel C., Lecourieux D., Serrano A., Torres E., Arce-Johnson P., Delrot S. An update on sugar transport and signalling in grapevine. J Exp Bot. 65(3):821-832 (2014)

118. Deluc L.G., Grimplet J., Wheatley M.D., Tillett R.L., Quilici D.R., Osborne C., Schooley D.A., Schlauch K.A., Cushman J.C., Cramer G.R. Transcriptomic and metabolite analyses of Cabernet Sauvignon grape berry development. BMC Genomics 8:429-471 (2007)

119. Gollop R., Farhi S., Perl A. Regulation of the leucoanthocyanidin dioxygenase gene expression. Plant Science 161:579-588 (2001)

120. Gollop R., Even S., Colova-Tsolova V., Perl A. Expression of the grape dihydroflavonol reductase gene and analysis of its promoter region. Trends in Plant Science 53:1397-1409 (2002)

121. Zheng Y., Tian L., Liu H., Pan Q., Zhan J., Huang W. Sugars induce anthocyanin accumulation and flavanone 3-hydroxylase expression in grape berries. Plant Growth Regulators 58:251-260 (2009)

122. Dai Z.W., Meddar M., Renaud C., Merlin I., Hilbert G., Delrot S., Gomès E. Long-term in vitro culture of grape berries and its application to assess the effects of sugar supply on anthocyanin accumulation. J Exp Bot. 65(16):4665-4677 (2014)

123. Fung R.W., Gonzalo M., Fekete C., Kovacs L.G., He Y., Marsh E., McIntyre L.M., Schachtman D.P., Qiu W. Powdery mildew induces defense-oriented reprogramming of the transcriptome in a susceptible but not in a resistant grapevine. Plant Physiol. 146(1):236-249 (2008)

124. Schnee S., Viret O., Gindro K. Role of stilbenes in the resistance of grapevine to powdery mildew. Physiol Mol Plant Pathol. 72(4):128-133 (2008)

125. Langcake P. & Pryce R. The production of resveratrol by *Vitis vinifera* and other members of the Vitaceae as a response to infection or injury. Physiol Plant Pathol. 9(1):77-86 (1976)

126. Adrian M., Jeandet P., Veneau J., Weston L.A., Bessis R. Biological activity of resveratrol, a stilbenic compound from grapevines, against *Botrytis cinerea*, the causal agent for gray mold. J Chem Ecol 23(7):1689-1702 (1997)

127. Bézier A., Lambert B., Baillieul F. Study of defense-related gene expression in grapevine leaves and berries infected with *Botrytis cinerea*. Eur J Plant Pathol. 108(2):111-120 (2002)

128. Chiron H., Drouet A., Lieutier F., Payer H., Ernst D., Sandermann H. Gene induction of stilbene biosynthesis in Scots pine in response to ozone treatment, wounding, and fungal infection. Plant Physiol. 124(2):865-872 (2000)

129. Wang W., Tang K., Yang H., Wen P., Zhang P., Wang H., Huang W. Distribution of resveratrol and stilbene synthase in young grape plants (*Vitis vinifera* L. cv. Cabernet Sauvignon) and the effect of UV-C on its accumulation. Plant Physiol Biochem. 48(2):142-152 (2010)

130. Rosemann D., Heller W., Sandermann H. Biochemical plant responses to ozone II. Induction of stilbene biosynthesis in Scots pine (*Pinus sylvestris* L.) seedlings. Plant Physiol. 97(4):1280-1286 (1991)

131. Klipper-Aurbach Y., Wasserman M., Braunspiegel-Weintrob N., Borstein D., Peleg S., Assa S., Karp M., Benjamini Y., Hochberg Y., and Laron Z. Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. Med. Hypotheses 45:486-490 (1995)

Chapter 9

RNA profiling of the phenylpropanoid/ flavonoid biosynthetic pathway-related genes

INTRODUCTION

Expression profiling, also called RNA profiling, is a logical next step after sequencing a genome: the sequence provides information about the function of the gene in the cell – even a specific isoform derived from alternative splicing, while gene expression profile suggests when/how this function is carried out consistently with experimental conditions. In particular, measuring the relative amount of mRNA in at least two or more conditions – such time points, developmental stages, control versus treatment – permits to deeper characterize genes involved in specific pathways that are known and have interest in the specific field of research.

Grape phenolic compounds are a various group of secondary products with a wide range of biological functions. These functions include roles in plant reproduction through attraction of pollinators and seed dispersers, in protection against biotic and abiotic stresses, and in responses to diverse environmental stimuli. Berry phenolics are also important in the determination of wine style and quality, and have beneficial effects in many aspects of human health[1,2]. In the past decade, significant advances towards a better understanding of the genetics, biochemistry and physiology governing the synthesis of this class of secondary metabolite have been made allowing the identification of most of the genes involved in the biosynthesis.

Phenolics are synthesised from the amino acid phenylalanine[3]. Phenylalanine is a product of the shikimate pathway, which links carbohydrate metabolism with the biosynthesis of aromatic amino acids and secondary metabolites. Three main classes of phenolic compounds are synthesised in grape berry: phenolic acids, stilbenes, and flavonoids[4]. Simple phenols are molecules with a single C6 ring, which include benzoic acids (C6-C1) with a carbon bonded to the phenyl group (gentisic, salicylic, and gallic acid)[5-7], and hydroxycinnamic acids (C6-C3) with a C3 skeleton attached to the aromatic ring like caftaric acid[8]. The stilbene with the simplest molecular structure is the *trans*-resveratrol (3, 5, 4'-trihydroxystilbene) that could be either glycosylated, methylated or oligomerized[9]. Flavonoids are C6-C3-C6 polyphenolic compounds, in which two hydroxylated benzene rings, A and B, are joined by a three carbon chain that is part of a heterocyclic C ring. Flavonoids are divided into structural classes that include flavonols, flavan-3-ols, and anthocyanins, according to the oxidation state of the C ring[10,11]. The number of structural variants within each class is increased by modifications of stereochemistry, the position and type of substituents (including

hydroxyl, methyl, galloyl, and glycosyl groups, and aliphatic and aromatic acids), and the degree of polymerisation of basic units.

In this work, an example of how useful could be the data set generated from this first 10-variety berry RNA-Seq assay is provided. Indeed, RNA-profiles of genes belonging to well-characterized pathways in grape berry – namely the phenylpropanoid, flavonoid and stilbenoid biosynthetic pathways – were compared using hierarchical clustering analysis in order to link gene expression and phenolic compound biosynthesis, as the expression of phenylpropanoid pathway-related genes with the downstream pathways.

Biosynthesis of phenylpropanoids

The general phenylpropanoid pathway converts phenylalanine into 4-coumaroyl-CoA through three successive enzymatic reactions. The first reaction involves deamination of phenylalanine by the enzyme phenylalanine ammonia lyase (PAL), which produces cinnamic acid. In the second step, cinnamic acid is converted into p-coumaric acid by a hydroxylation at the 4 position by cinnamate-4-hydroxylase (C4H). Then, the enzyme 4-coumaroyl:CoA-ligase (4CL) catalyses the esterification of p-coumaric acid with CoA, which produces 4-coumaroyl-CoA. Intermediates of the general phenylpropanoid pathway give rise to phenolic acids. The end-product of the phenylpropanoid pathway, 4-coumaroyl-CoA, is the substrate of two enzymes, chalcone synthase (CHS) and stilbene synthase (STS), which control the enzymatic reactions at the entrance of the flavonoid pathway and stilbene pathway, respectively (**Fig. 1**). Modifications to intermediates of the phenylpropanoid pathway generate hydroxycinnamic acids. Nonetheless, hydroxycinnamic acid biosynthesis will not be described in this work.

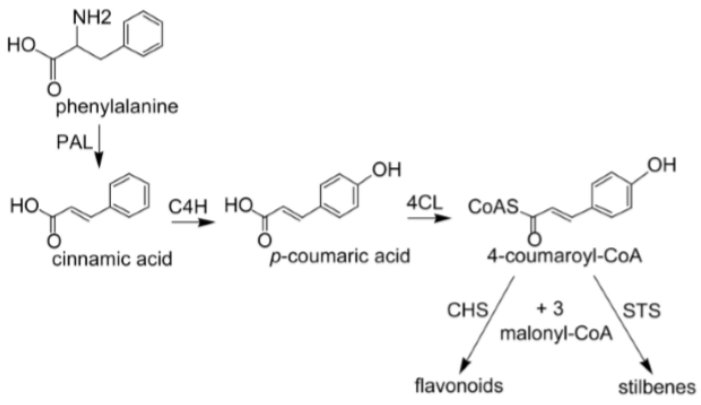


Figure 1: General phenylpropanoid pathway.

Biosynthesis of flavonoids

The flavonoid pathway leads to the synthesis of different classes of metabolites: flavonols, flavan-3-ols, proanthocyanidins and anthocyanins (**Fig. 2**).

Flavonols are yellow pigments that contribute to the colour of white wines. In the grape berry, flavonols are synthesised only in the skin, where they serve as UV-protecting agents. Synthesis of flavonols begins in flower buttons, it levels off during early fruit development, and it resumes during fruit ripening. Flavonols reach the peak amount per berry a few weeks after véraison[12].

Synthesis of flavan-3-ols and proanthocyanidins commences in flower buttons and proceeds in developing seeds and skins during fruit growth. Their synthesis is completed by véraison in skins and a few weeks after véraison in seeds. As a result, the content of flavan-3-ols and proanthocyanidins remains stable or decreases slightly during fruit ripening[13-17]. Flavan-3-ols and proanthocyanidins are predominantly found in the hypodermal cell layers of berry skin and in the soft parenchyma of the seed coat, within the vacuole or bound to cell wall polysaccharides[18-20]. Flavan-3-ols, proanthocyanidins and their complexes with polysaccharides have a significant impact on wine quality. All of them contribute to mouth-feel and smooth sensory perceptions[21].

Anthocyanins are responsible for red, purple, and blue pigmentation of grape berries. They are accumulated in berry skins of red varieties from véraison until full maturity, when the rate of synthesis levels off. Anthocyanins are synthesised in the cytosol of berry epidermal cells and then stored in the vacuole.

All flavonoids stem from tetrahydroxychalcone, which is synthesised by the enzyme chalcone synthase (CHS) using one molecule of 4-coumaroyl-CoA, the end-product of the phenylpropanoid pathway, and three molecules of malonyl-CoA (**Fig. 2**).

CHS catalyses the condensation and the partial cyclisation of three acetate units onto 4-coumaroyl-CoA, forming the A and B rings. Chalcone isomerase (CHI) accelerates the formation of the C ring to produce naringenin, the earliest intermediate with the typical flavonoid core.

Naringenin is a substrate for several enzymes. It may undergo either an hydroxylation at the 3 position by the activity of flavanone-3-hydroxylases (F3H) forming dihydrokaempferol, or an hydroxylation at the 3’ and 3’,5’ positions by the activity of flavonoid-3’-hydroxylase (F3’H) and flavonoid-3’,5’-hydroxylase (F3’5’H),

which catalyse the conversion into eriodictyol and pentahydroxyflavanone, respectively. Eriodictyol and pentahydroxyflavanone are then converted into dihydromyricetin or dihydroquercetin by the activity of F3H. Dihydrokaempferol is also a substrate for F3'H and F3'5'H, which catalyse the hydroxylation into dihydromyricetin or dihydroquercetin.

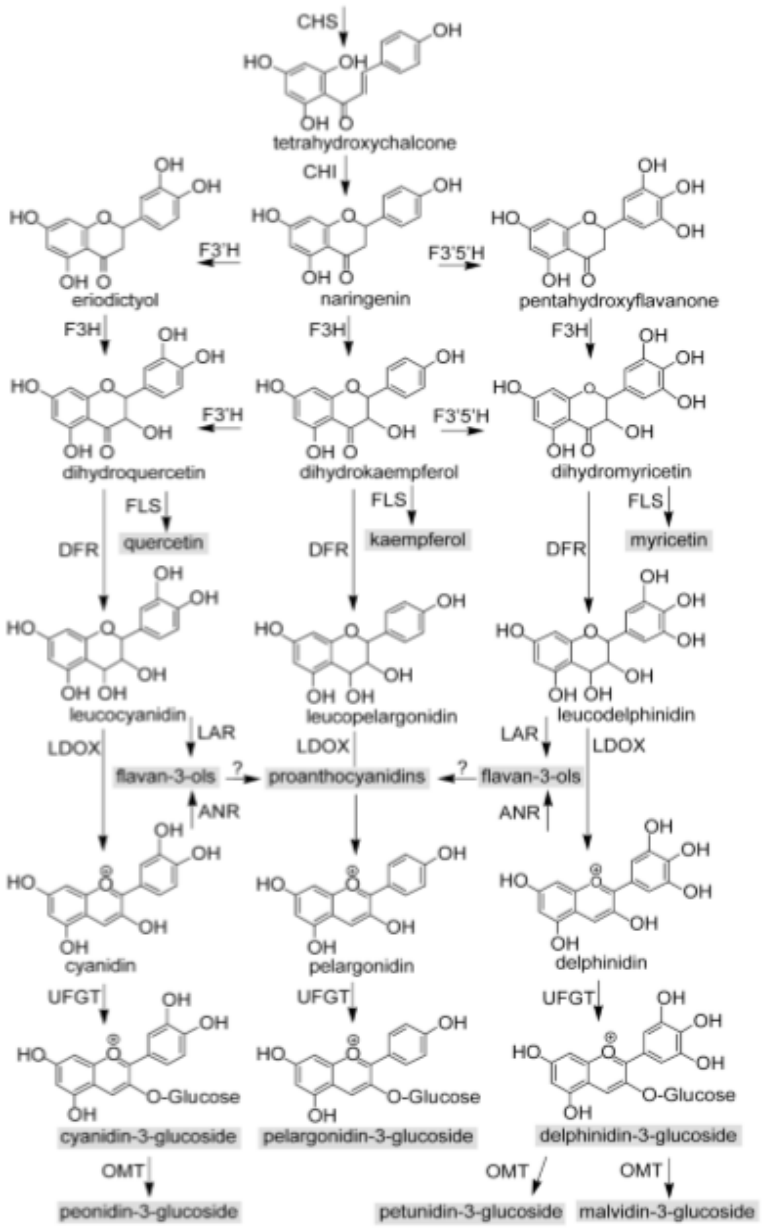


Figure 2: Flavonoid pathway.

At this point in the pathway, the three dihydroflavonols, i.e. dihydrokaempferol, dihydroquercetin, and dihydromyricetin, are substrates for a number of enzymes which

partition them into different downstream pathways. Dihydroflavonols may become substrates for the enzyme flavonol synthase (FLS)[12,22] that catalyses the formation of the flavonols kaempferol, quercetin, and myricetin. Flavonols are normally glycosylated at the 3 position of the C ring. Nevertheless, substrate specific 3-glycosyl transferases have not been isolated so far, but a UDP-glucose:flavonoid 3-O-glucosyl transferase (UFGT) with a high affinity for anthocyanidins also accepts flavonol substrates. Dihydroflavonols may also enter the pathways leading to the formation of anthocyanins and flavan-3-ols. The first committed step is the reduction of dihydroflavonols to flavan-3,4-ols, which is catalysed by a dihydroflavonol reductase (DFR)[23]. The activity of DFR on 3'-OH dihydrokaempferol sends the product into the branch of the pathway that leads to pelargonidin, the activity of DFR on 3',4'-OH dihydroquercetin leads to cyanidin, and the activity of DFR on 3',4',5'-OH dihydromyricetin leads to delphinidin.

Flavan-3,4-ols – also known as leucoanthocyanidins – are converted into anthocyanidins by the enzyme leucoanthocyanidin dioxygenase (LDOX) – also known as anthocyanidin synthase[24]. Leucoanthocyanidins and anthocyanidins may enter the flavan-3-ol pathway. The two enzymes: leucoanthocyanidin reductase (LAR) and anthocyanidin reductase (ANR) convert leucoanthocyanidins and anthocyanidins into (+)-catechin and (-)-epicatechin respectively. Flavan-3-ols can then polymerise into proanthocyanidins. This process has important implications for wine quality, but the underlying biochemistry of polymerisation remains largely unknown. Anthocyanidins with a hydroxyl group at the 3 position of the C ring are unstable. Instead of being reduced by ANR into epicatechin, anthocyanidins can be also be stabilised by the addition of a glucose at the 3 position of the C ring, resulting in anthocyanin production. The formation of glucosides is catalysed by UFGT. This enzyme accepts both anthocyanidins and flavonols as substrates but its catalytic efficiency is much higher for anthocyanidins[25].

UFGT controls the entry point into the terminal part of the flavonoid pathway, leading to the formation of anthocyanins[26,27]. The primary di-hydroxylated and tri-hydroxylated anthocyanins that are synthesised by UFGT are cyanidin-3-O-glucoside and delphinidin-3-O-glucoside. The 3'-OH of cyanidin and delphinidin and their 3-O-glucosides, and eventually the 5'-OH of delphinidin-3-O-glucoside, can be methylated by an anthocyanin *O*-methyltransferase (AOMT) generating peonidin, petunidin, and malvidin, and their 3-O-glucosides [28,29]. Furthermore, acylated anthocyanins are found in most red grapes, suggesting the presence of anthocyanin acyltransferases (AAT).

Acylation of anthocyanins consists in the addition of an aliphatic acetyl group or an aromatic p-coumaroyl (or occasionally a caffeoyl) group to the 6´ position of the 3-O-glucoside[30]. It is proposed that these modifications increase the stability of anthocyanins and change their water solubility[31]. Anthocyanin and flavonoid acyltransferases have been cloned from several plant species[32-34]. The enzymes show specificity for the acyl donor, but tend to be more promiscuous with regards to the acyl acceptor[32], suggesting that at least two AATs will be responsible for the acylation of grape anthocyanins.

Anthocyanin transport

Anthocyanins are synthesised in the cytosol and delivered into the vacuole, where they are stored as coloured coalescences called anthocyanic vacuolar inclusions. Vacuolar uptake may depend on multiple mechanisms, either mediated by tonoplast transporters or based on vesicular trafficking. Indeed, Gomez and colleagues[35] demonstrated that gluthatione S-transferase (GST)-mediated transport, vesicle trafficking and anthoMATE transporter mechanisms are all necessary for the final vacuolar accumulation of anthocyanins in grapevine. Two MATE proteins, anthoMATE1 (AM1) and anthoMATE3 (AM3), that specifically mediated acylated anthocyanin transport *in vitro*, were identified in grapevine[36]. AM1 and AM3 were localized mainly at the tonoplast and in membrane vesicles attached to the nucleus[36], suggesting that they participate in the transport of cytoplasmic-synthesized anthocyanins across the tonoplast. Among the GST gene family, GST4 has overlapping patterns of transcription with anthocyanin accumulation[37,38].

Regulation of the flavonoid pathway

Regulation of the flavonol pathway

Expression of *FLS* is specifically activated by the R2R3-MYB transcription factor VvMYBF1. The promoter of *VvMYBF1* contains light regulatory elements, which renders *VvMYBF1* expression extremely sensitive to light induction[39].

Regulation of flavan-3-ol and proanthocyanidin pathways

Expression of *LAR1* and *ANR* is regulated by the Myb transcription factors VvMybPA1 and VvMybPA2 [40,41]. VvMybPA1 and VvMybPA2 activate the promoters of *LAR1* and *ANR* triggering their expression, along with the expression of early genes of the

flavonoid pathway, but they do not activate the promoter of *LAR2* (the other terminal step in flavan-3-ols synthesis) or the *UFGT* promoter.

Another MYC-like basic Helix-Loop-Helix transcription factor, *VvMYC1*, participates in the control of flavan-3-ol synthesis[42]. *VvMYC1* can independently activate the *ANR* promoter, but it also physically interacts with *VvMybPA1* and, in this combination, activation of the *ANR* promoter increases 8-fold compared with *VvMYC1* alone.

Regulation of the anthocyanin pathway

Transcription of *VvUFGT* is regulated by the additive activity of two MybA-type transcription factors, *VvMybA1* and *VvMybA2*, which reside within a MybA gene cluster of four copies. This locus accounts for 62 % of the variance in anthocyanin content in linkage mapping experiments[43]. In a highly diverse core collection of grapevine varieties, five DNA polymorphisms within *VvMybA1*, *VvMybA2*, and *VvMybA3* account for 84% of the variance in anthocyanin content in those varieties. The allelic variation caused by the insertion of the *Gret1* Gypsy-type retrotransposon into the promoter of *VvMybA1* prevents *VvMybA1* transcription[44]. *VvMybA1* regulates expression of *UFGT* and other early genes of the flavonoid pathway and, much more strongly, a narrow set of genes relevant to anthocyanin metabolism: *F3'5'H*, *OMT*, *GST* and a vacuolar anthocyanin MATE-transporter gene[45,46].

Knowledge from studies of model systems would predict the participation of basic Helix-Loop-Helix and WD-repeat proteins in the transcriptional regulatory complex with Myb factors that activates anthocyanin synthesis. This is known to occur between the VvMYC1 transcription factor and VvMybA1 through physical interaction. The promoter of *VvUFGT* is induced ~40-fold after the co-expression of *VvMYC1* and *VvMybA1*[42]. *VvMYC1* also interacts with *Myb5a/Myb5b*, enhancing their activation of early genes in the flavonoid pathway [47,48] as *LAR*, *ANR*, *F3'H* and *CHI*.

Biosynthesis of stilbenes

Some stilbenes, particularly resveratrol, have long been known for their benefits to human health. Stilbenes are constitutively synthesised in healthy grape berries[49]. Stilbene content increases from véraison to ripening, and significant differences in final

content exist among varieties[50]. Their synthesis also increases upon pathogen infection and in response to abiotic stresses.

The first committed step of the stilbene pathway is controlled by stilbene synthase (STS). STS accepts the same substrates as chalcone synthase namely 4-coumaroyl-CoA, the end-product of the phenylpropanoid pathway, and 3 molecules of malonyl-CoA. In a similar way as CHS, STS carries out three reactions of condensation that produce resveratrol, a molecule with two aromatic rings. In the STS reaction, the terminal carboxyl group is removed prior to closure of the A ring, which causes different ring-folding in resveratrol compared to tetrahydroxychalcone, the product of CHS. CHS and STS thus compete for the same substrates and control the entry point into the flavonoid pathway and the stilbene pathway, respectively.

Little is known about other enzymatic steps that convert resveratrol into downstream derivatives. A resveratrol O-methyltransferase (ROMT) catalyses the conversion of resveratrol into the highly fungitoxic pterostilbene in grapevine leaves undergoing different stresses, including downy mildew (*Plasmopara viticola*) infection, ultraviolet light, and $AlCl_3$ treatment[51]. But, no evidence of its activity in grape berry has been demonstrated yet.

Regulation of the stilbene pathway

Expression of at least some *STS* copies is under the control of two R2R3-MYB–type transcription factors MYB14 and MYB15, which strongly coexpress with *STS* genes, both in leaf tissues under biotic and abiotic stress and in the skin and seeds of healthy developing berries during maturation[52]. In transient gene reporter assays, *MYB14* and *MYB15* were demonstrated to specifically activate the promoters of *STS* genes, and the ectopic expression of *MYB15* in grapevine hairy roots resulted in increased *STS* expression and in the accumulation of glycosylated stilbenes *in planta*[52].

RESULTS AND DISCUSSION

RNA-profiling of genes involved in phenylpropanoid biosynthetic pathway

Among the 21,746 transcripts identified as expressed (**chapter 4**), sixteen *PAL* genes, three *C4H* genes and ten *4CL* genes were identified. In order to evaluate their expression profiles in the ten variety-berries throughout development, a hierarchical

clustering analysis was performed using Pearson distance as the metric (TMeV 4.3; http://www.tm4.org/mev) (**Fig. 3**).

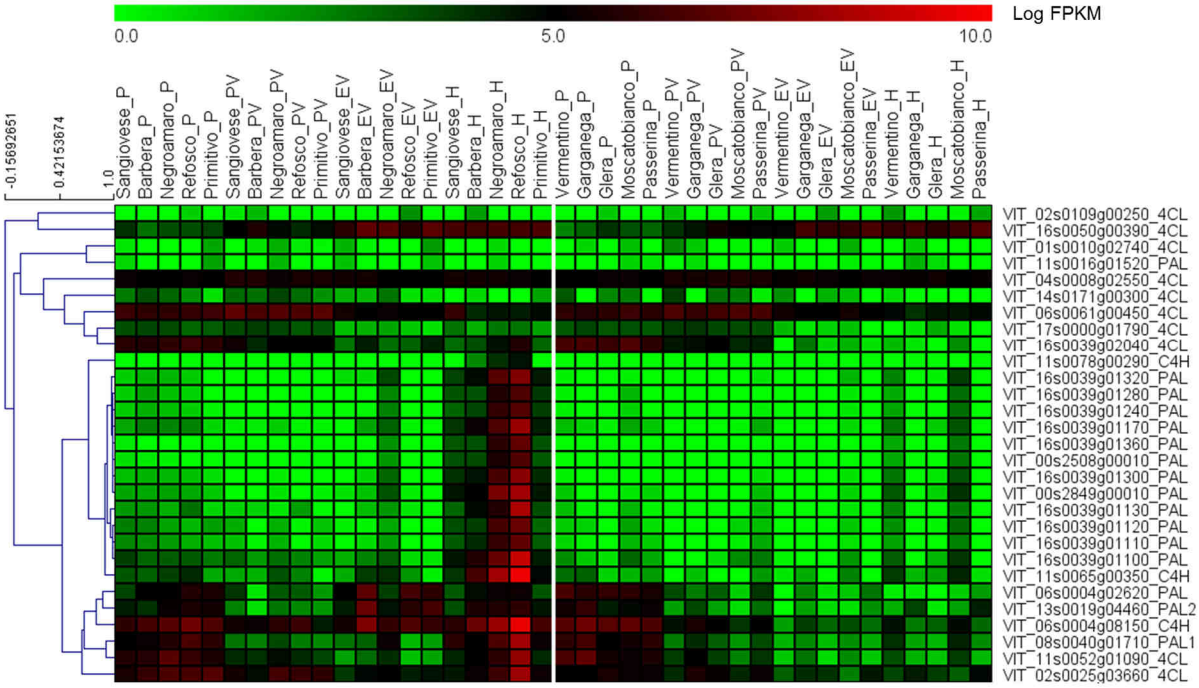


Figure 3: Hierachical clustering analysis of the genes involved in the general phenylpropanoid biosynthesis pathway. Abbreviations: phenylalanine ammonia lyase (PAL), cinnamate-4-hydroxylase (C4H), 4-coumaroyl:CoA-ligase (4CL). Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

The formation of four main clusters – i.e. the main branches of the dendrogram – could be observed. The first cluster was composed of the two *4CL* genes (VIT_02s0109g00250; VIT_16s0050g00390), which showed an increasing gene expression pattern during berry development in all varieties. The second cluster included one *PAL* gene (VIT_11s0016g01520) and six *4CL* genes (VIT_01s0010g02740; VIT_04s0008g02550; VIT_14s0171g00300; VIT_06s0061g00450; VIT_17s0000g01790; VIT_16s0039g02040). These transcripts exhibited a higher expression level during the green phase compared to the maturation phase, accordingly to the time of accumulation of flavan-3-ols, proanthocyanidins[13-17] and hydroxycinnamates[53].

The third cluster was composed of twelve *PAL* genes and two *C4H* genes (VIT_11s0078g00290; VIT_11s0065g00350) having a peak of expression in full-ripe red-skinned berries, suggesting their involvement in anthocyanin and stilbene biosynthesis[49]. Moreover, these genes showed a particularly higher expression in

Barbera, Negro amaro and Refosco mature berries, which exhibited the greatest accumulation of anthocyanins (**chapter 8**). However, the increase of expression at harvest also suggests that the precursors might be channelled to stilbenoid biosynthesis. Consequently, these three varieties may benefit from the activation of a specific group of the large PAL gene family determining a reinforcement of anthocyanin and stilbene biosynthesis.

Finally, three *PAL* genes including *PAL1* and *PAL2* (VIT_06s0004g02620; VIT_08s0040g01710; VIT_13s0019g04460), one *C4H* gene (VIT_06s0004g08150) and two *4CL* genes (VIT_11s0052g01090; VIT_02s0025g03660) composed the fourth cluster. Their gene expression patterns consisted in a peak of expression at pea-sized berry stage in all varieties, followed by a second peak occurring after véraison for two *PAL* genes and the *C4H* gene and at harvest for *PAL1* and the two *4CL* genes only in the five red-skinned berries. Consequently, we can suppose that these genes are related to proanthocyanidin accumulation in all varieties, and to anthocyanin and/or stilbene biosynthesis in red grapes.

RNA-profiling of flavonoid biosynthesis pathway-related genes

Regarding the flavonoid biosynthesis pathway, three *CHS* genes, two *CHI* genes, two *F3H* genes, two *F3'H* genes, twelve *F3'5'H* genes, seven *FLS* genes, two *LAR* and one *ANR* genes, one *DFR* gene, one *LDOX* gene, one *UFGT* gene, two *AOMT* genes and one predicted *AAT* gene were identified among the 21,746-transcript data set (**chapter 4**). We also found genes involved in anthocyanin transport as *GST4*, *AM1* and *AM3* and the transcription factor genes regulating the pathway: *MybA1* and *MybA2*, *MybPA1*, *MybPA2*, *MYBF1*, *Myb5a* and *Myb5b*, and *MYC1*. As for the general phenlypropanoid pathway-related genes, a hierarchical clustering analysis was carried out using Pearson distance as the metric (TMeV 4.3; http://www.tm4.org/mev) (**Fig. 4**), which highlighted again four main clusters.

The first cluster included the three *CHS* genes, the two *CHI* genes, the two *F3'H* genes, a *F3'5'H* gene (VIT_08s0007g05160), the *DFR* and *LDOX* genes. These genes showed a peak of expression at pea-sized berry stage of all varieties and exhibited a greater expression value in ripening red-skinned berries than in white grapes, suggesting an involvement in both flavan-3-ol, proanthocyanidin[13-17] and anthocyanin biosynthesis.

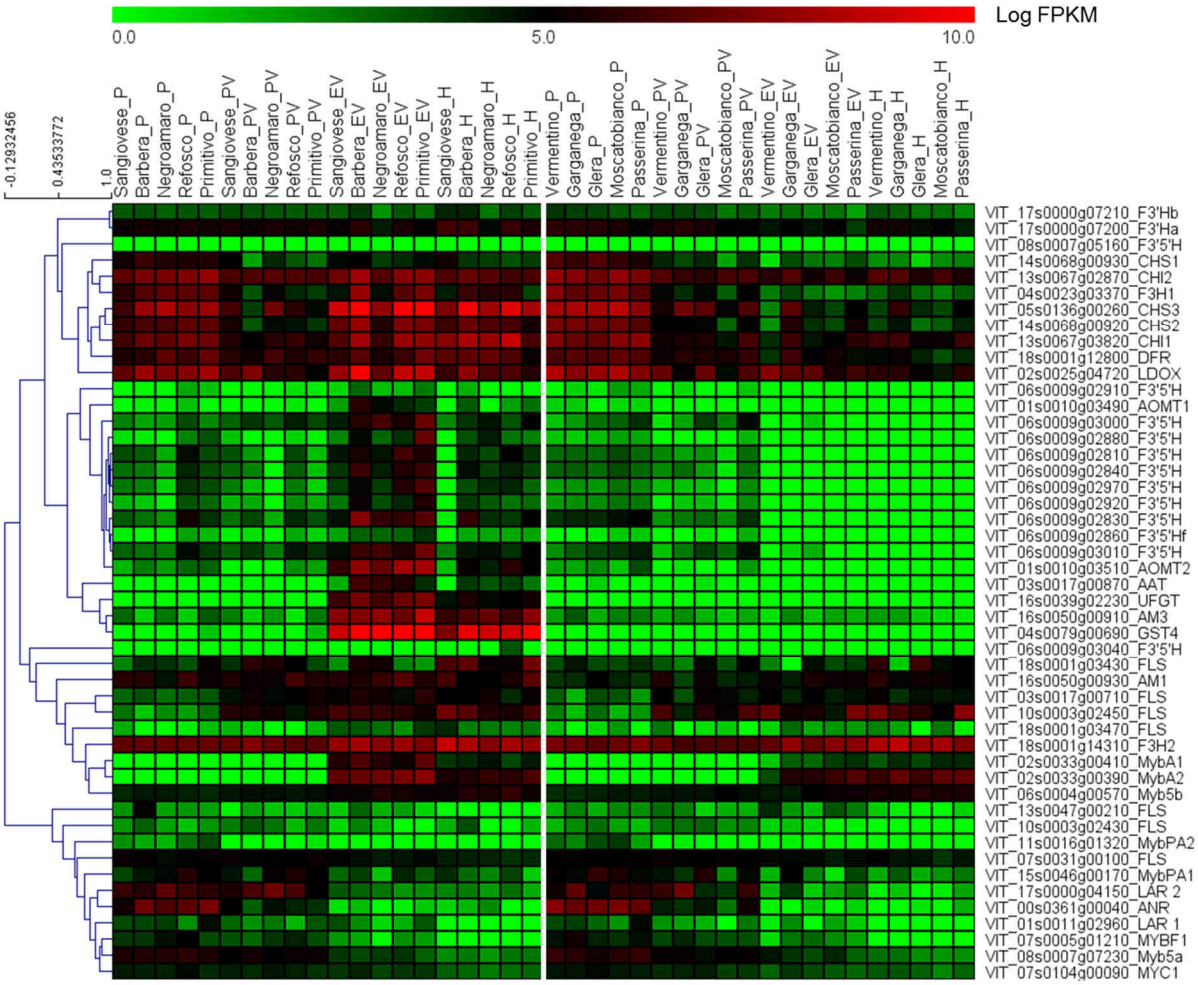


Figure 4: Hierachical clustering analysis of the genes involved in flavonoid biosynthesis pathway. Abbreviations: chalcone synthase (CHS), chalcone isomerase (CHI), flavanone-3-hydroxylase (F3H), flavonoid-3'-hydroxylase (F3'H), flavonoid-3',5'-hydroxylase (F3'5'H), flavonol synthase (FLS), leucoanthocyanidin reductase (LAR), anthocyanidin reductase (ANR), dihydroflavonol reductase (DFR), leucoanthocyanidin dioxygenase (LDOX), UDP-glucose:flavonoid 3-O-glucosyl transferase (UFGT), anthocyanin *O*-methyltransferase (AOMT), anthocyanin acyltransferase (AAT), gluthatione S-transferase (GST), antho-MATE transporter (AM). Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

Concerning the second cluster, it was composed of ten *F3'5'H* genes, *UFGT* gene, both *AOMT* and *AAT* genes, *AM3* and *GST4* genes. These genes showed a peak of expression at the end of véraison followed by a slight decline at maturity in red-skinned berries, confirming their involvement in anthocyanin biosynthesis and transport.

Genes belonging to the third cluster presented a higher gene expression level during the maturation independently of the skin colour. This cluster was composed of

F3H2 gene (VIT_18s0001g14310), a *F3'5'H* gene (VIT_06s0009g03040), four *FLS* genes (VIT_18s0001g03430; VIT_03s0017g00710; VIT_10s0003g02450; VIT_18s0001g03470) correlating with flavonol accumulation, *AM1* gene (VIT_16s0050g00930) and the two transcription factor genes *MybA1* and *MybA2* (VIT_02s0033g00410 and VIT_02s0033g00390 respectively). Surprisingly, *MybA1* gene was found as expressed in all varieties berries after véraison. As the *MybA1* allelic variation that prevents its transcription is known to be the cause of anthocyanins biosynthesis lack in white varieties[44], our result was surely due to a mismapping issue that will be described later in this chapter. Furthermore, *F3H2* was highly expressed in all the 40-sample berries (more than 100 FPKM), showing its involvement in the different pathways occurring downstream. Moreover, *AM1* was also expressed in white-skinned berries during maturation phase, suggesting another employment of this vacuolar transporter in white grapes.

Finally, the fourth cluster included three *FLS* genes (VIT_13s0047g00210; VIT_10s0003g02430; VIT_07s0031g00100), the two *LAR* genes, the *ANR* gene and the two transcription factor genes *MybPA1* and *MybPA2*. These genes exhibited a declining gene expression profile throughout berry development, confirming their involvement in the accumulation of flavan-3-ols and proanthocyanidins[13-17] in all varieties.

RNA-profiling of stilbene synthase genes

Stilbene synthases (STS) control the entry point into the stilbene pathway. Forty-four genes encoding STS were detected among the 21,746-transcript data set (**chapter 4**). In **Figure 5**, all *STS* genes have a high peak of expression in mature red-skinned berries, and more particularly in Barbera, Negro amaro and Refosco berries suggesting a greater resveratrol content in these berries. Indeed the content of stilbenes in the final product depends mainly on the grape variety[49] and, in particular, resveratrol concentration in berries was found to be quite high among other varieties[50]. The expression trend of the *STS* genes - alike the twelve *PAL* and the two *C4H* genes previously discussed - suggests that they might influence transcriptome of Barbera, Negroamaro and Refosco berries at harvest.

Moreover, *Myb14* and *Myb15* genes followed a similar pattern of expression to *STS* genes in full-ripe red grapes. Nevertheless, *Myb15* was also induced in red grapes at the end of véraison and Moscato bianco berries at both ripening stage. Consequently, we can suppose that Myb15 might regulate other target genes.

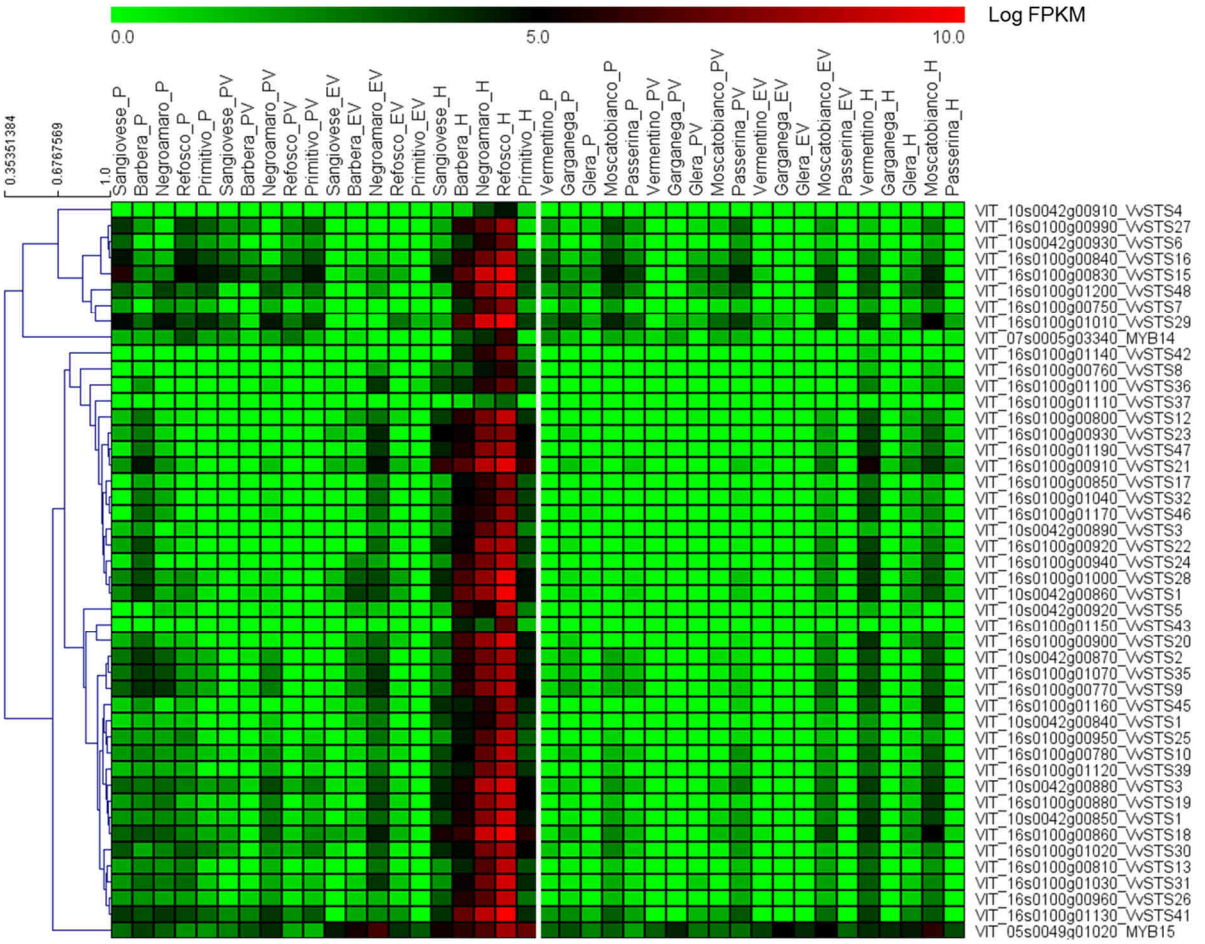


Figure 5: Hierachical clustering analysis of the genes involved in flavonoid biosynthesis pathway. Abbreviation for stilbene synthase (STS). Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest).

MybA1 gene expression in ripening white-skinned berries is due to a mismapping

In this RNA-Seq study, *MybA1* gene (VIT_02s0033g00410) exhibited a low but consistent expression in white-skinned berries at both post-véraison stages, which was not in line with the current literature[44]. In order to unravel the reason of this result, reads of selected samples (Garganega EV replicate 1; Glera EV replicate 2; Moscatobianco EV replicate 3; Vermentino EV replicate 2; Passerina EV replicate 3) that were not uniquely aligned to the *MybA1* gene locus of the reference genome PN40024_12x (chr2: 14,239,589-14,241,011) were filtered out. Then, unique reads mapped to *MybA1* gene locus were visualized using the Integrative Genomics Viewer software (version 2.3.32) (**Fig. 6**).

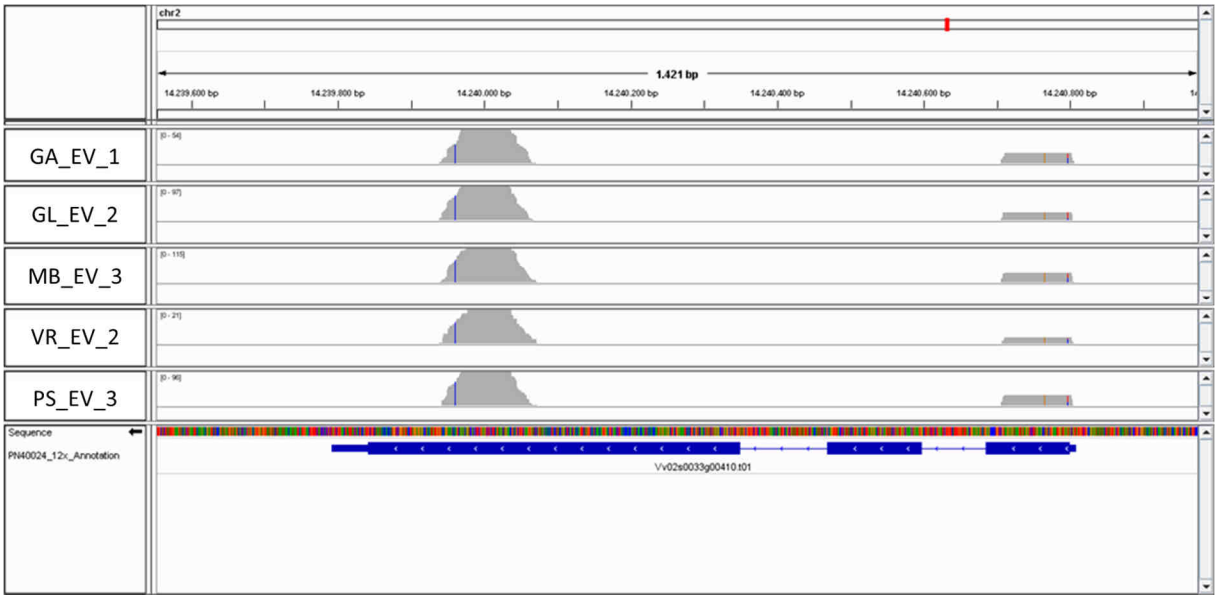


Figure 6: Unique reads mapped to *VvMybA1* gene locus in white grapes at the end of véraison (EV). Abbreviations for the five white-grape varieties: VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina).

Reads showed to cover two contiguous regions: the first region spanning across the exon 1 (chr2[-]:14,240,707-14,240,807) and the second region across the exon 3 (chr2[-]:14,239,943-14,240,072). The mapping profile resulted to be perfectly reproducible in all tested samples.

The unique reads mapped to exon 1 of *MybA1* gene locus exhibited two Single Nucleotide Polymorphisms (SNPs): a heterozygous SNP G > A at position chr2[-]:14,240,797 and a homozygous SNP T > C at position chr2[-]:14,240,766 in all the five white grape varieties. Considering that *MybA2* (VIT_02s0033g00390) differs from *MybA1* for those two SNPs only, the most likely explanation is that those reads were originated from *MybA2* transcripts and were wrongly mapped to *MybA1* instead.

Similarly, the unique reads mapped to exon 3 of the *MybA1* gene locus exhibited a homozygous SNP A > G at position chr2[-]:14,239,961 in all the five white grape varieties while the sequence of VIT_02s0033g00380 differs from *MybA1* for that SNP only, within the covered region, those reads are likely originated from the VIT_02s0033g00380 transcript. In addition, a close look at locus VIT_02s0033g00380 at position chr2[-]:14,179,390-14,179,733 showed that it overlapped VIT_02s0033g00390 chr2[-]:14,179,527-14,180,934, suggesting that VIT_02s0033g00380 is the terminal region of the *MybA2* gene – terminal region which is not translated in the white varieties[54].

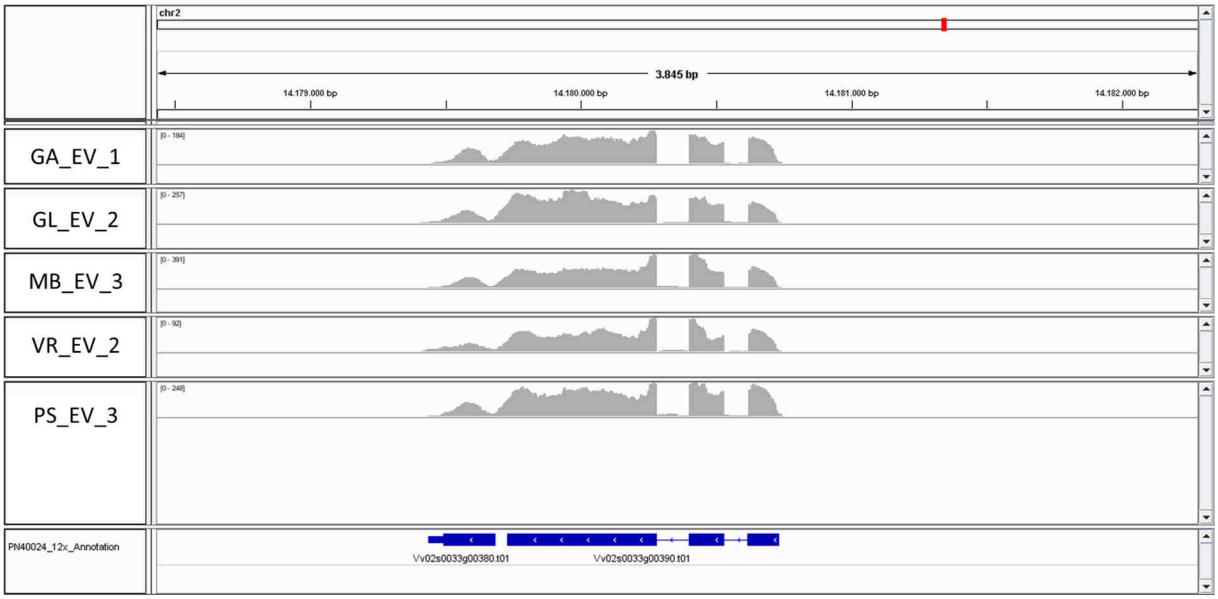


Figure 7: VIT_02s0033g00380 and VIT_02s0033g00390 gene loci are a unique gene: *MybA2*. Unique mapped reads to VIT_02s0033g00380 and VIT_02s0033g00390 gene loci were visualised for white grapes at the end of véraison (EV). Abbreviations for the five white-grape varieties: VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina).

In **Figure 7**, visualization of VIT_02s0033g00380 and VIT_02s0033g00390 mapped reads allows noticing the mapping of some reads onto the genomic region between the end of the third exon of VIT_02s0033g00380 from the start of VIT_02s0033g00390. Moreover, Walker and colleagues[54] demonstrated that a double STOP codon truncates the MybA2 protein (white allele). Its position lies at the right end of that space. This might have originated the wrong annotation of *MybA2* gene locus in the reference genome annotation V1[55] (http://genomes.cribi.unipd.it/DATA). Furthermore, in pre-véraison stages, *MybA2* was not expressed in all 10-variety berries, as *MybA1*, which correlates with literature[54].

Finally, we suggest that TopHat software mismapped those reads to *MybA1* sequence due to a problem of annotation. In fact, TopHat software first mapped the reads to the transcriptome, which was extracted from V1 annotation[55]. If a read originated from *MybA2* was longer than the exon, the alignment to *MybA1* was considered to have a higher score, despite one or two mismatches, because the alignment was longer.

CONCLUSION

This first RNA-Seq assay performed on 10-variety berries at four growth stage allowed gaining insights into the expression profiling of genes involved in phenylpropanoid, flavonoid and stilbenoid biosynthesis pathways. In fact, we were able to correlate each gene related to phenolic compound biosynthesis with the already-known biosynthesis timing, thus to identify phenylpropanoid and flavonoid biosynthetic pathway-related genes participating in the downstream pathways. Regarding the general phenylpropanoid pathway, gene expression of four *PAL* genes, including *PAL1* and *PAL2* genes (VIT_06s0004g02620; VIT_08s0040g01710; VIT_13s0019g04460; VIT_11s0016g01520), one *C4H* gene (VIT_06s0004g08150) and the all *4CL* genes could be associated with the accumulation of flavan-3-ols, proanthocyanidins[13-17] and hydroxycinnamates[53]. Moreover expression of all *PAL* genes except VIT_11s0016g01520, the three *C4H* genes, and two *4CL* genes (VIT_11s0052g01090; VIT_02s0025g03660) correlated with anthocyanin and stilbene biosynthesis in red grapes[49]. Concerning the flavonoid biosynthesis pathway, all *CHS*, *CHI* genes and *F3'H* genes, a *F3'5'H* gene (VIT_08s0007g05160), the *DFR* and *LDOX* genes seemed to be involved in both flavan-3-ols, proanthocyanidins[13-17] and anthocyanin biosynthesis. Furthermore, *F3H2* gene was highly expressed in all the 40-sample berries (more than 100 FPKM), showing its involvement in all the different downstream pathways.

Finally, *MybA1* expression in white grapes during maturation phase was probably due to an annotation issue of *MybA2* gene in the gene annotation V1[55], suggesting that the grapevine gene annotation needs to be improved in the future.

METHODS

Hierarchical Clustering

Hierarchical clustering analyses were performed on genes involved in phenlypropanoid, flavonoid and stilbene synthesis pathways using Pearson's correlation distance as the metric (TMeV 4.3; http://www.tm4.org/mev).

Read visualization

After filtering out reads that were not uniquely aligned to the *MybA1* gene locus of the reference genome PN40024_12x (chr2: 14,239,589-14,241,011), unique reads mapped to *MybA1* gene locus of selected white-skinned berry samples were visualized

using the Integrative Genomics Viewer software (version 2.3.32). The same approach was used to visualize VIT_02s0033g00380 and *MybA2* (VIT_02s0033g00390) gene loci.

REFERENCES

1. Corder R., Mullen W., Khan N.Q., Marks S.C., Wood E.G., Carrier M.J., Crozier A. Oenology: red wine procyanidins and vascular health. Nature 444(7119):566 (2006)
2. Jang M., Cai L., Udeani G.O., et al. Cancer chemopreventive activity of resveratrol, a natural product derived from grapes. Science 275(5297):218-220 (1997)
3. Sparvoli F., Martin C., Scienza A., Gavazzi G., Tonelli C. Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (*Vitis vinifera* L.). Plant Mol Biol. 24(5):743-55 (1994)
4. Castellarin S.D., Bavaresco L., Falginella L., Gonçalves M.I.V.Z., Di Gaspero G. Phenolics in grape berry and key antioxidants. In The Biochemistry of the Grape Berry, H. Geros, M.M. Chavez, and S. Delrot, eds (Sharjan, United Arab Emirates: Bentham Science Publishers), pp. 89-110 (2012)
5. Vanhoenacker G., de Villiers A., Lazou K., de Keukeleire D., Sandra P. Comparison of high-performance liquid chromatography—Mass spectroscopy and capillary electrophoresis—Mass spectroscopy for the analysis of phenolic compounds in diethyl ether extracts of red wines extraction of phenolic compounds LC-UV-MS high-per. Chromatographia 54:309-315 (2001)
6. Pozo-Bayón M.A., Hernández M.T., Martín-Alvarez P.J., Polo M.C. Study of low molecular weight phenolic compounds during the aging of sparkling wines manufactured with red and white grape varieties. J. Agric. Food. Chem. 51:2089-2095 (2003)
7. Ali K., Maltese F., Choi Y.H., Verpoorte R. Metabolic constituents of grapevine and grape-derived products. Phytochem. Rev. 9:357-378 (2010)
8. Singleton V., Zaya J., Trousdale E. Caftaric and coutaric acids in fruit of Vitis. Phytochemistry 25:2127-2133 (1986)
9. Conde C., Silva P., Fontes N., et al. Biochemical changes throughout grape berry development and fruit and wine quality. Food 1:1-22 (2007)
10. Braidot E., Zancani M., Petrussa E., Peresson C., Bertolini A., Patui S., Macrì F. Transport and accumulation of flavonoids in grapevine (*Vitis vinifera* L.). Plant Signal. Behav. 3:626-632 (2008)
11. Adams, D. Phenolics and ripening in grape berries. Am. J. Enol. Vitic. 3:249-256 (2006)
12. Downey M.O., Harvey J.S., Robinson S.P. Synthesis of flavonols and expression of flavonol synthase genes in the developing grape berries of Shiraz and Chardonnay (*Vitis vinifera* L.). Aust J Grape Wine R. 9(2):110-121 (2003a)
13. Downey M.O., Harvey J.S., Robinson S.P. Analysis of tannins in seeds and skins of Shiraz grapes throughout berry development. Aust J Grape Wine R. 9(1):15-27 (2003b)
14. Hanlin R.L., Downey M.O. Condensed tannin accumulation and composition in skin of Shiraz and Cabernet Sauvignon grapes during berry development. Am J Enol Viticult. 60(1):13-23 (2009)
15. De Freitas V.A.P., Glories Y., Monique A. Developmental changes of procyanidins in grapes of red *Vitis vinifera* varieties and their composition in respective wines. Am J Enol Viticult. 51(4):397-403 (2000)
16. del Rio J.L.P., Kennedy J.A. Development of proanthocyanidins in *Vitis vinifera* L. cv. Pinot noir grapes and extraction into wine. Am J Enol Viticult. 57(2):125-132 (2006)

17. Harbertson J.F., Kennedy J.A., Adams D.O. Tannin in skins and seeds of Cabernet Sauvignon, Syrah, and Pinot noir berries during ripening. Am J Enol Viticult. 53(1):54-59 (2002)
18. Amrani-Joutei K., Glories Y., Mercier M. Localization of tannins in grape berry skins. Vitis 33:133-138 (1984)
19. Geny L., Saucier C., Bracco S., Daviaud F., Glories Y. Composition and cellular localization of tannins in grape seeds during maturation. J Agr Food Chem 51(27): 8051-8054 (2003)
20. Hanlin R.L., Hrmova M., Harbertson J.F., Downey M.O. Condensed tannin and grape cell wall interactions and their impact on tannin extractability into wine. Aust J Grape Wine Res. 16(1):173-188 (2010)
21. Cheynier V., Prieur C., Guyot S., Rigaud J., Moutounet M. The structures of tannins in grapes and wines and their interactions with proteins. Acs Sym Ser. 661:81-93 (1997)
22. Fujita A., Goto-Yamamoto N., Aramaki I., Hashizume K. Organ-specific transcription of putative flavonol synthase genes of grapevine and effects of plant hormones and shading on flavonol biosynthesis in grape berry skins. Biosci Biotechnol Biochem. 70(3):632-638 (2006)
23. Gollop R., Even S., Colova-Tsolova V., Perl A. Expression of the grape dihydroflavonol reductase gene and analysis of its promoter region. J Exp Bot. 53(373):1397-1409 (2002)
24. Gollop R., Farhi S., Perl A. Regulation of the leucoanthocyanidin dioxygenase gene expression in *Vitis vinifera*. Plant Sci. 161(3):579-588 (2001)
25. Ford C.M., Boss P.K., Hoj P.B. Cloning and characterization of *Vitis vinifera* UDP-glucose:flavonoid 3-Oglucosyltransferase, a homologue of the enzyme encoded by the maize Bronze-1 locus that may primarily serve to glucosylate anthocyanidins in vivo. J Biol Chem. 273(15):9224-9233 (1998)
26. Boss P.K., Davies C., Robinson S.P. Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L cv Shiraz grape berries and the implications for pathway regulation. Plant Physiol. 111(4):1059-1066 (1996)
27. Boss P.K., Davies C., Robinson S.P. Expression of anthocyanin biosynthesis pathway genes in red and white grapes. Plant Mol Biol. 32(3):565-569 (1996)
28. Hugueney P., Provenzano S., Verries C., et al. A novel cation-dependent O-methyltransferase involved in anthocyanin methylation in grapevine. Plant Physiol. 150(4):2057-2070 (2009)
29. Lucker J., Martens S., Lund S.T. Characterization of a *Vitis vinifera* cv. Cabernet Sauvignon 3 '5 '-O-methyltransferase showing strong preference for anthocyanins and glycosylated flavonols. Phytochemistry 71(13):1474-1484 (2010)
30. Mazzuca P., Ferranti P., Picariello E., Chianese G.L., Addeo F. Mass spectrometry in the study of anthocyanins and their derivatives: differentiation of *Vitis vinifera* and hybrid grapes by liquid chromatography/electrospray ionization mass spectrometry and tandem mass spectrometry. J Mass Spect. 40:83-90 (2005)
31. Strack D. & Wray V. The anthocyanins. In: Harborne JB (ed) The Flavonoids. Chapman & Hall/CRC Press, Washington DC (1994)
32. Nakayama T., Suzuki H., Nishino T. Anthocyanin acyltransferases:specificities, mechanism, phylogenetics, and applications. J Mol Catal B Enzymatic. 23:117-132 (2003)

33. D'Auria JC. Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol. 9:331-340 (2006)
34. Luo J., Nishiyama Y., Fuell C., Taguchi G., Elliott K., Hill L., Tanaka Y., Kitayama M., Yamazaki M., Bailey P., Parr A., Michael A.J., Saito K., Martin C. Convergent evolution in the BAHD family of acyl transferases:identification and characterization of anthocyanin acyltransferases from *Arabidopsis thaliana*. Plant J. 70:6678-6695 (2007)
35. Gomez C., Conejero G., Torregrosa L., Cheynier V., Terrier N., Ageorges A. In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST. Plant J. 67(6):960-970 (2011)
36. Gomez C., Terrier N., Torregrosa L., et al. Grapevine MATE-type proteins act as vacuolar H+-dependent acylated anthocyanin transporters. Plant Physiol. 150(1):402-415 (2009)
37. Ageorges A., Fernandez L., Vialet S., Merdinoglu D., Terrier N., Romieu C. Four specific isogenes of the anthocyanin metabolic pathway are systematically co-expressed with the red colour of grape berries. Plant Sci. 170(2):37283 (2006)
38. Castellarin S.D. & Di Gaspero G. Transcriptional control of anthocyanin biosynthetic genes in extreme phenotypes for berry pigmentation of naturally occurring grapevines. BMC Plant Biol. 7:46 (2007)
39. Czemmel S., Stracke R., Weisshaar B., et al. The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. Plant Physiol. 151(3):1513-30 (2009)
40. Bogs J., Jaffe F.W., Takos A.M., Walker A.R., Robinson S.P. The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. Plant Physiol. 143(3):1347-1361 (2007)
41. Terrier N., Torregrosa L., Ageorges A., Vialet S., Verries C., Cheynier V., Romieu C. Ectopic expression of *VvMybPA2* promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. Plant Physiol. 149(2):1028-1041 (2009)
42. Hichri I., Heppel S.C., Pillet J., et al. The basic Helix-Loop-Helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. Mol Plant. 3(3):509-523 (2010)
43. Fournier-Level A., Le Cunff L., Gomez C., et al. Quantitative genetic bases of anthocyanin variation in grape (*Vitis vinifera* L. ssp sativa) berry: A Quantitative Trait Locus to Quantitative Trait Nucleotide integrated study. Genetics 183(3):1127-1139 (2009)
44. Kobayashi S., Goto-Yamamoto N., Hirochika H. Retrotransposon-induced mutations in grape skin color. Science 304(5673):982 (2004)
45. Cutanda-Perez M.C., Ageorges A., Gomez C., et al. Ectopic expression of VlmybA1 in grapevine activates a narrow set of genes involved in anthocyanin synthesis and transport. Plant Mol Biol. 69(6):633-648 (2009)
46. Poudel P.R., Azuma A., Kobayashi S., Goto-Yamamoto N. Coordinate induction of anthocyanin biosynthetic pathway genes by VvMybAs In: Proceedings of the 10th International Conference on Grapevine Breeding and Genetics. Geneva, 1-5 August 2010 (2010)

47. Deluc L., Barrieu F., Marchive C., et al. Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. Plant Physiol. 140(2):499-511 (2006)
48. Deluc L., Bogs J., Walker A.R., Ferrier T., et al. The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. Plant Physiol. 147(4): 2041-2053 (2008)
49. Gatto P., Vrhovsek U., Muth J., et al. Ripening and genotype control stilbene accumulation in healthy grapes. J Agr Food Chem. 56(24):11773-11785 (2008)
50. Vincenzi S., Tomasi D., Gaiotti F., Lovat L., Giacosa S., Torchio F., Río Segade S., Rolle L. Comparative study of the resveratrol content of twenty-one Italian red grape varieties. S. Afr. J. Enol. Vitic. 34(1):30-35 (2013)
51. Schmidlin L., Poutaraud A., Claudel P., et al. A stress-inducible resveratrol O-methyltransferase involved in the biosynthesis of pterostilbene in grapevine. Plant Physiol. 148(3):1630-1639 (2008)
52. Höll J., Vannozzi A., Czemmel S., D'Onofrio C., Walker A.R., Rausch T., Lucchin M., Boss P.K., Dry I.B., Bogs J. The R2R3-MYB transcription factors MYB14 and MYB15 regulate stilbene biosynthesis in *Vitis vinifera*. Plant Cell 25(10):4135-4149 (2013)
53. Ong B.Y. & Nagel C.W. Hydroxycinnamic acid tartaric acid ester content in mature grapes and during maturation of white riesling grapes. Am J Enol Viticult. 29(4):277-281 (1978)
54. Walker A.R., et al. White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant J. 49:772 (2007)
55. Forcato C. Gene prediction and functional annotation in the *Vitis vinifera* genome. PhD Thesis, Universita' Degli Studi Di Padova (2010)

Chapter 10

Very low-expressed genes throughout berry development

INTRODUCTION

RNA sequencing (RNA-Seq) is a powerful tool for quantifying gene expression[1-3]. However, RNA-Seq produces a wide range of read counts per gene, and genes with a low coverage of reads can produce artificially high fold-change values[4]. In **chapter 4**, we evaluated that a reliable FPKM threshold for our data set could be 1 FPKM. Therefore, the 5,061 genes showing maximal expression intensity among the 40 samples below 1 FPKM were classified as very low-expressed.

In this work, we gained insights into the functional category distribution and biological processes, which the very low-expressed genes were involved in. Furthermore, the unsupervised Principal Component Analysis revealed different patterns among sample distribution.

RESULTS AND DISCUSSION

Functional category distribution and Gene Ontology enrichment of the very low-expressed genes

In order to gain an insight into biological functions of the 5,061 very low-expressed genes during berry development, transcripts were annotated against the V1 version[5] of the 12x draft annotation of the grapevine genome (http://genomes.cribi.unipd.it/DATA/). Transcripts were then grouped into the sixteen most represented functional categories (**Fig. 1**). We observed that several transcripts belonged to the cellular process, the response to stress, the secondary metabolic process, the signal transduction, the transcription factor activity and the transport categories.

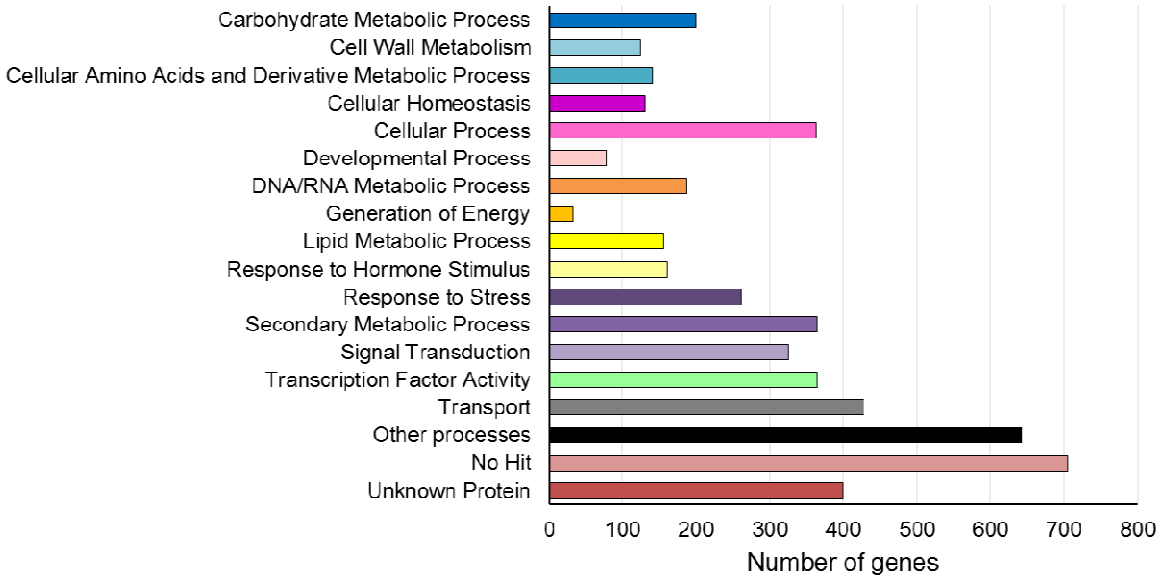


Figure 1: Functional category distribution of the 5,061 transcripts with mean FPKM values lower than 1 among all the 40 samples. Transcripts were grouped into the eighteen most represented functional categories, based on Plant GO Slim biological processes classification.

In order to find the biological processes hidden bellow the functional categories, these 5,061 transcripts were analysed for overrepresented functions using the GO enrichment tool BiNGO[6] and the GoSlim Plants annotation (**Fig. 2**). Four GO categories were classified as significantly overrepresented: ripening (GO:0009835), protein modification process (GO:0006464), pollen-pistil interaction (GO:0009875) and pollination (GO:0009856). The overrepresentation of the pollen-pistil interaction and pollination categories indicates, not surprisingly, that genes involved in these processes were very low expressed during berry development, suggesting either a leftover transcripts present before fecundation or a basal expression of these genes. Furthermore, the enriched protein modification process GO categories included many kinases considered as pollen-pistil interaction- and pollinisation-related as S-receptor kinases[7]. Regarding the overrepresented ripening category, three polygalacturonase GH28 genes (VIT_01s0127g00120; VIT_13s0064g00750; VIT_13s0064g00780) and four polygalacturonase PG1 genes (VIT_08s0007g07750; VIT_08s0007g07760; VIT_08s0007g07770; VIT_08s0007g07780) were identified. Polygalacturonase and pectin methylesterase are two of the most significant pectin-degrading enzymes during ripening[8]. This result indicates that these members of the polygalacturonase family might show only a basal expression during berry development and that other genes play a more remarkable role in this process. Moreover, the four polygalacturonase PG1

sequences were places one next to each other along chromosome 8, suggesting an eventual epigenetic regulation.

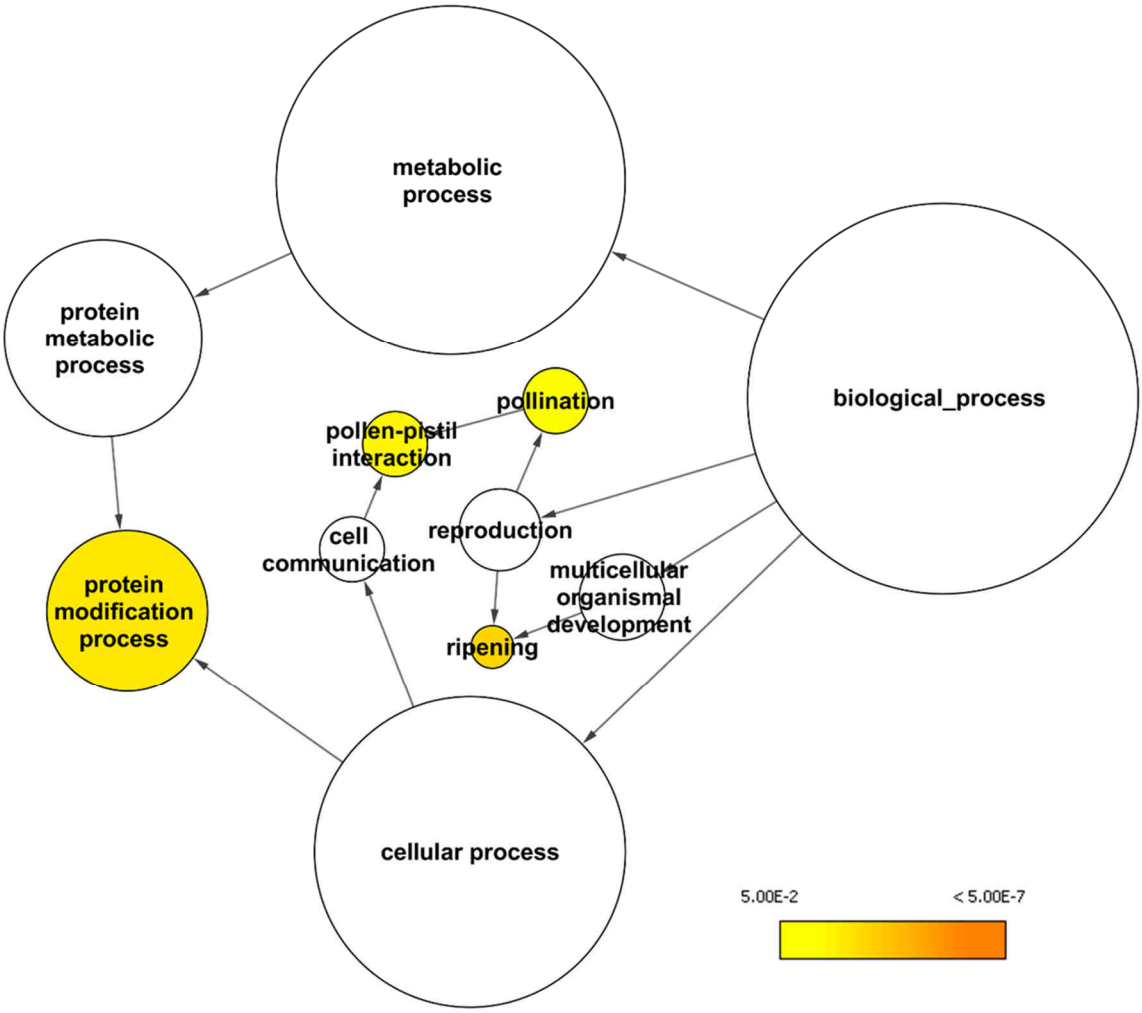


Figure 2: Enriched GO terms in the 5,061 very-low expressed genes.
The network graph shows BiNGO visualizations of the overrepresented GO terms for the 5,061 very-low expressed genes during berry development. Node size is positively correlated with the number of genes belonging to the category. Non-coloured nodes are not overrepresented, but they may be the parents of overrepresented terms. Coloured nodes represent GO terms that are significantly overrepresented (Benjamini and Hochberg corrected p-value < 0.05), with the shade indicating significance as shown in the colour bar.

Low-transcriptome relationships

For simplicity, this group of very low expressed genes will be called low-transcriptome. In order to evaluate the relationships between low-transcriptomes, namely all the expression values of the 5,061 very low-expressed genes in each berry sample, a

Principal Component Analysis (PCA) was performed on the data set composed of 40-sample FPKM values (**Fig. 3**).

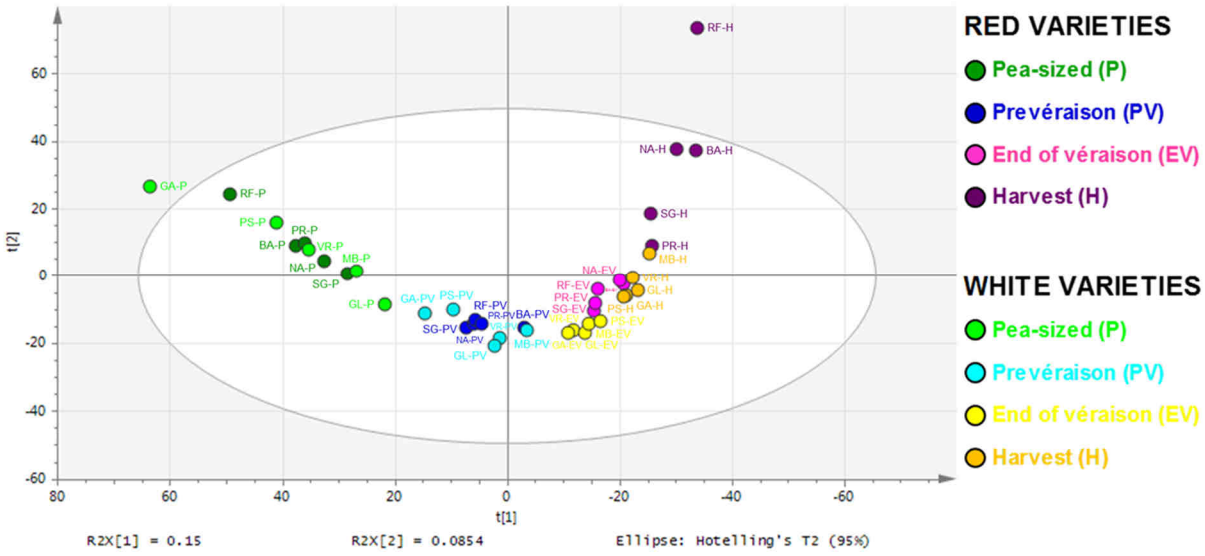


Figure 3: Plot of 40-sample data set PCA showing the first two principal components t[1]=15.0% and t[2]=8.54%.

The dataset used for this analysis was composed of 5,061 transcripts corresponding to the genes having a FPKM value inferior to 1 in all samples.

Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest) and for each variety: SG (Sangiovese), BA (Barbera), NA (Negro amaro), RF (Refosco), PR (Primitivo) for the red grape varieties and VR (Vermentino), GA (Garganega), GL (Glera), MB (Moscato bianco), PS (Passerina) for the white grape varieties.

The first two principal components (PC) explained 23.5% of the total transcriptional variance. Compared to the PCA performed with the 21,746 genes having expression intensity greater than 1 FPKM (**chapter 5**), PC1 could again explain the time course of berry development. Nevertheless, PC1 did not clearly separate pre- and post-véraison samples. Moreover, PC1 did not permit the grouping of the pea-sized berry samples. In addition, PC2 did not distinguish clearly pea-sized berry stage from prevéraison stage low-transcriptomes, whereas end of véraison and harvest stages still slightly overlapped. PC2 seemed again to describe a separation between white- and red-skinned berry samples starting from the end of véraison and more evidently at harvest with Barbera, Negro amaro and Refosco full-ripe berry transcriptomes separated from the other harvest transcriptomes. Thus, these results suggest that pea-sized berry low-transcriptomes had different relationship patterns than pea-sized berry transcriptomes

and might have influenced the results if they had not been removed them from the data set. In fact, even if the transcriptome patterns were similar on the plot of the PCA performed with the 26,807-transcript data set – including all the genes having a FPKM value above 0 – the first two PCs explained 40% of the total transcriptional variance instead of 44.3% with the 21,746-transcript data set – without low-transcriptomes (**chapter 5**). In addition, keeping low-transcriptomes could have slightly decreased the cumulative variance of the two principal components of the O2PLS-DA model used for the biomarker discovery (**chapter 5**) – from 42.4% to 38.7% – and then might potentially have screened the discovery of well-expressed biomarker transcripts.

CONCLUSION

Very low-expressed genes were mostly involved in the cellular process, the response to stress, the secondary metabolic process, the signal transduction, the transcription factor activity and the transport categories. The GO enrichment analysis allowed identifying four significantly overrepresented GO categories: the ripening, the protein modification process, the pollen-pistil interaction and pollination. Therefore, we suggested the detection of the possible residual expression of genes involved in pollen-pistil interaction and pollination or their constitutive basal expression at a very low intensity. Regarding the overrepresented ripening GO category, the low expression level of some polygalacturonase genes could be due to the epigenetic regulation of some member of the gene family, which might be rather more active in other tissues or conditions concerning cell wall polymers modification. Finally, PCA revealed that pea-sized berry low-transcriptomes had different relationship patterns than pea-sized berry transcriptomes, indicating that very-low expressed genes could have slightly influenced the results if they had not been removed from the data set.

METHODS

Functional Category Distribution and GO Enrichment Analysis

All 5,061 transcripts were annotated against the V1 version of the 12x draft annotation of the grapevine genome (http://genomes.cribi.unipd.it/DATA/). This was verified manually and integrated using gene ontology (GO) classifications. Transcripts were then grouped into the sixteen most represented functional categories: carbohydrate metabolic process (GO:0005975), cell wall metabolism (GO:0044036), cellular amino

acids and derivative metabolic process (GO:0006520), cellular homeostasis (GO:0019725), cellular process (GO:0009987), developmental process (GO:0032502), DNA/RNA metabolic process (GO:0090304), generation of energy (GO:0006091), lipid metabolic process (GO:0006629), response to hormone stimulus (GO:0009725), response to stress (GO:0006950), secondary metabolic process (GO:0019748), signal transduction (GO:0007165), transcription factor activity (GO:0051090), transport (GO:000681), other processes (GO:0008150) based on GO biological processes. Genes with unknown functions or with a “no hit” annotation were also included.

GO enrichment analysis was performed by using BiNGO 2.4 plug-in tool in Cytoscape version 3.1.0 with PlantGOslim categories, as described by Maere *et al.*[6]. Overrepresented PlantGOslim categories were identified using a hypergeometric test with a significance threshold of 0.05 after a Benjamini & Hochberg False-Discovery Rate (FDR) correction[9].

REFERENCES

1. Garber M., et al. Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Meth. 8(6):469-477 (2011)
2. Oshlack A., Robinson M., Young M. From RNA-seq reads to differential expression results. Genome Biol. 11(12):220 (2010)
3. Wang Z., Gerstein M., Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 10(1):57-63 (2009)
4. Warden C.D., Yuan Y.C., Wu W. Optimal Calculation of RNA-Seq Fold-Change Values. International Journal of Computational Bioinformatics and In Silico Modeling 2(6):285-292 (2013)
5. Forcato C. Gene prediction and functional annotation in the Vitis vinifera genome. PhD Thesis, Universita' Degli Studi Di Padova (2010)
6. Maere S., Heymans K., and Kuiper M. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21:3448-3449 (2005)
7. Li H.Y., Gray J.E. Pollination-enhanced expression of a receptor-like protein kinase related gene in tobacco styles. Plant Mol Biol. 33(4):653-65 (1997)
8. Deytieux-Belleau C., Vallet A., Donèche B., Geny L. Pectin methylesterase and polygalacturonase in the developing grape skin. Plant Physiol Biochem 46(7):638-46 (2008)
9. Klipper-Aurbach Y., Wasserman M., Braunspiegel-Weintrob N., Borstein D., Peleg S., Assa S., Karp M., Benjamini Y., Hochberg Y., and Laron Z. Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. Med. Hypotheses 45:486-490 (1995)

Chapter 11

Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development

Maria Concetta Palumbo, Sara Zenoni, Marianna Fasoli, **Mélanie Massonnet**, Lorenzo Farina, Filippo Castiglione, Mario Pezzotti, and Paola Paci

A modified version of this chapter has been published in Plant Cell 26:4617-4635 (2014).

INTRODUCTION

Fruit development and ripening occur during the growing season, which involves a double-sigmoid pattern of growth featuring unique developmental, physiological, and biochemical processes that influence the colour, texture, flavour, and aroma of the berries[1]. The onset of ripening is called véraison, and it represents a striking metabolic transition phase, not only for the berry but also for other clustered organs such as the rachis and seeds. Harvest usually marks the end of ripening, but in some cultivars, this is followed by a period of postharvest withering for the production of particular styles of wine. The economic importance of grapevine has prompted the investigation of molecular factors that regulate growth, development, berry ripening, and particularly the impact of gene expression on quality traits[2]. Recent analysis of the grapevine global gene expression atlas[3] revealed a clear distinction between vegetative/green and mature/woody sample transcriptomes, suggesting a fundamental shift in global gene expression as the plant switches from the immature to the mature developmental program. Elucidating this global transcriptomic reprogramming, which can be considered a characteristic feature of perennial fruit crops, represents a significant step toward the large-scale characterization of genes governing the key developmental and metabolic processes in grapevine.

Tomato (*Solanum lycopersicum*) is another valuable fruit crop due to the nutritional properties of the ripe berries[4]. Like grape berries, tomato fruits become edible after a complex ripening process involving changes in physical and biochemical properties, characterized by a clear transition from a vegetative to a mature phase. However, unlike grapevine, tomato produces fruits that can be used as a model system to study the molecular basis of fruit development and ripening thanks to the availability of efficient transformation methods[5] and well-characterized ripening mutants[6-8]. The characterization of tomato development has been facilitated by the recently completed genome sequence and transcriptome map[9] (Tomato Genome Consortium, 2012). Ripening in fleshy fruit involves profound changes in metabolism of the tissue surrounding the seed, and both the tomato and grapevine transcriptomes are characterized by a transition from immature to mature development[10]. A few genes induce this overall transcriptome reprogramming at the onset of ripening, and gene network analysis offers a powerful tool to unveil the key players. Unlike statistical approaches that focus on global properties of groups of genes as separate entities

(clustering methods, principal component analysis, t test, F-test, or nonparametric versions of the Wilcoxon test and Kruskal Wallis test), network analysis is a global approach that also takes into account relationships among genes. There are two primary methods used to infer gene networks: Bayesian networks[11,12] and gene coexpression networks (weighted or unweighted)[13]. Bayesian networks are created by finding all of the causal relationships among gene expression levels, whereas coexpression networks assume all genes are connected and their strength of connection is quantified by the correlation between expression profiles of gene pairs. Unlike Bayesian networks, gene coexpression networks are undirected. Gene coexpression network analysis is based on the concept of scale-free networks that are in turn founded on connectivity, which describes the relative importance of each gene in the network and reflects how frequently a node interacts with other nodes (for a weighted network, connectivity is defined as the sum of the weights across all edges of a node). According to node connectivity, genes can be further classified as hub genes (with an extremely high level of connectivity) and non-hub genes. Hubs have proven useful for the characterization of key biological properties such as lethality, modularity, hierarchical organization, and robustness[14,12]. Because protein interactions are regulated dynamically both in time and space, two fundamental types of hubs have been identified in protein-protein interaction (PPI) networks[15]: “party hubs,” which interact with most of their partners simultaneously, and “date hubs”, which bind their different partners at different times or locations. Another important aspect of gene coexpression networks is modularity: highly connected genes in the network are usually involved in the same biological modules or pathways[15,16]. In order to study modularity in complex networks, a new approach known as cartography methodology[17] has been developed recently, in which nodes are classified into sets of predefined roles depending on their between/within-module connections. Here, we integrated the cartography[17] and date/party[15] approaches in a single method used to analyse coexpression networks. We investigated the deep transcriptome shift that occurs in grapevine, corresponding to the immature-to-mature transition in most organs and tissues[3]. This method allowed us to classify genes according to the extent of correlation with interaction partners. We also identified a class of genes that we termed “fight-club genes,” reflecting their predominantly negatively correlated profile, and within this group a subclass that we termed “switch genes,” which are likely to be key players in large-scale transcriptomic transitions. The same procedure was also applied to a new

RNA-Seq data set representing the pre-véraison and post-véraison developmental phases of five red-berry grapevine varieties, in order to identify switch genes potentially involved in the transition from immature to mature berry development. To demonstrate proof of concept, we also applied the same procedure to two different tomato expression datasets representing different organs (Tomato Genome Consortium, 2012) and a ripening-deficient mutant[9]. The identification of known tomato master regulators of fruit maturation allowed us to predict the efficiency with which our method can help to identify key regulators of organ development in fleshy fruit crops. As a final comment, we note that our methodology is quite general, and it can be used to analyse, in principle, any gene expression data set. We speculate that it can also be used to identify key regulators in different conditions and organisms.

RESULTS AND DISCUSSION

Network analysis of genes involved in the immature-to mature transition during grapevine development reveals an important type of hub

A previous correlation analysis covering the comprehensive grapevine transcriptome map, which represents the expression of 29,549 genes in 54 different samples, revealed a clear separation between samples representing vegetative/green and mature/ woody tissues[3]. This indicated the existence of a major regulatory switch that promotes the transition from the immature to the mature developmental phase. Key players in this transcriptome reprogramming event were sought by analysing two groups of samples, 25 from mature/woody organs and 27 from green/vegetative organs. Two further samples (pollen and leaves undergoing senescence) were excluded due to their atypical transcriptomes.

We identified 1,686 genes showing significant differential expression between vegetative/green and mature/woody samples. Among them, 1,220 genes were down-regulated but only 466 were up-regulated during the developmental transition.

To provide an overview of their roles, the differentially expressed genes (DEGs) were classified on the basis of a manually improved functional annotation[18]. The significantly overrepresented ($P < 0.05$) functional categories identified for down-regulated and up-regulated genes showed that maturation involves the suppression of diverse metabolic processes, including photosynthesis, energy metabolism,

carbohydrate metabolism, cellular component organization, and the cell cycle, which are related to vegetative growth. In contrast, only a few pathways were specifically induced, including secondary metabolic processes and responses to biotic stimuli. This evidence suggests that the transition to mature growth predominantly involves the suppression of vegetative pathways rather than the activation of mature pathways.

Any of the DEGs discussed above could be directly involved in the regulation of transcriptome reprogramming during the developmental transition to mature growth or the véraison step in berry development. In order to identify potential master regulators of this phase transition, we generated a coexpression network using Pearson correlation as a distance metric. Nodes in the coexpression network represent genes, and the presence of an edge linking two genes means that the correlation between their expression profiles exceeds a given threshold in terms of absolute value. The coexpression network comprised 1,660 nodes and 118,784 edges (**Fig. 1A**).

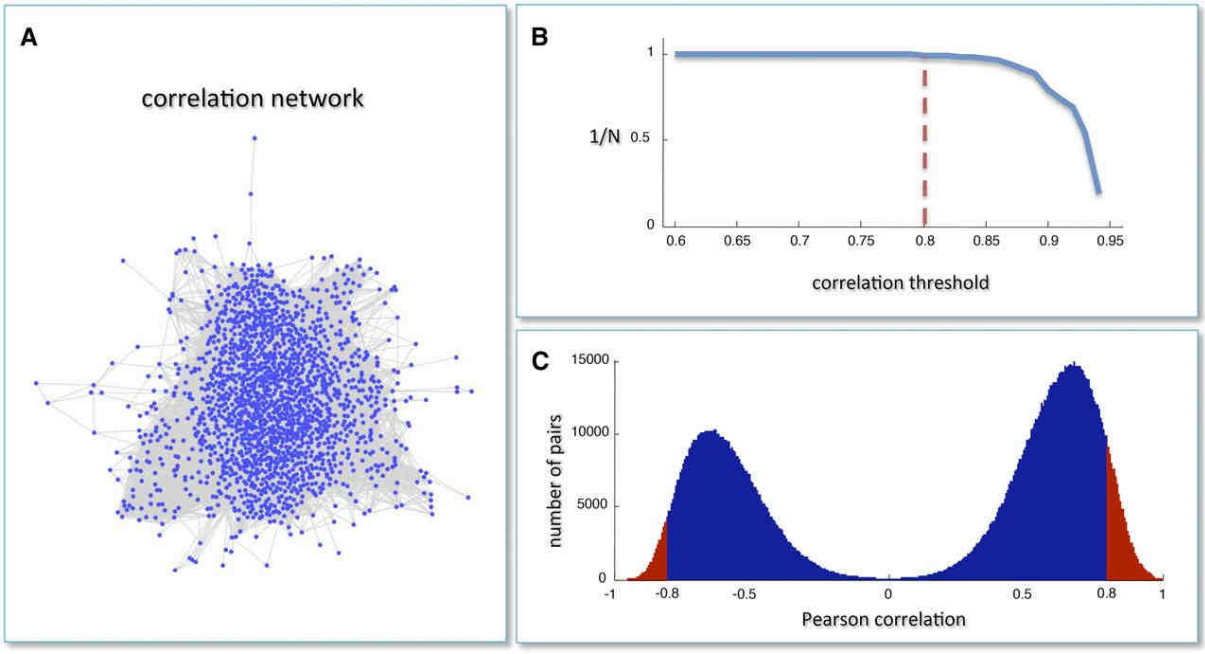


Figure 1: Grapevine Coexpression Network.

(**A**) The correlation network obtained using the grapevine gene expression atlas[3]. Nodes are mRNAs, and a link is present when the absolute value of the Pearson correlation between the expression profiles of two mRNAs is greater than 0.8.

(**B**) The inverse of the number N of network-connected components is plotted against the correlation threshold. The value 0.8 is the largest threshold value that maintains network integrity.

(**C**) Pearson correlation distribution of all mRNA profile pairs. Red regions correspond to the most correlated pairs in the network.

We estimated the number of strongly connected components for different correlation values and predicted that the Pearson correlation threshold of 0.8 would largely maintain network integrity (**Fig. 1B**). The Pearson correlation distribution of all mRNA profile pairs revealed a clear bimodal profile (**Fig. 1C**). This supported the presence of a large group of inversely correlated profiles, contrasting with the typical unimodal distribution but in agreement with the gene/tissue bicluster that was found to mirror the transcriptome reprogramming that occurs during the developmental transition[3].

We next searched for specific topological properties of the coexpression network using the date/party hub classification system based on the average Pearson correlation coefficients (APCCs) between the expression profiles of each hub and its nearest neighbours[15]. The extent to which hubs are coexpressed with their interaction partners leads to two classifications with characteristic topological properties, i.e., date hubs (low APCC) and party hubs (high APCC). Date hubs have a coordinating role within the network, whereas party hubs act as local hubs[15]. Interestingly, we found that the distribution of APCCs was trimodal, with two peaks representing low and high positive APCC values (mirroring the date and party hub distributions in PPI networks[15]; and a third peak representing negative APCC values (**Fig. 2**). This third peak revealed the presence of a class of hubs that we named fight-club hubs because they are characterized by an average inversely correlated profile with their partner genes. We identified 517 party hubs, 797 date hubs, and 234 fight-club hubs in the grapevine transcriptome. In order to show that the data/party dichotomy is also observed in coexpression networks and in particular that the trimodal behaviour of the APCCs was not obtained by chance but is a peculiar feature of these coexpression networks, we calculated the distribution of APCCs in a randomized network generated by keeping node labels constant while shuffling their edges but preserving the degree of each node[15]. The resulting distribution was unimodal with a peak equivalent to a moderate positive APCC value of ~0.4 (**Fig. 2**). This positive value reflects the predominance of positive compared with negative correlations in the network.

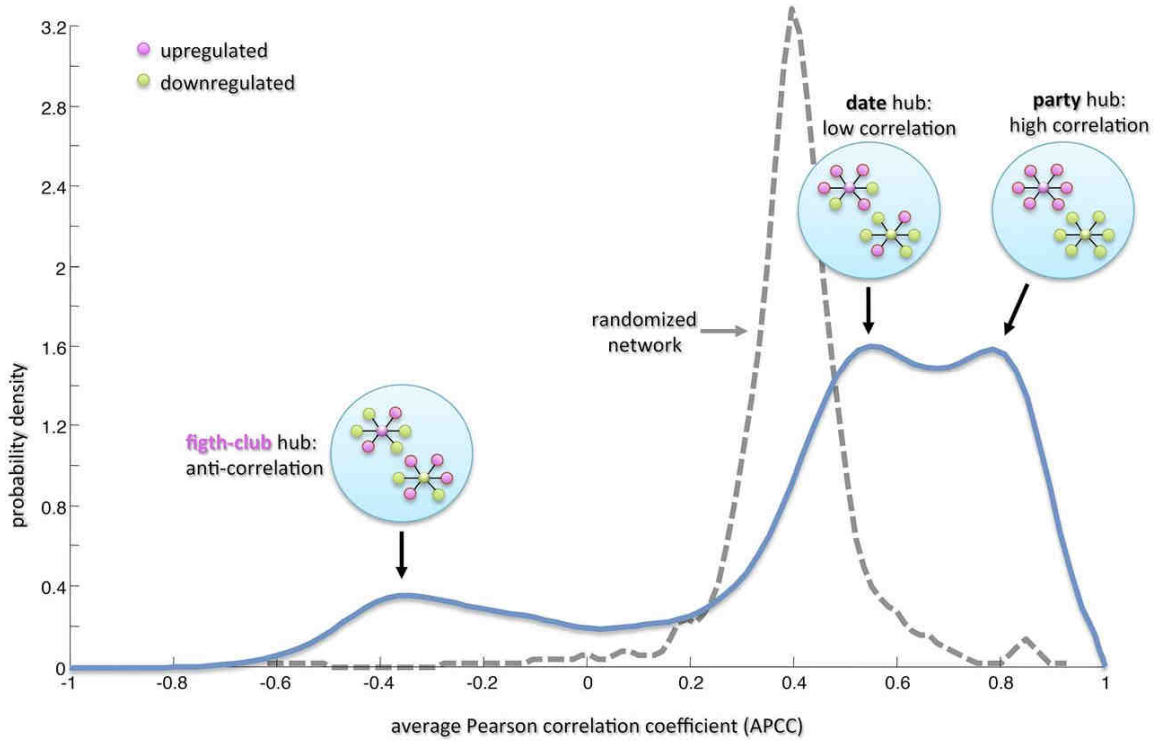


Figure 2: Average Pearson correlation coefficient in grapevine.

The blue curve is the estimated probability density using a smoothing algorithm with a Gaussian kernel of the APCC for each hub (i.e., node with degree greater than 5) in the correlation network. The distribution of APCCs is trimodal, with two peaks representing low and high positive APCC values (mirroring the date and party hub distributions in PPI networks[15]) and a third peak representing negative APCC values (fight-club hubs). Circular insets illustrate schematic hub interactions: Party hubs are highly correlated with the expression of their partners (i.e., they are coloured as their partners), date hubs show more limited coexpression (i.e., they are mostly coloured as their partners), and fight-club hubs show mostly negative correlation with their partners (i.e., if the fight-club hub is up-regulated [magenta], then most of its partners are down-regulated [green] or vice versa). The dashed grey curve represents the APCC for the hubs in a randomized network generated by keeping node labels constant while shuffling their edges but preserving the degree of each node.

Heat cartography in grapevine reveals switch genes as network bottlenecks

The bimodal correlation distribution for grapevine genes that are differentially expressed between vegetative/green and mature/ woody tissues reveals the intrinsic modular structure of the coexpression network, wherein each module can be regarded as a subnet work with a distinct structure and function. The relationship between structure and function within this coexpression network was investigated by k-means clustering[21-25] leading to the identification of five modules. Node classification can also be reconsidered in terms of topological roles defined by intramodule and intermodule

connections[17]. Usually, topological role assignment is based on the computation of two parameters for each node: the within-module degree *z* and the participation coefficient *P* [17]. We modified these parameters (renamed z_g and K_π, respectively) to identify genes representing the immature-to-mature transition. The first parameter is a normalized measure of intramodule communication, whereas the second characterizes the mode of communication between nodes in different modules. As previously discussed[17], the plane is divided into seven regions each defining a specific node role (**Fig. 3**). High z_g values correspond to nodes that are hubs within their module (local hubs), whereas low z_g values correspond to nodes with few connections within their module. Nodes characterized not only by low z_g values but also by high K_π values show many connections outside their module and, thus, a high node degree. According to a hub definition based on the node degree, these nodes can be classified as global hubs.

These nodes fall into the so-called R4 region of the plot (**Fig. 3**), showing that even crucial nodes can populate this region. By assigning a colour proportional to the APCC value of each node[25], we obtained a heat cartography map of the modular network in which party, date, and fight-club hubs were easily identified by red, orange, and blue colouring, respectively (**Fig. 3**). This map revealed that many fight-club hubs form a small and tightly localized group in the R4 region, characterized by z_g values lower than -2 and high K_π values, describing nodes that are highly connected outside their module. This group comprises 113 genes that we classified as switch genes because they are more strongly linked to inversely correlated rather than positively correlated genes in the network. By drawing the heat cartography for the nodes of the grapevine randomized network, we found a less populated R4 region. However, we observed a predominance of positive correlation and an absolute absence of switch genes.

The importance of nodes within a complex network can be determined by evaluating their topological resilience to random breakdown (failure) and to the removal of key hubs (attack)[14]. Resilience can be evaluated by observing the effect of failure or attack on the Characteristic Path Length (CPL), which is the shortest path connecting each node to other nodes in the network. The removal was implemented by ranking nodes with decreasing degree values and then deleting the first *N* nodes, where *N* is the number of the switch genes. In this way, we evaluated how the removal of *N* switch genes and *N* hubs affects the CPL. Removing a crucial node (such as a hub) in a scale-free network may cause fundamental edges to break down, thus increasing the CPL.

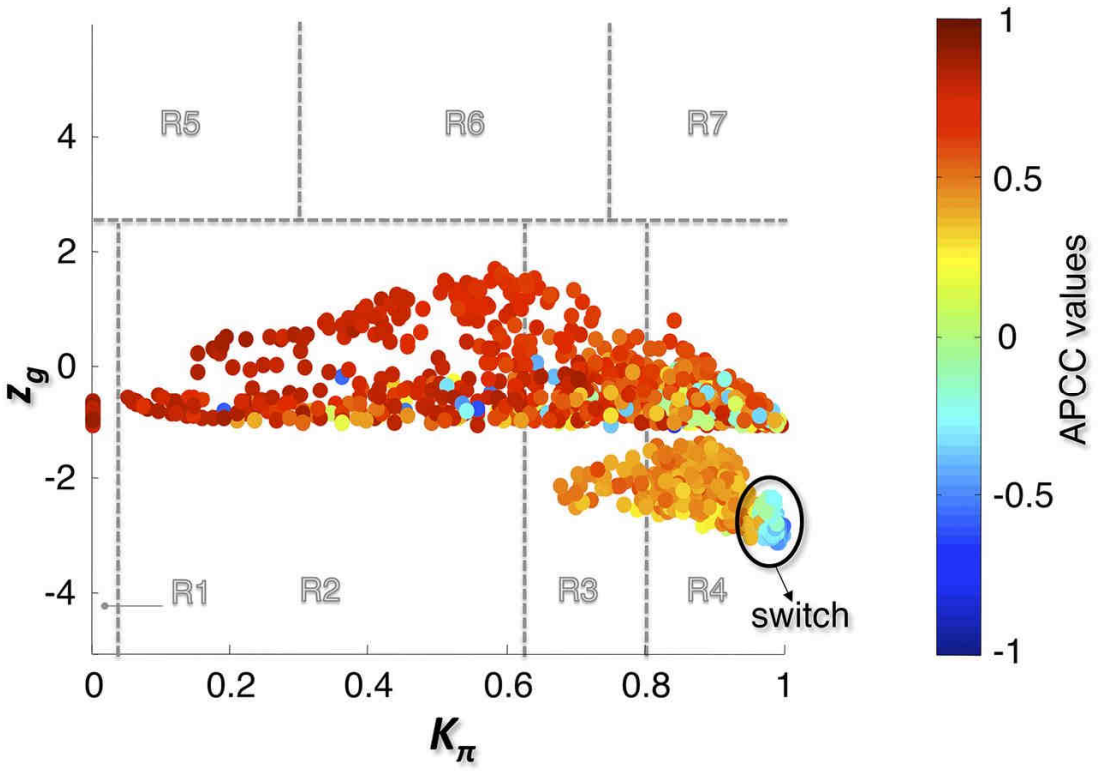


Figure 3: Grapevine atlas heat cartography map.

The parameters z_g and K_π represent a normalized measure of intramodule communication and the mode of communication between nodes in different modules, respectively. The plane identified by these two parameters is divided into seven regions each defining a specific node role. Each point represents a node in the correlation network of Figure 1A, and the colour of the each node corresponds to its APCC value. Roles have been assigned to each node of the correlation network of Figure 1A, according to the heat cartography (each circle corresponds to a node in the network which has been coloured according to its APCC value).

A scale-free network is characterized by a power law distribution of nodal connectivity, $p(k) = k^{-\gamma}$, where γ is a parameter whose value is typically in the range $2 < \gamma < 3$ [26]. Thus, the main peculiar feature of a scale-free network is the existence of a few highly connected nodes (hubs) and a lot of poorly connected nodes (non-hubs). In such networks, the removal of standard nodes does not increase the CPL, whereas the removal of one of the few key hubs results in a net increase of the network diameter. Scale-free networks therefore appear to be robust against accidental failures but vulnerable to coordinated attacks (i.e., against hubs)[14]. After verifying that the grapevine coexpression network is scale-free, we evaluated its resilience in the face of attack, failure, and the breakdown of the key nodes (switch genes). The removal of switch genes had a profound impact on network connectivity by substantially

increasing the CPL (**Fig. 4**). This suggests that switch genes are key players in the transcriptional network and play a critical role in plant development.

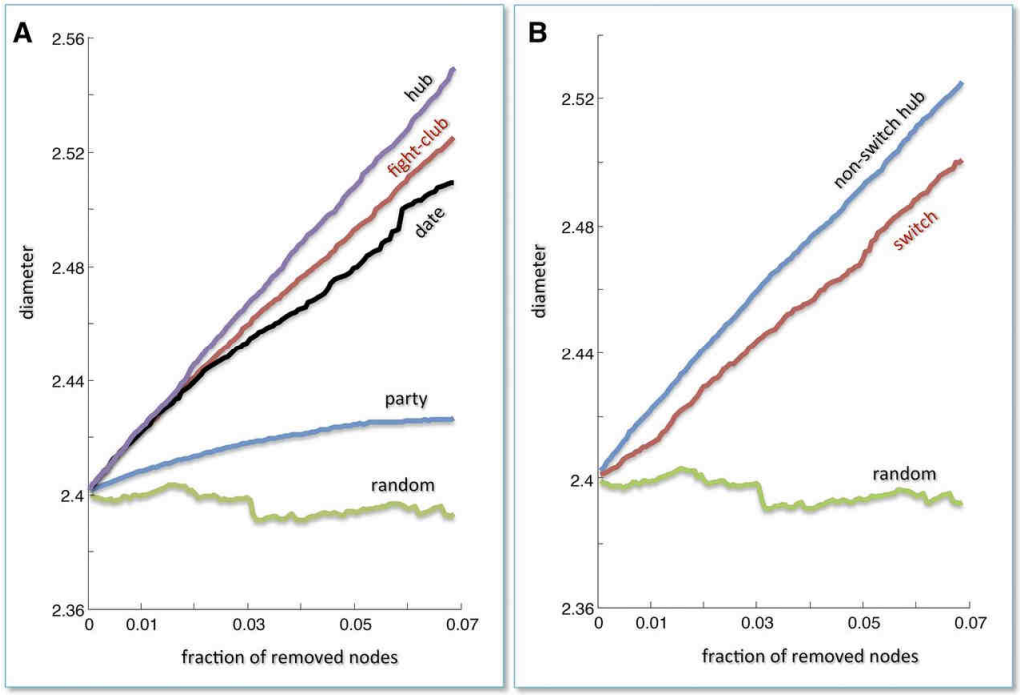


Figure 4: Error and attack tolerance of the grapevine atlas coexpression network.
Both panels show changes in the CPL, which is the shortest path connecting each node to other nodes in the network, as a function of the fraction of removed nodes. The total number of removed nodes is 113, which is the number of switches.
(**A**) Comparison of hub removal (attack) and random removal (failure) on party, date, and fight-club hubs.
(**B**) Comparison of hub removal that are non-switch genes, random removal (failure) on switch genes. Removing a crucial node (such as a hub) may cause fundamental edges to break down, thus increasing the CPL. The removal of switch genes (red curve) appears to be comparable with the removal of hubs (not including switch genes), highlighting their importance in the network.

Switch genes are negatively correlated to vegetative genes and may represent putative microRNA targets

We found that all 113 switch genes were expressed at a low level in vegetative/green tissues but at significantly higher levels in mature/woody organs, suggesting they participate in the regulation of the transition from immature to mature development. We assigned functional annotations to 86 of the switch genes and found that they covered diverse functions, with only secondary metabolism significantly overrepresented, accounting for 14 of the genes ($P < 0.05$). Fifteen other genes encoded transcription

factors, and nine were assigned a role in carbohydrate metabolism, suggesting these functions are particularly important during the developmental transition. Transcription factors often regulate developmental processes and would be expected to account for many switch genes, so it is surprising that many switch genes encode enzymes.

The switch genes were also connected to 921 other genes, representing more than the 50% of the entire coexpression network. The most connected switch gene was linked to 469 other genes, whereas the least connected was linked to 53 others. All 921 of the neighbouring genes showed the opposite expression profile to that of the switch genes. The overrepresented functional categories ($P < 0.05$) of neighbouring genes were identical to those revealed for down-regulated genes during the immature-to-mature developmental transition. These processes are typically associated with vegetative growth, suggesting the genes are required during the immature phase and are switched off during the transition to mature growth.

The well-established role of microRNAs (miRNAs) in the regulation of plant development suggested that the switch genes might include potential miRNA targets. To test this hypothesis, we screened the switch genes with the psRNAtarget tool (http://plantgrn.noble.org/psRNATarget/) and duly identified 49 switch genes that may be regulated by miRNAs. We also identified that in the entire grapevine predicted transcriptome, 540 transcripts may be miRNA targets. By applying a hypergeometric test, we found that the list of switch genes was statistically significant enriched in miRNA targets ($P < 0.001$).

The 49 switches that are regulated by miRNAs belong to diverse functional categories. Interestingly, among the putative miRNA:mRNA pairs, we identified Vv-miR164:NAC33, whose interaction has already been experimentally validated in grapevine[27].

Identification of switch genes representing the immature-to-mature transition in grapevine berry development

We applied our pipeline, which consists of identifying fight-club hubs and then applying the heat cartography approach to highlight switch genes, to a new transcriptomic data set, focusing on the berry. In order to identify key genes in the transition from immature to mature development, we used RNA-Seq analysis to generate berry transcriptome profiles from five grapevine red-berry varieties during the two growth phases. Berries

were collected at four phenological stages: two representing the immature phase and two representing the maturation phase.

By comparing transcriptomic profiles of berries during the immature and maturation phases, we identified 1,883 genes showing significant differential expression. Among them, 1,464 genes were down-regulated but only 419 were up-regulated after the developmental transition. Significantly overrepresented categories ($P < 0.05$) among the down-regulated and up-regulated genes showed that the transition from green/immature to ripe/mature berry development involved the modulation of different metabolic processes from those described in the atlas. In particular, photosynthesis, energy metabolism, lipid metabolism, transcriptional regulation, responses to endogenous stimuli, and responses to abiotic stress were down-regulated after véraison, whereas only stress response pathways (particularly the response to biotic stress), secondary metabolism, and carbohydrate metabolism were up-regulated. These results were also obtained by studying the common DEGs throughout development (**chapter 6**), the putative stage- and phase-specific positive biomarker transcripts (**chapter 5**) and the genes belonging to the berry development transcriptomic route (**chapter 7**), suggesting again that the shift to maturation in the berry predominantly involves the suppression of vegetative pathways rather than the activation of maturation-specific pathways.

In order to identify key genes regulating the metabolic shift in berry development, we generated a coexpression network comprising 1,792 nodes and 333,700 edges. The threshold correlation was chosen as a balance between network integrity and the smallest number of links. As observed for the atlas data set, the Pearson correlation distribution of all mRNA profile pairs revealed a clear bimodal profile, supporting the existence of a large group of inversely correlated profiles.

We next searched for specific topological properties of the coexpression network, and again we identified three types of hubs, specifically 1,344 party hubs, 80 date hubs, and 251 fight-club hubs. We used k-means clustering to identify three modules, and by assigning a colour proportional to the APCC value of each node, we obtained a heat cartography map of the modular network in which the switch genes were easily identified (**Fig. 5**). Switch genes disappear in the randomized network. In this case, we identified 190 switch genes that are likely to be involved in the regulation of the grape berry developmental transition.

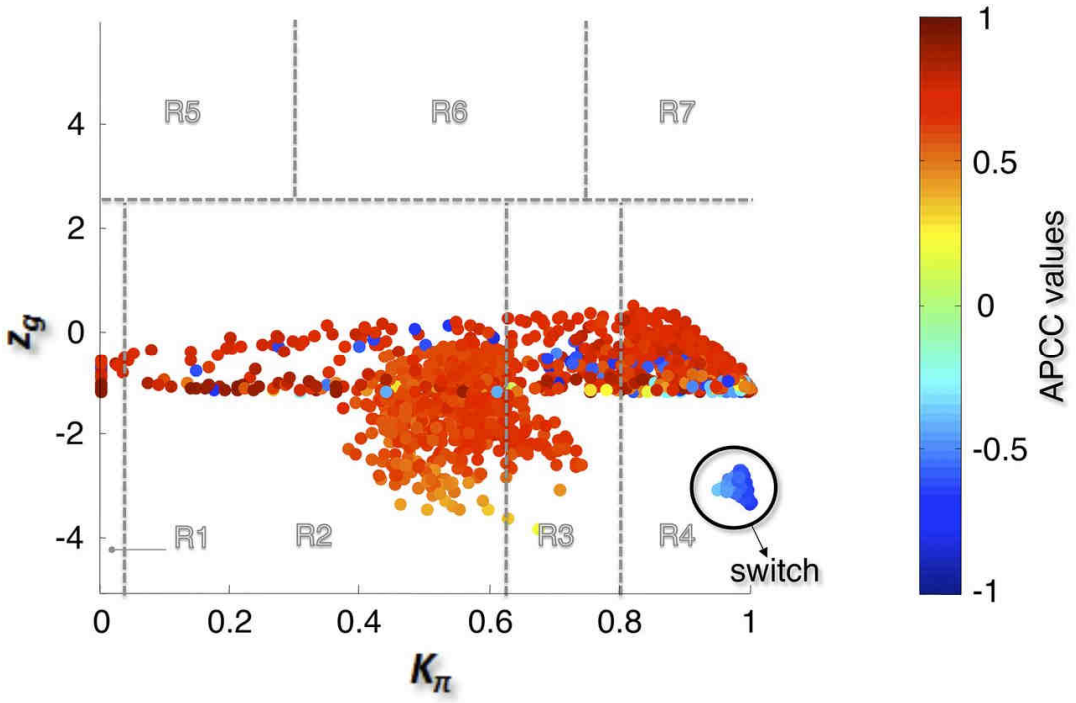


Figure 5: Heat cartography map for the grapevine berry transcriptome of five red-berry varieties. The parameters z_g and K_π represent a normalized measure of intramodule communication and the mode of communication between nodes in different modules, respectively. The plane identified by these two parameters is divided into seven regions each defining a specific node role. Each point represents a node in the correlation network and the colour of each node corresponds to its APCC value. Roles have been assigned to each node, according to the heat cartography described in Methods.

There are many common switch genes in the grapevine atlas and berry transcriptomes

All 190 switch genes in the berry transcriptome were expressed at a low level during the immature phase and were significantly induced at véraison (**Supplementary table 1**). Indeed, among the switch genes, 13 and 5 were identified as putative maturation phase- and harvest stage-specific positive biomarkers (**chapter 5**). Moreover, 18 switch genes belong to cluster 5 of the berry development transcriptomic route (**chapter 7**) – genes belonging cluster 5 exhibit an increasing gene expression pattern throughout development. By assigning functional annotations to 156 of the switch genes, we found they covered diverse functions but that transcription factor activity, cell wall metabolism, and developmental process were significantly overrepresented ($P < 0.05$).

The berry developmental switch genes included many that have already been characterized for their major role during berry ripening. For example, we identified

switch genes encoding the transcription factors MYBA1, MYBA2, and MYBA3, which induce anthocyanin synthesis at the onset of the ripening by activating the UDP-glucose:flavonoid-3-O-glucosyltransferase gene *UFGT* [29-31]. We also identified *GST4* (glutathione S-transferase 4), which is also induced by the transcription factors listed above and which contributes to anthocyanin transport to the vacuole[32,33]. Further switch genes encoded thaumatin and a major latex-like ripening protein (grip61) involved in stress responses, which are both considered to be markers of the onset of the ripening process[34]. The switch genes also included ethylene responsive factors (ERFs), such as *ERF019* and *ERF045,* the latter of which is a putative harvest stage-specific positive biomarker transcripts (**chapter 5**) and known to be up-regulated during berry ripening[35]. *EXPB4* (expansin-like B4) was also classified as switch gene, confirming previous reports of its specific transcriptional activation at the onset of ripening and its putative role in berry softening caused by cell wall modifications[36]. The identification of transcription factors and enzymes with fundamental roles in berry development at the onset of ripening strongly support the suitability of our approach and indicates that switch genes represent master regulators of development and their direct targets, thus comprising a core function that allows plant organs to complete developmental transition. The berry switch genes were also connected to 1,266 other genes, representing ~70% of the entire coexpression network. Among these neighbouring genes, 82 and 325 were identified as putative pea-sized berry stage- and green phase-specific biomarker transcripts respectively in **chapter 5**. In fact, all the 1,266 neighbouring genes were found down-regulated between the immature and maturation phases, suggesting that they could represent good candidates as direct targets of the transcription factors identified as switch genes. As for the grapevine atlas, the overrepresented functional categories of neighbouring genes ($P < 0.05$) were identical to those of down-regulated genes after véraison.

Next, we compared the grapevine berry and atlas data sets to compile a list of shared switch genes, revealing 40 switch genes present in both transcriptomes (**Table 1**). This list included important transcription factors, such as two NAC-domain proteins (NAC33 and NAC60)[37], three zinc fingers, one Myb protein, and the lateral organ boundaries domain proteins LOB15 and LOB18, which are required for organ shape determination. These genes are likely to represent the master regulators of the developmental phase transition between vegetative and mature metabolism.

Other common switches were related to secondary metabolism, including those representing the carotenoid and phenylpropanoid pathways (e.g., carotenoid cleavage dioxygenase), carbohydrate metabolism (e.g., alcohol dehydrogenase, stachyose synthase, and pyruvate dehydrogenase kinase), and stress responses (e.g., two osmotins and one dehydration-responsive protein). Interestingly, we found only one gene related to hormone responses, encoding the auxin-responsive SAUR29 (SMALL AUXIN UP RNA29) protein, which suggests a key role for this hormone in the grapevine developmental transition. We also found that 23 of the common switch genes were candidates for regulation by miRNA (**Table 1**).

Table 1: List of switch genes shared between the grapevine atlas and berry transcriptome.

Switch Gene ID	Annotation	Biological Process	Putative miRNA
VIT_07s0005g01680	Stachyose synthase	Carbohydrate Metabolic Process	Vv-miR390
VIT_09s0002g06420	Lactoylglutathione lyase	Carbohydrate Metabolic Process	
VIT_14s0060g00420	Pyruvate dehydrogenase kinase	Carbohydrate Metabolic Process	Vv-miR156b/Vv-miR156c/Vv-miR156d/ Vv-miR156e/Vv-miR156h/Vv-miR394a/ Vv-miR394c/Vv-miR845a/Vv-miR845b
VIT_14s0068g01760	Alcohol dehydrogenase	Carbohydrate Metabolic Process	
VIT_00s0323g00070	Pectin methylesterase inhibitor	Cell Wall Metabolism	
VIT_08s0040g01200	Short-chain type alcohol dehydrogenase	Cellular Process	
VIT_07s0005g02730	Myb Radialis	Developmental Process	Vv-miR2111-5p
VIT_14s0066g01710	Leaf senescence protein	Developmental Process	Vv-miR156f/Vv-miR156g/Vv-miR156h/ Vv-miR156i/Vv-miR398a/Vv-miR398c
VIT_01s0011g03670	Bifunctional nuclease	DNA/RNA Metabolic Process	Vv-miR3629a-3p/Vv-miR3629b/Vv-miR3629c
VIT_18s0072g01010	Peptide chain release factor eRF subunit 1	DNA/RNA Metabolic Process	
VIT_08s0058g00410	ferritin 1 (FER1)	Generation of Energy	Vv-miR3631b-3p
VIT_08s0058g00440	ferritin	Generation of Energy	Vv-miR3631b-3p
VIT_04s0044g01230	Unknown	No Hit	
VIT_14s0108g00450	No hit	No Hit	Vv-miR156b/Vv-miR156c/Vv-miR156d/ Vv-miR408
VIT_16s0098g01150	Auxin-responsive SAUR29	Response to Hormone Stimulus	Vv-miR390
VIT_16s0100g00570	Dehydration-responsive protein	Response to stress	Vv-miR3633a-5p/Vv-miR482
VIT_02s0025g04280	Osmotin	Response to Stress	Vv-miR408
VIT_02s0025g04340	Osmotin	Response to Stress	Vv-miR3626-5p/Vv-miR479
VIT_01s0127g00680	SRO2 (similar to rcd one 2)	Secondary Metabolic Process	Vv-miR3634-3p
VIT_02s0025g02920	caffeic acid 3-O-methyltransferase	Secondary Metabolic Process	Vv-miR172d
VIT_02s0087g00930	(9,10) (9`,10`) cleavage dioxygenase (CCD4) (VvCCD4b)	Secondary Metabolic Process	
VIT_16s0050g00390	4-coumarate-CoA ligase	Secondary Metabolic Process	Vv-miR159a/Vv-miR159b/ Vv-miR169r/Vv-miR169t/Vv-miR169u
VIT_14s0068g01360	GEM-like protein 5	Signal Transduction	
VIT_06s0004g07790	Lateral organ boundaries Domain 15	Transcription Factor Activity	Vv-miR172d
VIT_15s0048g00830	LOB domain-containing 18 (Asymmetric leaves 2-like protein 20)	Transcription Factor Activity	
VIT_19s0027g00230	NAC domain-containing protein (VvNAC33)	Transcription Factor Activity	Vv-miR164a/Vv-miR164b/Vv-miR164c/ Vv-miR164d
VIT_08s0007g07670	NAC domain-containing protein (VvNAC60)	Transcription Factor Activity	Vv-miR3626-5p
VIT_08s0040g01950	Zinc finger (C3HC4-type ring finger)	Transcription Factor Activity	Vv-miR3633b-5p/Vv-miR447b-3p
VIT_14s0219g00040	Zinc finger (C3HC4-type ring finger)	Transcription Factor Activity	
VIT_18s0001g01060	Zinc finger (C3HC4-type ring finger)	Transcription Factor Activity	
VIT_06s0009g01140	Amino acid permease	Transport	Vv-miR2111-5p
VIT_19s0014g04790	Organic cation/carnitine transporter4	Transport	Vv-miR172d/Vv-miR3629a-3p/Vv-miR3629b/ Vv-miR3629c
VIT_00s0214g00090	F-box protein PP2-B10 (Protein phloem protein 2-like B10)		Vv-miR477
VIT_08s0007g08840	Glycosyl transferase HGA1		
VIT_12s0057g00310	Gamete expressed1 (GEX1)		
VIT_06s0004g03910	Unknown protein		
VIT_11s0118g00580	Unknown		
VIT_14s0066g01190	Unknown protein		Vv-miR3630-3p
VIT_14s0219g00110	Unknown protein		
VIT_18s0001g10480	Unknown protein		

Heat cartography in tomato identifies genes that regulate fruit ripening

The availability of natural ripening-deficient mutants in tomato has facilitated the identification and characterization of master regulators of ripening, allowing us to test the efficiency of our integrated network strategy. We therefore applied the heat cartography approach to a tomato (*S. lycopersicum* cv Heinz) transcriptomic data set obtained by next-generation sequencing (Tomato Genome Consortium, 2012).

We matched the procedure used to construct the coexpression network for the grapevine by dividing the samples into two groups: vegetative/green organs (leaf, root, flower, bud, and immature berry at 1, 2, and 3 cm) and mature organs (mature green berry, breaker fruit, and fruit after 10 days of ripening). We identified 1,961 genes that were differentially expressed between vegetative/green and mature organs. As observed in the grapevine networks, we identified more down-regulated than up-regulated genes in mature tissue samples (1,736 versus 225), indicating once again that the transition to mature development is characterized more by the suppression of vegetative genes than the induction of genes specific for the mature phase. The resulting coexpression network therefore comprised 1,961 nodes and 350,958 edges. Analysing the distribution of APCCs, we recorded the peak at negative APCC values corresponding to the fight-club hubs. In this case, the peak corresponding to low APCCs (date hubs) was less evident because the number of date hubs was much lower than the number of party hubs.

By assigning a role to each node of the correlation network and a colour proportional to its APCC value, we built the heat cartography (**Fig. 6A**) and identified a group of 217 switch genes (corresponding to ~11% of all DEGs). These were all up-regulated during tomato fruit ripening and included many genes with a known direct role in fruit ripening and its regulation (**Fig. 6B**). In particular, we identified the MADS box master regulator *RIPENING INHIBITOR* (*RIN*), the *FRUITFULL1* (*FUL1*), the NAC factor *NON-RIPENING* (*NOR*), and the SQUAMOSA promoter binding protein *COLORLESS NON-RIPENING* (*CNR*) that, together with *FUL2* and the *AGAMOUS-LIKE TAGL1*, constitute the transcription factor network that directly regulate tomato fruit ripening[38]. Interestingly, the number of master regulators found among the switch genes was tested for statistical significance using a hypergeometric test. The resulting P value ($P < 0.001$) shows that the number of master regulator switches we identified

is more than would be obtained by chance, validating our network integration approach.

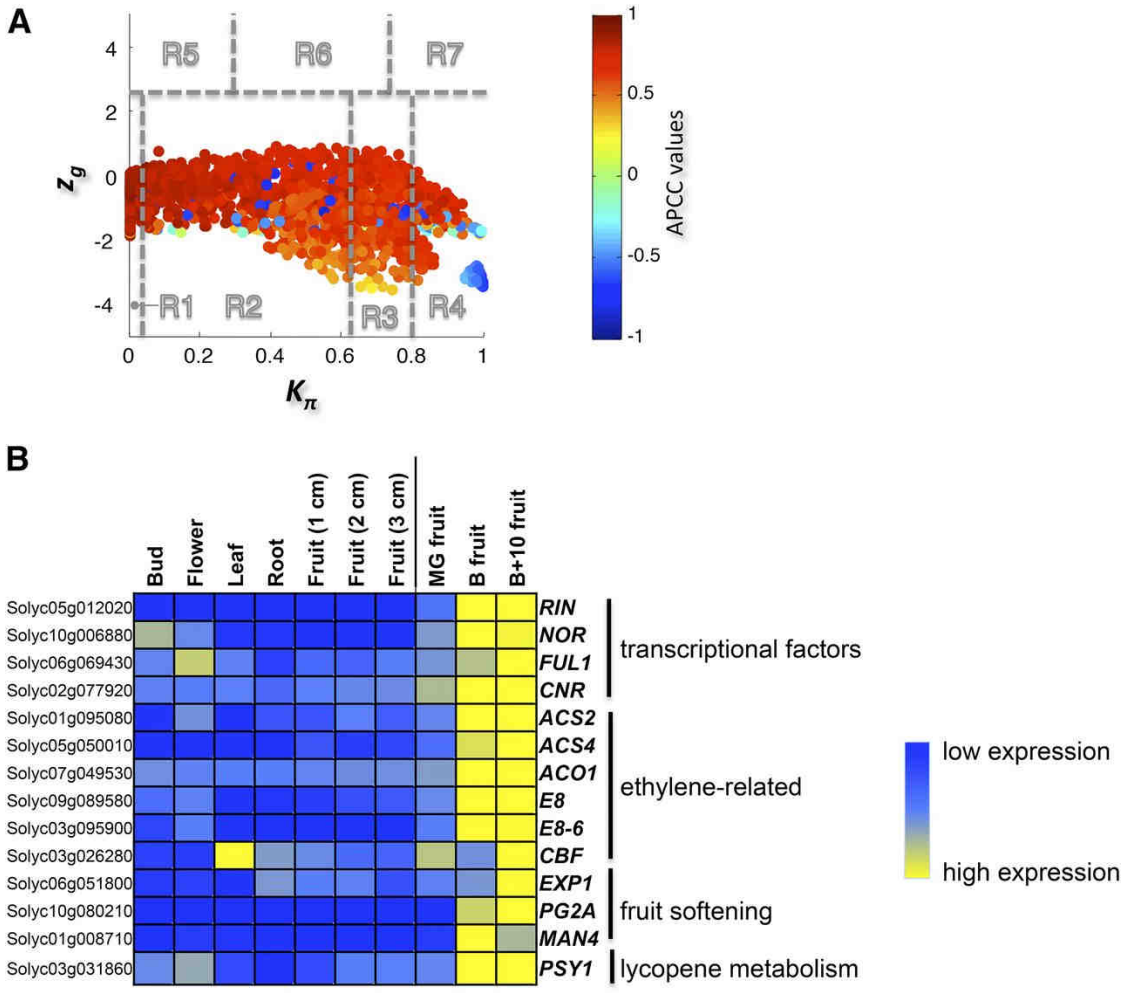


Figure 6: Heat cartography map for tomato.

(A) Heat cartography map for wild-type tomato fruit.

(B) Heat map for some switch genes that regulate fruit ripening in wild-type tomato plants. Blue shows genes expressed at low levels, and yellow shows genes expressed at high levels. Switch genes are all up-regulated during tomato fruit ripening. *RIN*, *RIPENING INHIBITOR*; *NOR*, *NON-RIPENING*; *FUL1*, *FRUITFULL1*; *CNR*, *COLORLESS NON-RIPENING*; *ACS2* and *ACS4*, genes for ACC synthase; *ACO1*, gene for ACC oxidase; *E8* and *E8-6*, genes for ACC oxidase homologs; *CBF*, *C-repeat binding factor*; *EXP1*, expansin; *PG2A*, polygalacturonase; *MAN4*, mannan endo-1,4-b-mannosidase; *PSY1*, phytoene synthase.

Tomato ripening mutants can be used as models to test the roles of switch genes identified by integrated network analysis and to characterize developmental programs that may still occur normally in the mutant. The *rin* mutant has a severe ripening defective phenotype characterized by the suppression of genes involved in

respiration and ethylene biosynthesis and the inability of fruits to soften and accumulate lycopene[7]. Recent microarray analysis has identified genes that are differentially expressed between wild-type and *rin* mutant fruit at three ripening stages (35, 40, and 45 days after planting), corresponding in the wild-type fruit to the immature green, breaker, and red-ripe fruit stages, respectively[9]. In this expression data set, the *rin* mutant is equivalent to a constitutive immature growth stage, whereas the wild-type fruits represent the switch to mature growth. We therefore tested our network integration strategy on this data set to confirm the heat cartography data described above and to determine whether *RIN* acts as a switch gene during ripening.

We built a coexpression network comprising 1,243 nodes and 440,734 edges. A role was assigned to each of these nodes to produce a heat cartography map (**Fig. 7A**). The distribution of APCCs was trimodal, with two peaks representing low and high positive APCC values (the date/party dichotomy) and a third peak representing negative APCC values corresponding to the fight-club hubs.

We found 354 switch genes corresponding to ~30% of the DEGs, almost all of which were up-regulated in wild-type fruits during ripening but down-regulated in the *rin* mutant. Interestingly, many of the switch genes (34 genes) we identified were the same as those found in the previous experiment. Among common switches we identified *RIN*, as expected, and genes encoding other ripening-related transcription factors (*NOR* and *FUL1*), together with genes involved in ethylene biosynthesis, fruit softening, and lycopene metabolism, again validating our network integration approach (**Fig. 7B**). However, among the switch genes not identified in the *RIN* data set were those encoding the transcription factor CNR and three proteins (E8, E8-6, and CBF) all involved in ethylene metabolism. This suggests that these genes also operate in the absence of a functional *RIN* allele.

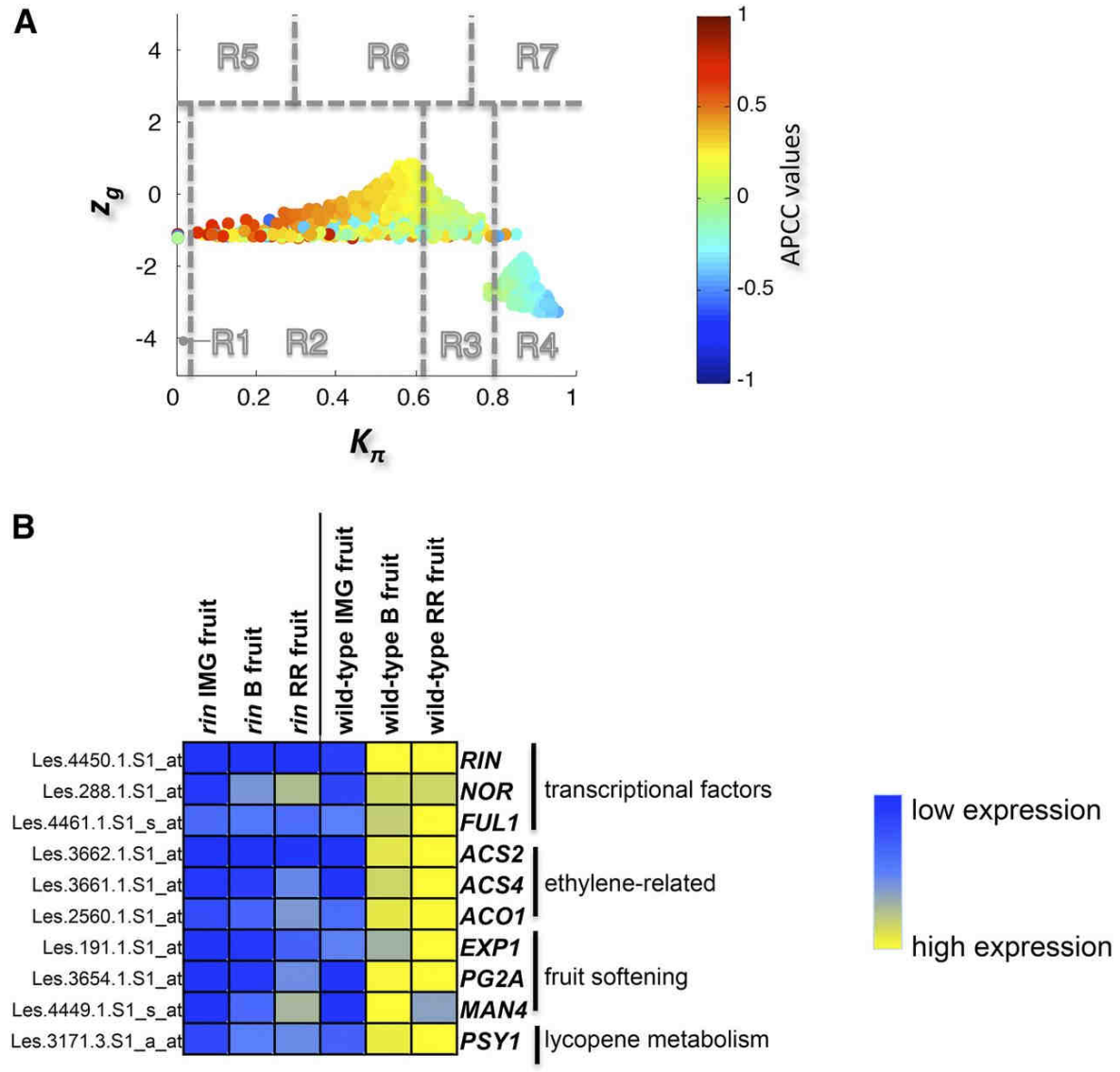


Figure 7: Heat cartography map for the tomato *rin* mutant.

(A) Heat cartography maps for rin mutant tomato fruit.

(B) Heat map for some switch genes that regulate fruit ripening in *rin* mutant tomato plants. Blue shows genes expressed at low levels, and yellow shows genes expressed at high levels. Switch genes were up-regulated in wild-type fruits during ripening but down-regulated in the *rin* mutant.

DISCUSSION

We developed an integrated approach based on the analysis of topological coexpression network properties to identify putative key regulators of organ phase transition during plant development. We tested our approach on the recently published grapevine global gene expression atlas, which follows the transcriptomic changes in 52 tissues/organs during development, including the transition from vegetative to mature growth[3]. This analysis identified a category of genes, named switch genes,

which are significantly up-regulated during the developmental transition and inversely correlated with a large number of genes that are suppressed during the mature growth phase. The same approach was used to analyse a new grapevine berry transcriptomic data set in order to identify switch genes representing the transition from the green phase to the maturation phase of berry development. Finally, our topological network analysis was applied to two tomato transcriptomic data sets, allowing the identification of genes that are already known as master regulators of fruit ripening, thus validating the basis of our methodology.

Switch genes in the vegetative-to-mature transition may represent the key mediators of transcriptome reprogramming during grapevine organ development

One of the most important insights gained from the analysis of our grapevine coexpression network was the identification of hubs, herein described as fight-club hubs, characterized by negative average Pearson correlation coefficients with their nearest neighbours. Unlike party hubs, it is therefore likely that fight-club hubs could play a negative regulatory role in grapevine development.

The fight-club hub genes included 113 so-called switch genes, which were predominantly connected outside their module. The analysis of network robustness revealed that switch genes are crucial nodes for network integrity and their removal causes a rapid decrease in the CPL. All of the switch genes were up-regulated during the transition to mature/woody development, and they showed strong inverse correlations with genes that are down-regulated during the same developmental transition. We found that the transition to the mature development phase was characterized by a larger number of down-regulated than up-regulated genes, suggesting that many processes are inhibited rather than activated to implement this developmental program.

The switch genes encoded a number of transcription factors, including zinc-fingers, ZFWD2 proteins, and the NAC (NAM, ATAF1/2, and CUP-SHAPED COTYLEDON2 [CUC2]) family proteins NAC33 and NAC60. ZFWD proteins contain a C3H-type zinc finger and seven WD40 repeats, and a role in development has been suggested based on their structural similarity to the *Arabidopsis thaliana* CONSTITUTIVE PHOTOMORPHOGENESIS1 and PLEIOTROPIC REGULATORY

LOCUS1 proteins, which regulate photomorphogenesis[39,40], whereas NAC transcription factors have been shown to regulate vegetative and reproductive development in Arabidopsis[41], tomato[42], papaya (*Carica papaya*)[43], and grapevine[27, 37], suggesting a broad role as master regulators of fruit ripening.

Interestingly, the linked genes that are inversely correlated with the transcription factors described above include several positive markers of vegetative/green tissues and several negative markers of mature/woody tissues[3]. Genes involved in photosynthesis are significantly overrepresented among these neighbouring genes, particularly ZFWD2, which is inversely correlated with 23 genes encoding photosystem I and II light-harvesting complex and reaction center components. This suggests that ZFWD2 represses photosynthesis-related genes during the transition to mature growth. This transition also involves the down-regulation cell cycle regulators, such as the NIMA-related serine/threonine kinases (NEKS). In particular, *NEK6* (whose expression is inversely correlated with *ZFWD2*) is involved in growth, development, and stress responses in Arabidopsis[45]. Other examples include cyclins, mitotic proteins, kinesin-related proteins, fibrins, tubulins, and Rho (ROP) family proteins, indicating that the network of regulators and effectors modulate key signalling pathways affecting cell proliferation, apoptosis and gene expression[45].

Many switch genes were inversely correlated with members of the transcription factor family SQUAMOSA PROMOTER BINDING PROTEIN (SBP)[46]. In particular, *SBP11* and the high-similarity pair *SBP2/SBP15* (nearest neighbours to *NAC33*) play diverse roles in vegetative development and early-stage berry development, but their expression declines during fruit maturation[46]. Furthermore, At-*SPL9* and At-*SPL15* are homologs of Vv-*SBP8*, and these are expressed in the vegetative shoot apex where they facilitate the juvenile-to-adult phase transition in Arabidopsis[47].

In addition to the transcription factors discussed above, many switch genes also represented metabolic functions (principally carbohydrate and secondary product metabolism) as well as stress responses during the major shift from vegetative to mature development. The presence of an alcohol dehydrogenase switch gene is one of the most interesting metabolic features of the network because the up-regulation of ADH genes is thought to represent the oxidative burst that occurs at the onset of berry ripening[48]. The secondary metabolic processes represented by switch genes include phenylpropanoid biosynthesis (two flavonol synthases) and modification (three caffeic acid *O*-3-methyltransferases), reflecting the developmental shift toward the production

of aromatic compounds. Another metabolic switch gene encoded (9,10) (9',10') carotenoid cleavage dioxygenase (*CCD4b*), whose expression increases dramatically during berry development, proportional to the loss of carotenoids[49]. *CCD4b* may protect plants against oxidative stress during fruit maturation by promoting the biosynthesis of abscisic acid[50]. Another switch gene encodes ALTERNATIVE OXIDASE 1A, which mediates abscisic acid signalling in response to oxidative stress. Additional switch genes involved in stress responses include those encoding thaumatins and osmotins, also known as grape ripening-induced proteins, which protect plants against changes in osmotic potential[51].

Taken together, these results suggest that the 113 switch genes represent key metabolic components of the major transcriptome reprogramming during the developmental transition to mature growth. The switch genes seem to act mainly as negative regulators of vegetative metabolic processes such as photosynthesis and cell proliferation.

miRNAs are potential regulators of grapevine switch genes: *NAC33:miR164* as a case study

MicroRNAs play important roles in plant development and responses to abiotic and biotic stress by pairing with specific mRNAs and causing their cleavage and degradation or blocking protein synthesis[52,53,54,13]. They are often expressed in a developmentally regulated or tissue-specific manner[37,53-65], and their target genes include those encoding other regulators such as transcription factors[66]. This suggests that plant miRNAs are master regulators of the regulatory hierarchy[54,56-58,62,65-74], so we hypothesized that specific miRNAs are likely to act as posttranscriptional regulators of the switch genes.

The 49 switch genes of the vegetative-to-mature transition that were predicted as miRNA targets included the NAC-domain gene *NAC33* as a putative target of miR164. Interactions involving miR164 and NAC-domain genes have been experimentally validated and appear to be conserved among many plants. For example, Arabidopsis miR164 targets five NAC-domain mRNAs, including *NAC1* (which is directly involved in lateral root emergence), *CUC1* and *CUC2* (functionally redundant genes that promote boundary formation and maintenance throughout vegetative and reproductive development), and *At5g07680* and *At5g61430* [65,75,76].

Furthermore, the tomato sly-*miR164* spatially and temporally regulates the expression of the *CUC2* homolog *GOBLET* [77,78] and uncharacterized *NAM* genes such as Sl-*NAM2*, which are involved in the establishment of floral boundaries[42]. Finally, maize (*Zea mays*) miR164 is one of the trans-acting factors that inhibit *NAC1*, which is involved in lateral root development.

The interaction between miR164 and NAC-domain mRNAs has been experimentally validated in grapevine[27], revealing that Vv-miR164 is expressed strongly in leaves (vegetative/green tissues) but only weakly in fruits (mature tissues). The NAC33 expression profile is generally inversely correlated with that of Vv-miR164, which would be expected for a miRNA target[27]. The recent transcriptome-wide analysis of grapevine miRNAs in six different organs during development revealed the spatiotemporal expression profile of the miRNAs involved in the predicted miRNA/switch gene pairs[79]. We were therefore able to confirm that 26 switch genes showed the opposite expression profiles to those of their putative miRNA regulators, e.g., Vv-miR166 and ZFWD2 (VIT_13s0047g01130), Vv-miR172d and caffeic acid 3-*O*-methyltransferase (VIT_02s0025g02920), and Vv-miR482 and dehydration-responsive protein (VIT_16s0100g00570). On the basis of these data, we propose a transcriptional regulatory network in which tissue-specific and stage-specific miRNAs regulate the expression of several switch genes identified by our integrated network analysis (**Fig. 8**). In particular, during the vegetative/green phase of organ development, switch genes are negatively regulated by miRNAs and vegetative genes are expressed. During the switch to the maturation phase, these miRNAs are deactivated, the switch genes are expressed and vegetative genes that are negatively correlated with the switches are suppressed (**Fig. 8**).This hypothesis provides an additional and more immediate level of regulation in the transition from vegetative to mature development.

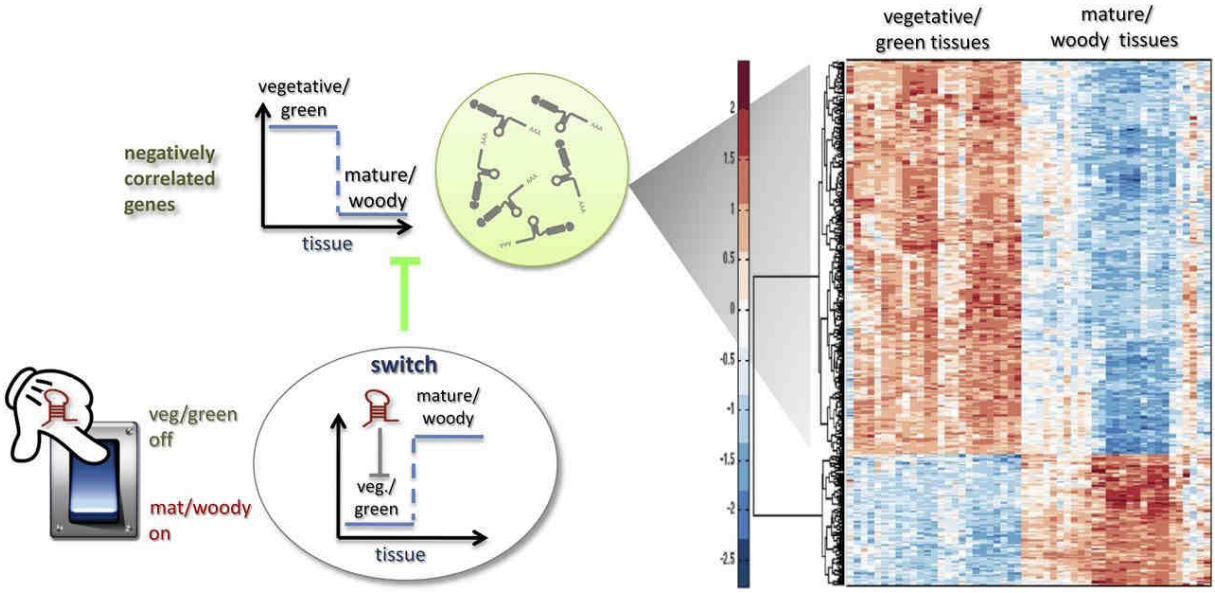


Figure 8: Proposed switch gene regulation by miRNAs.
The electric switch on the left side is a schematic representation of the regulation of the switch genes by miRNAs. The analysis of grapevine expression revealed the down-regulation of switch genes in vegetative/green tissues and their up-regulation in mature/woody tissues. Such genes therefore seem to act as switches in the green/mature transition by changing their state from off to on. Because target cleavage seems to be the dominant mode of action of miRNAs in plants, we hypothesize that the presence of specific miRNAs silences the switch mRNAs (negatively correlated mRNAs), but when the miRNA is down-regulated (or absent) the switch gene mRNA accumulates (mature/woody tissues) and as consequence its targets are down-regulated. The heat map shows the transcription level of positively and negatively correlated mRNAs with a typical switch. The number of inversely correlated genes far outnumbers the correlated ones, which is consistent with the strong negative value of the APCC, a unique feature of the switch genes.

The grapevine berry switch genes mainly encode transcription factors

We applied our method to a new berry transcriptomic data set following the expression profiles of five red-berry varieties during the transition from the green phase to the maturation phase. We identified 190 switch genes, all up-regulated during maturation and connected to 1,266 negatively correlated genes related to green/vegetative metabolism. Interestingly, transcription factors were significantly overrepresented as a functional category of the berry switch genes. The identification of *MYBA1*, *MYBA2*, and *MYBA3*, already well characterized for their direct and crucial role during the transition to berry ripening[29,31], strongly supports the robustness of our approach and allows us to propose other transcription factors as potential regulators of the metabolic

shift during berry development. Interestingly, we identified four NAC proteins (*NAC33* and *NAC60*, also identified as switch genes in the atlas transcriptome, and *NAC11* and *NAC13*) whose regulatory role in organ development has been proposed in many plant species[41-43], including grapevine[27,37]. We also identified two MYB proteins, eight zing-finger proteins (three shared with the atlas switch genes), two WRKY proteins, one bHLH protein, one Agamous-like MADS box protein, and three LATERAL ORGAN BOUNDARIES (LOB) proteins (two shared with the atlas switch genes). All these transcription factors are likely to represent master regulators of the phase transition during berry development.

Several switch genes were found to be known targets of MYB transcription factors, including *UFGT*, which is a master regulator of berry colour[28-31], and the anthocyanin transporter *GST4* [32,33]. These findings show that the switch genes include not only putative master regulators of the developmental phase transition but also the direct targets of these regulators. Another 12 switch genes encoded transporters, including the anthoMATE transporter *AM2*, which was recently shown to regulate the malvidin content of Malbec berries[80]. Further switch genes were found to encode enzymes catalyzing early steps in the phenylpropanoid pathway, e.g., caffeic acid 3-O-methyltransferase and 4-coumarate-CoA ligase, or components of terpenoid and carotenoid metabolism, e.g., *CCD4b* encoding a carotenoid cleavage dioxygenase that generates typical grape berry aromatic compounds[49]. All three enzymes were also represented by switch genes in the atlas data set.

The berry switch genes also include several encoding proteins involved in auxin and abscisic acid biosynthesis and signalling pathways, which induce berry ripening and colour development[81]. Three ethylene response factors were also identified, two of which have already been described[35]. Other switch genes encoded proteins involved in cell wall metabolism, which is consistent with the tissue softening and cell expansion that marks the onset of ripening[1]. These included genes encoding a cellulase, three xyloglucan endotransglucosylase/hydrolases, three korrigan-like endo-1,4-b-glucanases, and five pectin-modifying enzymes such as polygalacturonase, which is involved in fruit softening during apple and tomato maturation[82,83]. Another switch gene encoded expansin B4 (EXPB4), which is known to be up-regulated in berry flesh and skin during maturation[36]. Finally, we identified several genes involved in carbohydrate metabolism (glycolysis and sucrose biosynthesis), including an alcohol dehydrogenase, lactoylgluthatione lyase, pyruvate dehydrogenase kinase, and

stachiose synthase shared with the atlas switch genes, suggesting that sugar accumulation may promote the maturation phase of development.

Functional annotation of the 1,266 neighbours of switch genes showed that, as for the atlas, they include several positive markers of the immature berry, such as photosynthesis-related genes, genes involved in the cell cycle, the generation of energy, and the synthesis of secondary metabolites associated with green berries. Taken together, these data revealed that genes expressed during the immature green phase of development are inhibited after véraison, whereas genes related to sugar accumulation and secondary metabolism are induced. As in the atlas data set, the switch genes appear to regulate the immature-to-mature phase transition by the mass down-regulation of target genes rather than the activation of new pathways[29,84-87].

The common switch genes in the atlas and berry transcriptomes reveal a core set of regulators of the transition from vegetative to mature organ development. We found that more than half of these common switch genes are putative targets for miRNAs, thus providing a hypothetical higher level of regulation of the grapevine transcriptome during development and suggesting that the transcriptional reprogramming during development is fine-tuned by a relatively small number of miRNAs.

Analysis of topological coexpression networks in tomato reveals master regulators of fruit ripening and their hierarchical relationship

The efficiency of the integrated network analysis approach was difficult to judge in grapevine because little is known about the overall regulation of the developmental transition from vegetative to mature growth. This is not the case in tomato, where many of the key regulators of ripening have already been identified through the analysis of mutants and transgenic lines[7,10,88]. Therefore, the efficiency of our integrated network analysis approach in tomato can be validated externally against existing benchmarks.

The topological coexpression network derived from a tomato transcriptomic data set representing different vegetative organs and ripening fruit revealed the presence of 217 switch genes, including several with known roles in the regulation of fruit ripening[7,88,89]. In particular, we found the MADS box protein RIN, which is one of the earliest regulatory factors involved in ethylene-dependent and ethylene-independent fruit ripening[7], the NAC family protein NOR, the MADS box protein FUL1, and the

squamosal binding protein-like transcription factor CNR[88,89]. All these transcription factors are involved in a transcriptional network that promotes ripening by activating many ripening-related genes[10,90].

Typical tomato ripening-associated categories are represented by ethylene biosynthesis and perception, carotenoid accumulation, and cell wall metabolism. Tomato produces climacteric fruits because respiration and ethylene production are rapidly induced at the onset of ripening[90]. Accordingly, we identified genes directly involved in ethylene perception and metabolism, including two 1-aminocyclopropane-1-carboxylate synthases (ACS2 and ACS4), both directly activated by RIN[88], three 1-aminocyclopropane-1-carboxylate oxidases (ACO1, E8, and E8-6), and members of the ERF/AP2 transcription factor family such as LeCBF1[92]. Lycopene metabolism was represented by a switch gene encoding the enzyme phytoene synthase 1 (PSY1), which is the first committed step in carotenoid biosynthesis and represents a well-characterized tomato ripening-related gene regulated by *RIN* and FUL1[90]. Cell wall metabolism was also strongly represented because of its role in fruit softening, with switch genes encoding cell-wall-degrading proteins such as expansin A1 (EXP1), mannan-endo-1-4-β-mannosidase (MAN4), and polygalacturonase (PG2A).

The availably of well-characterized tomato ripening mutants provides an excellent opportunity for the analysis of fleshy fruit development[5]. In particular, the ripening deficient mutant *rin* provides insights into the transcriptional network that regulates ripening[7,9,88]. We applied our integrated network analysis method to the comparative transcriptome data set obtained from wild-type and *rin* mutant tomato fruits during ripening, generating a list of switch genes including most of the key regulators and of their direct targets as discussed above. The up-regulation of these switch genes during the ripening of wild-type fruit and their concurrent down-regulation in never-ripe *rin* mutant fruits demonstrated once again that our approach can identify putative switch genes that in some cases promote ripening directly but in most cases promote ripening indirectly by suppressing the vegetative development program. Interestingly, the *CNR* transcription factor gene and the *E8*, *E8-6*, and *CBF* genes involved in ethylene synthesis were not found among switch genes, suggesting they are not direct targets of *RIN*. Interestingly, although in previous studies *RIN* was proposed to interact with the *CNR* promoter[93], it was recently demonstrated that *RIN* did not directly bind the *CNR* regulative region[38], strongly supporting our findings.

Moreover, the inability of *RIN* to bind to its targets in the absence of *CNR* hints at a more complex relationship between these two ripening regulators.

The integration of different network methods allowed us to classify genes according to the extent of correlation with interaction partners in the same coexpression network, revealing hubs demonstrating predominantly negative correlations with neighbouring genes. Within this classification, we identified a subclass of so called switch genes that appear to encode key regulators of large-scale transcriptomic transitions, which may in turn be regulated by a small number of miRNAs. The integrated method was applied de novo in grapevine, providing insight into the overall regulation of the recently discovered major developmental transcriptome reprogramming, and specifically to the regulation of phase transition during berry ripening, but the basis of the method was also validated in tomato, resulting in the identification of known master regulators of fruit maturation. The integrated networking method therefore appears to be an ideal strategy to identify key regulators of organ development in fleshy fruit crops.

METHODS

Analyzing Gene Expression Data

Grapevine Atlas

The grapevine gene expression data set was obtained from the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) series entry GSE36128. It comprises 29,549 grapevine genes whose expression was quantified using microarrays from 54 samples taken from different tissues and stages. Two outlying samples (pollen and leaves undergoing senescence) were excluded, whereas the other 52 samples were divided into two groups: 25 vegetative/green tissues and 27 mature/woody tissues. We used a t-test and a $\log_2$-change filter to select DEGs between the vegetative/green group and the mature/woody group. The $\log_2$-change filter corresponds to the 95th percentile of the entire fold-change distribution. Finally, we selected 1,686 statistically significant DEGs after a t-test corrected for multiple testing using the Benjamini-Hochberg method with a false discovery rate threshold of 0.001[94].

All grapevine transcripts were annotated against the V1 version of the 12x draft annotation of the grapevine genome (http://genomes.cribi. unipd.it/DATA). Gene Ontology annotations were assigned to switch genes and their neighbouring genes

using the BiNGO v2.3 plug-in tool in Cytoscape (http://www.cytoscape.org/) v2.6 with PlantGOslim categories. Overrepresented PlantGOslim categories were identified using a hypergeometric test with a significance threshold of 0.05, after Benjamini and Hochberg correction with a false discovery rate of 0.001[94].

Berry Sampling

Grape berries were collected from five red-skin grapevine (*Vitis vinifera*) cultivars: Sangiovese, Barbera, Negro amaro, Refosco, and Primitivo, all cultivated during 2011 in the same experimental vineyard (Conegliano, Veneto region, Italy). Samples of berries were harvested at four phenological stages: pea-sized berries (Bbch 75) at ~20 days after flowering (Pea), berries beginning to touch (Bbch77) just prior to véraison (Touch), softening berries (Bbch 85) at the end of véraison (Soft), and berries ripe for harvest (Bbch 89) (Harv). From each berry sample, three biological replicates were created by pooling three samples of ~50 berries from different clusters on different plants located in different parts of the vineyard.

Total RNA was extracted from ~400mg of berry pericarp tissue (entire berries without seeds) ground in liquid nitrogen, using the Spectrum Plant Total RNA kit (Sigma-Aldrich) with some modifications[3]. Then, all 60 non-directional cDNA libraries were prepared from 2.5 µg of total RNA using the Illumina TruSeq RNA Sample preparation protocol (Illumina). Single-ended reads of 100 nucleotides were obtained using an Illumina HiSeq 1000 sequencer (average value per sample ±SD: 33,438,747 ± 5,610,087).

The reads were aligned onto the grapevine 12x reference genome PN40024[95] using TopHat v2.0.6 with default parameters[96]. An average of 85.55% of reads was mapped for each sample. Mapped reads were used to reconstruct the transcripts using Cufflinks v2.0.2[38] and the reference genome annotation V1 (http://genomes.cribi.unipd.it/DATA). All reconstructed transcripts were merged into a single non-redundant list of 29,287 transcripts, using Cuffmerge from the Cufflinks package, which merges transcripts whose reads overlap and share a similar exon structure (or splicing structure) and generates a longer chain of connected exons. Using this novel list of transcripts, the normalized mean expression of each transcript was calculated in fragments per kilobase of mapped reads for each triplicate using the geometric normalization method.

The samples were divided into two groups comprising immature berries (Pea and Touch) and mature berries (Soft and Harvest).The log_2-fold changes were computed as the difference between the averages of the log_2-scale expression data of the first group and the averages of the log_2-scale data of the second group. We used a log_2-change filter to select DEGs between the two groups. The log_2-change filter corresponds to the 95th percentile of the entire fold-change distribution. We identified 1,883 DEGs.

Tomato

The processed and normalized tomato RNA-Seq collection (Tomato Genome Consortium, 2012) was downloaded from the NexGenEx database. The NexGenEx database collection consisted of cleaned and processed reads (H. Bostan and M.L. Chiusano, unpublished data). The data set comprises the genome-wide expression analysis of 20 samples representing 10 stages in duplicate. The stages correspond to the vegetative and reproductive organs at different developmental stages. Each sample was analysed as two biological replicates, and expression profiles were obtained for 34,727 genes. The samples were divided into two groups comprising vegetative/green tissues (leaf, root, flower, bud, 1-cm fruit, 2-cm fruit, and 3-cm fruit) and three stages of ripening fruit (mature green, breaker fruit, and fruit at 10 days). The log_2-fold changes were computed as the difference between the averages of the log_2-scale expression data of the first group and the averages of the log_2-scale data of the second group. We used a log_2-change filter to select DEGs between the two groups. The log_2-change filter corresponds to the 95th percentile of the entire fold-change distribution. We identified 1,961 DEGs.

Tomato *rin* Mutant

The tomato *rin* mutant microarray expression data[9] were sourced from the NCBI Gene Expression Omnibus data repository (http:// www.ncbi.nlm.nih.gov/geo/) under series entry GSE20720. Data represent a fruit stage-specific comparison between different stages of wild-type and *rin* mutant fruits (18 conditions in total), including three stages of tomato fruit ripening: immature green, breaker, and red-ripe fruit and the chronologically equivalent stages of the non-ripened *rin* mutant fruit. Three biological replicates were available for each stage. The log_2-fold changes were computed as the difference between the averages of the log_2-scale expression data of the *rin* mutant

samples and the averages of the $\log_2$-scale data of the wild-type samples. We used a $\log_2$-change filter to select DEGs between the *rin* mutant and wild-type samples. The $\log_2$-change filter corresponds to the 95th percentile of the entire fold-change distribution. We identified 1,259 genes that were differentially expressed between the wild-type fruits and the *rin* mutant.

Building Gene Coexpression Networks

For each data set, we built a gene coexpression network of the DEGs using the Pearson correlation between gene expression profiles as the distance metric: two nodes/genes were connected if the absolute value of the Pearson correlation between their corresponding expression profiles was greater than a given threshold. The choice of the threshold was not trivial: the higher the threshold (i.e., the greater the correlation and the smaller the distance between gene expression profiles), the lower the number of links. Thus, increasing the threshold increases the number of connected components in the network. As a consequence, a reasonable threshold should reflect an appropriate balance between the number of connected components and the distances between gene expression profiles. Thus, we chose as a threshold the 85th percentile of the distribution of the Pearson correlation coefficients.

Finding the Community Structure of a Network

A traditional method for detecting community structure in networks is a k-means clustering algorithm[19-24], which partitions *n* observations/objects (corresponding, in our case, to the genes/nodes of the coexpression network) into *N* clusters, where *N* must be decided a priori. The goal of clustering is typically expressed by an objective function that depends on the proximities of the objects to each other or to the cluster centroids, e.g., by minimizing the squared distance of each object to its closest centroid. We used the following definition for our distance measurement:

$$dist(x, y) = 1 - \rho(x, y)$$

where $\rho(x,y)$ is the Pearson correlation between expression profiles of nodes *x* and *y*. Two nodes are close in the network ($dist(x,y) = 0$) if they are highly correlated ($\rho(x,y) = 1$), whereas two nodes are far apart in the network ($dist(x,y) = 2$) if they are highly

anti-correlated ($\rho(x,y)$ = -1). As the objective function, which measures the quality of clustering, we used the sum of the squared error (SSE). This is formally defined as follows:

$$SSE = \sum_{i=1}^{N} \sum_{x \in C_i} dist(c_i, x)^2$$

where N is the number of the clusters, C_i is the i^{th} cluster, x is an object in the i^{th} cluster, and c_i is the centroid of the i^{th} cluster given by

$$c_i = \frac{1}{m_i} \sum_{x \in C_i} x$$

where m_i is the number of objects in the i^{th} cluster. Given two different sets of clusters produced by two different runs of k-means, the one with the smallest SSE was preferred. Finally, the position of the elbow in the SSE plot as function of N suggests the correct number of clusters (i.e., the value of N).

Building the heat cartography map

Using a cartographic representation of a network[17], nodes were classified into a small number of system-independent universal roles on the basis of their intercluster and intracluster connectivity. Nodes with similar roles are expected to have similar relative within-module connectivities.

We used modified z and P parameters[17], renamed z_g and K_π, respectively, to allow the identification of genes characterizing the immature-to-mature transition. These genes appear to interact mainly outside their community; thus, they will populate the R4 region, which is characterized by nodes with fewer than 35% of their links within the module (P > 0.8). However, the original definition of P appears to approach to a maximum value when links are uniformly distributed among modules (i.e., $P_{max} = 1 - \frac{1}{N}$, where N is the number of modules). In all our data sets, the maximum number of clusters was 5; thus, P_{max} = 0.8, corresponding to the lower bound of the R4 region. As a consequence, if we used the original approach, we would have obtained an empty R4 region.

We defined the within-module degree z_g, which measures the connectivity of a node *i* to other nodes in the module, as follows:

$$z_g^i = \frac{k_i^{in} - \bar{k}_{\mathrm{C_i}}}{\sigma_{\mathrm{C_i}}}$$

where k_i^{in} is the number of links of node *i* to other nodes in its module $\mathrm{C_i}$, $\bar{k}_{\mathrm{C_i}}$ and $\sigma_{\mathrm{C_i}}C_i$, are the average and standard deviation of the degree distribution of all nodes in the module $\mathrm{C_i}$.

Different roles can also arise because of the connections of a node to modules other than its own: two nodes with the same z_g can play different roles if one of them is connected to several nodes in other modules when the other is not. We defined the clusterphobic coefficient K_π of node *i* as follows:

$$K_\pi^i = 1 - \left(\frac{k_i^{in}}{k_i}\right)^2$$

where k_i^{in} is the number of links of node *i* to nodes in module $\mathrm{C_i}$, and k_i is the total degree of node *i* in the network. The clusterphobic coefficient of a node is ~1 if its links are mostly outside its module and 0 if its links are only within its own module. The role of a node can be determined by its within-module degree and its clusterphobic coefficient, which define how the node is positioned in its own module and in the context of other modules. Nodes having a $z_g > 2.5$ are identified as “module hubs”, whereas nodes with $z_g < 2.5$ are considered as “module non-hubs”.

Taking into account the clusterphobic coefficient, the module non-hubs can be divided into four regions and thus into four roles: (1) ultra-peripheral nodes (region R1): if a node has all links within its module $(K_\pi = 0)$; (2) peripheral nodes (region R2): if a node has at least 60% of its links within the module $(K_\pi < 0.625)$; (3) non-hub connectors (region R3): if a node has 45–60% of its links within the module $(0.625 < K_\pi < 0.8)$; (4) non-hub kinless nodes (R4): if a node has fewer than 45% of its links within the module $(0.8 < K_\pi < 1)$. Non-hub kinless nodes cannot be clearly assigned to a single module hence their name.

Taking into account the clusterphobic coefficient, the module hub nodes can be divided into three regions and thus into three roles: (1) Provincial hubs (region R5): if a hub has at least 5/6 of its links within the module$(K_\pi < 0.3)$; (2) Connector hubs

(region R6): if a node with a large degree has at least half of its links within the module ($0.3 < K_\pi < 0.75$); (3) Kinless hubs (region R7): if a hub has fewer than half of its links within the module ($0.75 < K_\pi < 1$) and it cannot be clearly associated with a single module. To build the heat cartography map, we considered only those nodes with a local degree not equal to zero. This slightly reduces the number of nodes shown in the heat cartography map with respect to the number of nodes in the original coexpression network.

The definition of module hubs reflects the definition of the within module degree z_g, which is a local feature. In contrast, the common definition of hubs as nodes with an extremely high level of connectivity represents a topological property on a global scale. In order to distinguish hubs from module hubs, we defined them as “global hubs”.

To classify global hubs, we extended the original date/party dichotomy used for PPI networks[15] to the gene coexpression networks, and similarly we defined global hubs as nodes with five or more connections. For each global hub, we computed the APCCs between the expression profiles of the global hub and of its nearest neighbours. These APCCs followed a trimodal distribution, which allowed us to identify fight-club hubs in addition to the previously described date and party hubs (**Fig. 2**).

In order to build the cartographic heat map, we evaluated the APCC value for each node of the co-expression network and coloured each node in the plane identified by z_g and K_π according to the APCCs (**Fig. 3**). Fight-club hubs in the heat cartography map are nodes with negative APCC values but a degree of at least five.

SUPPLEMENTARY MATERIAL

Supplementary table 1: List of the 190 switch genes in the berry transcriptome. For each transcript, a mean FPKM value of the five variety samples was calculated for each growth stage. Abbreviations for the four developmental stages: P (Pea-sized berry), PV (Prevéraison), EV (End of véraison), H (Harvest). Switch genes identified as putative biomarker transcripts specific to maturation phase (Mat) and harvest stage (Harv) (**chapter 5**) and switch genes belonging to cluster 5 of the berry development transcriptomic route (**chapter 7**) are indicated by a star (*).

Switch Gene ID	Annotation	P	PV	EV	H	Mat	Harv	Cluster 5
VIT_01s0011g04370	Phosphatidylserine synthase 2	9.2	10.4	38.3	118.4			*
VIT_01s0137g00550	CYP71B34	0.4	0.7	5.1	17.4			*
VIT_05s0020g03870	MLK/Raf-related protein kinase 1	1.8	3.3	14.2	31.4		*	*
VIT_05s0020g05040	Proteinase inhibitor 1 PPI3B2	0.7	0.3	49.5	525.2			*
VIT_06s0004g04650	Metallothionein	980.0	859.1	4,019.6	14,926.0			*
VIT_06s0009g01140	Amino acid permease	0.1	0.2	10.2	24.3			*
VIT_06s0061g00760	Zinc finger (C2H2 type) family	1.2	0.9	12.3	41.9			*
VIT_09s0002g05250	Aminotransferase class I and II	35.3	38.3	108.6	275.7			*
VIT_11s0118g00560	Nodulin	6.0	7.8	21.5	58.3		*	*
VIT_12s0035g01050	Octicosapeptide/Phox/Bem1p (PB1) domain-containing protein	1.0	2.5	33.5	96.0			*
VIT_12s0057g00310	Gamete expressed1 (GEX1)	0.2	0.4	12.8	75.5		*	*
VIT_14s0066g01710	Leaf senescence protein	0.9	0.9	15.8	45.4			*
VIT_14s0083g00940	Auxin-independent growth promoter	5.4	1.2	28.2	70.3			*
VIT_14s0108g01070	NAC domain-containing protein (VvNAC11)	1.7	2.6	13.5	42.4			*
VIT_15s0021g02700	Expansin (VvEXPB4)	9.3	3.8	998.9	1,764.3			*
VIT_15s0048g00830	LOB domain-containing 18	0.3	0.0	16.6	69.3			*
VIT_18s0001g10480	Unknown protein	1.0	2.8	12.0	30.0		*	*
VIT_18s0072g01010	Peptide chain release factor eRF subunit 1	3.8	13.2	34.9	99.9			*
VIT_00s0125g00210	Serine/theronine protein kinase 2	3.0	2.2	14.4	22.1			
VIT_00s0125g00220	Serine/theronine protein kinase 2	2.4	2.0	12.7	15.5			
VIT_00s0214g00090	F-box protein PP2-B10 (Protein phloem protein 2-like B10)	2.6	4.4	18.3	21.7			
VIT_00s0271g00110	flavodoxin-like quinone reductase 1	2.5	1.8	54.7	46.9			
VIT_00s0274g00080, VIT_00s0274g00090	Benzoquinone reductase	10.8	6.5	93.6	151.6			
VIT_00s0322g00030	Alpha-6-galactosyltransferase	38.6	41.0	256.0	142.5			
VIT_00s0323g00070	Pectin methylesterase inhibitor	0.8	1.0	235.0	245.0			
VIT_00s0340g00050	Endo-1,4-beta-glucanase korrigan (KOR)	1.6	1.8	18.6	18.8			
VIT_00s0340g00060	Endo-1,4-beta-glucanase korrigan (KOR)	1.3	1.3	15.1	13.8			
VIT_00s2526g00010	Endo-1,4-beta-glucanase korrigan (KOR)	1.0	1.0	12.2	11.1			
VIT_01s0010g02030	Gamma-thionin precursor	800.5	811.6	21,737.0	14,610.7			
VIT_01s0010g02460	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic 3	64.9	60.2	906.6	841.8			
VIT_01s0011g03670	Bifunctional nuclease	2.6	2.7	511.6	707.3			
VIT_01s0026g01020	Binding	3.6	5.5	35.5	29.7			
VIT_01s0026g02100	Unknown protein	16.3	20.3	84.0	87.5			
VIT_01s0127g00680	SRO2 (similar to rcd one 2)	3.4	7.1	143.2	235.8			
VIT_01s0137g00430	Cellulase	6.4	10.1	89.4	167.3			
VIT_01s0137g00520	CYP71B35	0.5	0.7	8.6	25.1			
VIT_01s0137g00540	CYP71E1	0.6	0.8	5.4	9.4			
VIT_01s0137g00560	CYP71B34	0.6	1.1	7.1	12.8			
VIT_01s0146g00260	Nodulin MtN3	16.1	10.1	1,426.5	1,985.2			
VIT_01s0150g00380	Unknown protein	2.5	2.5	22.9	42.7			
VIT_02s0012g00500	Invertase/pectin methylesterase inhibitor	6.2	4.6	501.4	583.0			
VIT_02s0012g01040	NAC domain-containing protein (VvNAC13)	1.7	6.8	72.9	41.6			
VIT_02s0012g02220	Xyloglucan endotransglucosylase/hydrolase 30	2.2	7.5	112.1	207.0			
VIT_02s0025g01450	Unknown protein	20.8	34.4	294.6	207.1			
VIT_02s0025g02920	caffeic acid 3-O-methyltransferase	1.3	2.7	21.9	65.6			
VIT_02s0025g03600	Phospholipid hydroperoxide glutathione peroxidase	117.2	172.0	887.5	1,472.9			
VIT_02s0025g03730	Translation initiation factor eIF-2B alpha subunit	1.3	0.8	23.7	20.2			
VIT_02s0025g04280	Osmotin	5.4	4.6	407.1	1,179.2			
VIT_02s0025g04330	Thaumatin VVTL1 [Vitis vinifera]	12.4	7.8	6,318.9	2,715.7			
VIT_02s0025g04340	Osmotin	7.1	4.3	186.3	386.4			

(Table continues on following page)

Supplementary table 1 (Continued from previous page)

Switch Gene ID	Annotation	P	PV	EV	H	Mat	Harv	Cluster 5
VIT_02s0025g04870	No hit	0.3	0.7	4.8	9.6			
VIT_02s0033g00380	Myb VvMYBA1	0.3	0.1	262.4	107.7			
VIT_02s0033g00390	VvMybA2	0.1	0.2	158.1	62.8			
VIT_02s0033g00410	VvMybA1	0.0	0.2	71.2	18.7			
VIT_02s0033g00450	VvMybA3	0.3	0.4	132.8	36.1			
VIT_02s0033g01330	Acyl-CoA binding protein	164.3	125.3	2,849.4	2,865.1			
VIT_02s0087g00930	(9,10) (9`,10`) cleavage dioxygenase (VvCCD4b)	0.2	1.1	43.0	40.6			
VIT_02s0234g00110	Oleosin OLE-4	0.4	0.6	36.8	23.6			
VIT_03s0038g01380	Calcium-binding EF hand	32.1	29.3	152.5	152.1			
VIT_03s0038g03570	Monocopper oxidase SKS5 (SKU5 Similar 5)	14.0	31.0	397.4	276.4			
VIT_03s0038g04390	Dehydrin 1	0.0	0.2	4.4	9.1			
VIT_03s0091g00260	Zinc finger protein 4	2.3	5.4	35.6	19.1			
VIT_03s0091g00670	Lateral organ boundaries protein 38	1.5	2.1	14.0	25.1			
VIT_04s0008g04990	Potassium channel AKT1	0.7	2.2	7.4	16.2			
VIT_04s0008g05770	CBL-interacting protein kinase 25 (CIPK25)	11.8	13.5	116.4	363.1		*	
VIT_04s0008g06000	ERF/AP2 Gene Family (VvERF045)	6.6	30.4	280.2	172.3	*		
VIT_04s0023g00130	Unknown protein	0.8	2.1	19.8	19.2			
VIT_04s0023g01800	Maf, septum formation protein	1.0	1.4	15.3	11.8			
VIT_04s0044g01230	Unknown	0.4	2.6	153.5	130.0			
VIT_04s0079g00690	Glutathione S-transferase (VvGST4)	0.6	0.7	1,428.5	621.7			
VIT_05s0020g04730	Zinc finger (C3HC4-type ring finger)	1.0	1.1	7.6	12.0	*		
VIT_05s0049g00510	Ethylene response factor ERF1	2.9	1.5	1,795.4	1,856.6			
VIT_05s0049g00770	No hit	0.9	1.7	1,817.7	900.3			
VIT_05s0049g01780	Caleosin	0.2	0.6	7.9	9.8			
VIT_05s0049g02240	AWPM-19	0.8	0.8	90.8	81.6			
VIT_05s0062g00610	Xyloglucan endotransglucosylase/hydrolase 23	2.8	5.5	56.9	57.9			
VIT_05s0062g01190	Pirin	13.1	12.0	99.1	133.5			
VIT_05s0077g01980	Unknown protein	1.7	1.5	41.2	44.7			
VIT_06s0004g03910	Unknown protein	4.2	8.3	114.1	80.8	*		
VIT_06s0004g07550	Wound-induced protein WI12	0.6	0.1	24.9	65.9			
VIT_06s0004g07790	Lateral organ boundaries Domain 15	0.2	0.5	29.4	57.4			
VIT_06s0009g03120	Cytochrome P450, family 79, subfamily A, polypeptide 2	0.0	0.0	9.3	16.8			
VIT_06s0061g00550	Xyloglucan endotransglucosylase/hydrolase 32	3.7	10.5	2,980.6	1,507.0			
VIT_07s0005g00140	Peripheral-type benzodiazepine receptor	0.1	0.2	4.9	5.7			
VIT_07s0005g00660	Late embryogenis abundant protein 5	471.5	399.8	3,406.5	4,132.6			
VIT_07s0005g00750	Sucrose synthase	123.1	175.1	769.0	782.3	*		
VIT_07s0005g01680	Stachyose synthase	0.6	2.0	31.7	33.0			
VIT_07s0005g01710	WRKY DNA-binding protein 23	9.2	19.1	123.1	78.1	*		
VIT_07s0005g02710	Unknown protein	6.9	22.5	156.9	116.1			
VIT_07s0005g02730	Myb Radialis	17.8	10.2	47.4	91.5			
VIT_07s0005g03310	Cofilin	154.5	380.9	1,268.8	1,452.3	*		
VIT_07s0005g03810	Unknown protein	2.5	2.8	12.5	29.8			
VIT_07s0005g04890	Glutathione S-transferase 25 GSTU7	1.9	4.8	29.9	61.5			
VIT_07s0031g01260	No hit	0.2	0.1	11.0	10.0			
VIT_07s0031g01930	myb TKI1 (TSL-KINASE INTERACTING PROTEIN 1)	1.8	1.2	11.8	24.6			
VIT_08s0007g00180	OBP3 (OBF-binding protein 3)	4.0	8.7	38.0	28.0			
VIT_08s0007g06040	Beta-1,3-glucanase	51.8	14.3	923.1	1,391.1			
VIT_08s0007g07670	NAC domain-containing protein (VvNAC60)	1.0	5.3	30.9	30.8			
VIT_08s0007g08310	Alkaline alpha galactosidase	105.6	102.3	780.8	864.6			
VIT_08s0007g08330	Polygalacturonase PG1	0.1	0.4	224.2	107.8			
VIT_08s0007g08840	Glycosyl transferaseHGA1	8.7	27.7	81.2	111.5			
VIT_08s0040g01200	Short-chain type alcohol dehydrogenase	0.8	0.6	392.2	292.9			
VIT_08s0040g01950	Zinc finger (C3HC4-type ring finger)	6.0	24.8	280.2	243.9	*		
VIT_08s0056g01240	No hit	2.5	4.2	568.1	538.6	*		
VIT_08s0056g01600, VIT_08s0056g01610	No hit	1,705.8	2,261.9	37,448.0	36,274.8			
VIT_08s0058g00400	No hit	0.4	2.8	30.0	39.0			
VIT_08s0058g00410	ferritin 1 (FER1)	0.7	7.9	93.4	96.4			
VIT_08s0058g00430	ferritin	0.3	3.4	45.1	55.3			
VIT_08s0058g00440	ferritin	0.4	2.8	96.1	110.5			
VIT_09s0002g06420	Lactoylglutathione lyase	2.4	12.6	217.7	214.8			
VIT_09s0002g07110	KEG (keep on going)	13.2	19.9	83.8	108.2	*		
VIT_09s0018g01370	STE20/SPS1 proline-alanine-rich protein kinase	1.3	2.6	12.6	19.2			
VIT_09s0018g01390	STE20/SPS1 proline-alanine-rich protein kinase	1.2	2.6	13.2	18.2			
VIT_10s0003g01410	CBL-interacting protein kinase 20 (CIPK20)	0.1	0.3	28.8	13.6			
VIT_10s0071g00320	GRAM domain-containing protein / ABA-responsive	4.0	4.3	22.2	21.9			
VIT_10s0116g00370	Unknown protein	2.9	2.6	29.3	25.4	*		
VIT_10s0116g00550	Oligosaccharide transporter OST3/OST6	66.3	66.7	294.3	350.1			

(Table continues on following page)

Supplementary table 1 (Continued from previous page)

Switch Gene ID	Annotation	P	PV	EV	H	Mat	Harv	Cluster 5
VIT_10s0116g00570	Nuclear transport factor 2B	40.5	89.3	2,355.9	2,738.4			
VIT_11s0016g03090	No hit	1.0	0.0	9.3	16.2			
VIT_11s0016g04920	Early nodulin 93	9.9	3.8	5,650.4	6,263.7			
VIT_11s0016g04990	Unknown protein	4.3	6.7	29.3	40.3			
VIT_11s0016g05530	Plastocyanin domain-containing protein	72.2	43.9	284.9	520.1			
VIT_11s0052g00630	Metallothionein	7.7	13.9	128.0	71.3			
VIT_11s0118g00200	Sucrose-phosphate synthase	40.7	65.8	306.4	172.0			
VIT_11s0118g00580	Unknown	0.8	2.0	13.8	12.6			
VIT_12s0028g03860	Zinc finger (C3HC4-type ring finger) protein (RMA1)	10.3	8.4	43.0	116.3			
VIT_12s0034g01870	Cupin	3.2	10.0	49.4	146.0			
VIT_12s0034g01910	Cupin family protein	0.2	0.1	6.5	19.6			
VIT_12s0034g01920, VIT_12s0034g01930	Globulin-like protein	0.2	0.2	3.8	9.9			
VIT_12s0034g01950, VIT_12s0034g01960	Legumin	0.9	0.5	34.4	117.3			
VIT_12s0034g01970	Cupin	3.0	1.5	332.4	1,011.4			
VIT_12s0059g02510	Zinc finger (B-box type)	4.0	13.2	68.7	75.9			
VIT_13s0064g01210	Zf A20 and AN1 domain-containing stress-associated protein 2	177.1	492.0	1,203.6	1,416.5			
VIT_13s0067g02300	Hypoxia-responsive	1.1	0.5	208.1	341.1			
VIT_13s0074g00570	Gamma-aminobutyric acid transporter	1.1	0.4	17.2	41.9			
VIT_13s0158g00100	MADS-box agamous-like 15	5.1	13.4	29.4	56.1			
VIT_14s0060g00420	Pyruvate dehydrogenase kinase	3.1	3.8	17.2	28.1			
VIT_14s0060g01490	DnaJ homolog, subfamily A, member 5	12.6	30.3	321.6	235.1			
VIT_14s0060g01980	Unknown protein	2.2	4.8	70.4	50.3			
VIT_14s0060g02300	Hypoxia-responsive	16.9	7.4	131.9	160.7			
VIT_14s0066g00700	Oleosin H-isoform	1.4	1.5	126.1	80.0			
VIT_14s0066g01140	Unknown protein	0.1	0.0	11.1	25.0			
VIT_14s0066g01190	Unknown protein	2.8	4.0	30.2	71.5			
VIT_14s0066g01600	NHL repeat-containing protein	4.6	7.7	137.1	77.4			
VIT_14s0068g01360	GEM-like protein 5	1.5	1.4	20.3	51.5			
VIT_14s0068g01400	UPF0497 family	1.0	0.4	44.0	116.7			
VIT_14s0068g01760	Alcohol dehydrogenase	10.4	11.6	132.0	174.7			
VIT_14s0081g00030	Pathogenesis-related protein-4 (Chitinase)	9.8	15.4	4,309.7	2,373.5			
VIT_14s0083g01140	B12D	21.1	58.0	280.9	230.3			
VIT_14s0108g00450	No hit	1.0	2.9	157.8	118.5			
VIT_14s0128g00570	Germin	0.3	0.1	49.8	140.8			
VIT_14s0171g00400	Lysine histidine transporter 1	12.0	10.4	46.4	138.8			
VIT_14s0219g00040	Zinc finger (C3HC4-type ring finger)	3.1	6.0	25.7	26.5			
VIT_14s0219g00110	Unknown protein	19.4	21.9	135.3	181.1			
VIT_15s0046g00150	DOF affecting germination 1	2.5	10.5	34.2	41.1			
VIT_15s0046g01500	Translation initiation factor eIF-1A	92.9	140.2	390.2	673.0			
VIT_15s0048g00740	No hit	1.2	0.9	20.7	13.1			
VIT_16s0022g00960	Invertase/pectin methylesterase inhibitor	187.5	116.9	11,757.5	8,967.5	*		
VIT_16s0022g01770	Phosphopyruvate hydratase.	169.8	149.9	2,161.4	2,083.2			
VIT_16s0039g02230	UDP glucose:flavonoid 3-O-glucosyltransferase (VvUFGT)	0.0	0.1	123.0	35.9			
VIT_16s0050g00390	4-coumarate-CoA ligase	11.7	30.6	90.1	83.0			
VIT_16s0050g00570	Pectinacetylesterase	1.6	2.1	13.9	29.5			
VIT_16s0050g00910	MATE efflux family protein	5.5	3.8	251.0	115.2			
VIT_16s0098g00800	Copper-binding family protein	1.6	3.0	47.0	52.6			
VIT_16s0098g01150	Auxin-responsive SAUR29	3.5	3.6	115.7	121.7			
VIT_16s0100g00400	ERF/AP2 Gene Family (VvERF019)	3.5	8.0	38.1	29.4			
VIT_16s0100g00570	Dehydration-responsive protein	0.2	3.1	26.1	25.3	*		
VIT_17s0000g00410	No hit	0.2	0.5	8.2	5.0			
VIT_17s0000g00430	basic helix-loop-helix (bHLH) family	11.5	86.7	996.7	421.0			
VIT_17s0000g01280	WRKY DNA-binding protein 75	11.6	22.9	73.9	81.9			
VIT_17s0000g03200	Unknown protein	4.0	6.0	47.2	35.9			
VIT_17s0000g03630	No hit	0.1	0.1	2.4	5.8			
VIT_17s0000g05110	CYP78A4	33.7	64.6	762.5	1,901.5			
VIT_18s0001g01060	Zinc finger (C3HC4-type ring finger)	4.7	7.0	55.1	76.5			
VIT_18s0001g05990	UDP-glycosyltransferase 85A1	0.3	1.6	45.6	32.1			
VIT_18s0001g06060	UDP-glycosyltransferase 85A1	0.3	7.8	253.0	191.8			

(Table continues on following page)

Supplementary table 1 (Continued from previous page)

Switch Gene ID	Annotation	P	PV	EV	H	Mat	Harv	Cluster 5
VIT_18s0001g07840	Unknown protein	9.7	21.4	138.9	106.0			
VIT_18s0001g08200	MATE efflux family protein ZF14	0.2	0.5	13.0	6.0			
VIT_18s0001g11930	Thaumatin SCUTL2	1.6	0.7	216.3	183.1			
VIT_18s0001g12710	No hit	0.3	0.1	7.4	7.0			
VIT_18s0001g15460	Stearyl acyl carrier protein desaturase	4.7	4.5	22.9	48.4			
VIT_18s0075g00340, VIT_18s0075g00350	Sucrose-phosphate synthase isoform C	0.8	5.7	23.3	24.1			
VIT_18s0089g00160	1,4-beta-mannan endohydrolase	0.3	0.0	10.1	9.3			
VIT_19s0014g04790	Organic cation/carnitine transporter4	0.7	1.3	21.0	79.8			
VIT_19s0015g01300	Amino acid permease 7	12.2	9.5	152.0	107.5	*		
VIT_19s0015g01310	Amino acid permease 7	2.4	2.5	26.0	14.3			
VIT_19s0027g00230	NAC domain-containing protein (VvNAC33)	0.8	2.9	21.9	26.5			
VIT_19s0090g00960	Unknown	22.7	27.7	152.7	133.6			
VIT_19s0090g01340	No hit	390.6	65.0	8,617.6	9,578.0			
VIT_19s0090g01370	No hit	50.4	75.6	10,800.8	19,597.0			

REFERENCES

1. Coombe B.G. & Hale C.R. The hormone content of ripening grape berries and the effects of growth substance treatments. Plant Physiol. 51:629-634 (1973)
2. Tornielli G.B., Zamboni A., Zenoni S., Delledonne M., Pezzotti M. Transcriptomics and metabolomics for the analysis of grape berry development. In The Biochemistry of the Grape Berry, H. Geros, M.M. Chavez, and S. Delrot, eds (Sharjan, United Arab Emirates: Bentham Science Publishers), pp. 218-240 (2011)
3. Fasoli M., et al. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. Plant Cell 24:3489-3505 (2012)
4. Weier T.E., Stocking C.R., Barbour M.G., Rost T.L. Botany: An Introduction to Plant Biology. (New York: Wiley) (1982)
5. Gupta V., Mathur S., Solanke A.U., Sharma M.K., Kumar R., Vyas S., Khurana P., Khurana J.P., Tyagi A.K., Sharma A.K. Genome analysis and genetic enhancement of tomato. Crit. Rev. Biotechnol. 29:152-181 (2009)
6. Moore S., Vrebalov J., Payton P., Giovannoni J. Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. J. Exp. Bot. 53:2023-2030 (2002)
7. Vrebalov J., Ruezinsky D., Padmanabhan V., White R., Medrano D., Drake R., Schuch W., Giovannoni J. A. MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. Science 296:343-346 (2002)
8. Manning K., Tör M., Poole M., Hong Y., Thompson A.J., King G.J., Giovannoni J.J., Seymour G.B. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat. Genet. 38:948-952 (2006)
9. Kumar R., Sharma M.K., Kapoor S., Tyagi A.K., Sharma A.K. Transcriptome analysis of *rin* mutant fruit and in silico analysis of promoters of differentially regulated genes provides insight into LeMADS-RIN-regulated ethylene-dependent as well as ethylene-independent aspects of ripening in tomato. Mol. Genet. Genomics 287:189-203 (2012)
10. Karlova R., Chapman N., David K., Angenent G.C., Seymour G.B., de Maagd R.A. Transcriptional control of fleshy fruit development and ripening. J. Exp. Bot. 65:4527-4541 (2014)
11. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.(San Francisco, CA: Morgan Kaufmann Publishers) (1988)
12. Pearl J. Causality: Models, Reasoning, and Inference. (Cambridge, UK: Cambridge University Press) (2000)
13. Zhang B., Horvath S. A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4:1 (2005)
14. Albert R., Jeong H., Barabasi A.-L. Error and attack tolerance of complex networks. Nature 406:378-382 (2000)
15. Han J.D.J., Bertin N., Hao T., Goldberg D.S., Berriz G.F., Zhang L.V., Dupuy D., Walhout A.J.M., Cusick M.E., Roth F.P., Vidal M. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430:88-93 (2004)

16. Stuart J.M., Segal E., Koller D., Kim S.K. A gene coexpression network for global discovery of conserved genetic modules. Science 302:249-255 (2003)
17. Guimerà R. & Amaral L.A.N. Cartography of complex networks: modules and universal roles. J. Stat. Mech. 2005:a35573 (2005)
18. Grimplet J., Van Hemert J., Carbonell-Bejerano P., Díaz-Riquelme J.,Dickerson J., Fennell A., Pezzotti M., Martínez-Zapater J.M. Comparative analysis of grapevine whole-genome gene predictions, functional annotation, categorization and integration of the predicted gene sequences. BMC Res. Notes 5:213 (2012)
19. Steinhaus H. (1957). Sur la division des corps matériels en parties. Bulletin de l'Académie Polonaise des Sciences 4:801-804.
20. Forgy E.W. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. Biometrics 2:768-769 (1965)
21. MacQueen J.B. Some methods for classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, L.M. Le Cam and J. Neyman, eds (Englewood Cliffs, NJ: University of California Press), pp. 281-297 (1967)
22. Hartigan J.A. Clustering Algorithms. (NewYork: JohnWiley&Sons) (1975)
23. Hartigan J.A. & Wong M.A. Algorithm AS 136: A K-Means Clustering Algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28:100-108 (1979)
24. Lloyd S.P. Least square quantization in PCM. IEEE Trans. Inf. Theory 28:129-137 (1982)
25. Agarwal S., Deane C.M., Porter M.A., Jones N.S. Revisiting date and party hubs: novel approaches to role assignment in protein interaction networks. PLOS Comput. Biol. 6:e1000817 (2010)
26. Barabasi A.-L. & Albert R. Emergence of scaling in random networks. Science 286:509-512 (1999)
27. Sun X., Korir N.K., Han J., Shangguan L.-F., Kayesh E., Leng X.P., Fang J.G. Characterization of grapevine microR164 and its target genes. Mol. Biol. Rep. 39:9463-9472 (2012)
28. Kobayashi S., Goto-Yamamoto N., Hirochika H. Retrotransposon-induced mutations in grape skin color. Science 304:982 (2004)
29. Lijavetzky D., Ruiz-García L., Cabezas J.A., De Andrés M.T., Bravo G., Ibáñez A., Carreño J., Cabello F., Ibáñez J., Martínez-Zapater, J.M.. Molecular genetics of berry colour variation in table grape. Mol. Genet. Genomics 276:427-435 (2006)
30. This P., Lacombe T., Cadle-Davidson M., Owens C.L. Wine grape (*Vitis vinifera* L.) color associates with allelic variation in the domestication gene VvmybA1. Theor. Appl. Genet. 114:723-730 (2007)
31. Walker A.R., Lee E., Bogs J., McDavid D.A., Thomas M.R., Robinson S.P. White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant J. 49:772-785 (2007)
32. Cutanda-Perez M.C., Ageorges A., Gomez C., Vialet S., Terrier N., Romieu C., Torregrosa L. Ectopic expression of VlmybA1 in grapevine activates a narrow set of genes involved in anthocyanin synthesis and transport. Plant Mol. Biol. 69:633-648 (2009)
33. Gomez C., Conejero G., Torregrosa L., Cheynier V., Terrier N., Ageorges A. In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST. Plant J. 67:960-970 (2011)

34. Robinson S.P. & Davies C. Molecular Biology of grape berry ripening. Aust. J. Grape Wine Res. 6:175-188 (2000)
35. Licausi F., Giorgi F.M., Zenoni S., Osti F., Pezzotti M., Perata P. Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*. BMC Genomics 11:719 (2010)
36. Dal Santo S., Tornielli G.B., Zenoni S., Fasoli M., Farina L., Anesi A., Guzzo F., Delledonne M., Pezzotti M. The plasticity of the grapevine berry transcriptome. Genome Biol. 14:r54 (2013)
37. Wang N., Zheng Y., Xin H., Fang L., Li S. Comprehensive analysis of NAC domain transcription factor gene family in *Vitis vinifera*. Plant Cell Rep. 32:61-75 (2013)
38. Roberts A., Pimentel H., Trapnell C., Pachter L. Identification of novel transcripts in annotated genomes using RNASeq. Bioinformatics 27:2325-2329 (2011)
39. Németh K., et al. Pleiotropic control of glucose and hormone responses by PRL1, a nuclear WD protein, in Arabidopsis. Genes Dev. 12:3059-3073 (1998)
40. Torii K.U., McNellis T.W., Deng X.W. Functional dissection of Arabidopsis COP1 reveals specific roles of its three structural modules in light control of seedling development. EMBO J. 17:5577-5587 (1998)
41. Raman S., Greb T., Peaucelle A., Blein T., Laufs P., Theres K. Interplay of miR164, CUP-SHAPED COTYLEDON genes and LATERAL SUPPRESSOR controls axillary meristem formation in *Arabidopsis thaliana*. Plant J. 55:65-76 (2008)
42. Hendelman A., Stav R., Zemach H., Arazi T. The tomato NAC transcription factor SlNAM2 is involved in flower-boundary morphogenesis. J. Exp. Bot. 64:5497-5507 (2013)
43. Fabi J.P., Seymour G.B., Graham N.S., Broadley M.R., May S.T., Lajolo F.M., Cordenunsi B.R., Oliveira do Nascimento J.R. Analysis of ripening-related gene expression in papaya using an Arabidopsis-based microarray. BMC Plant Biol. 12:242 (2012)
44. Zhang B., et al. NIMA-related kinase NEK6 affects plant growth and stress response in Arabidopsis. Plant J. 68:830-843 (2011)
45. Berken A. & Wittinghofer A. Structure and function of Rho-type molecular switches in plants 46:380-393 (2008)
46. Hou H., Li J., Gao M., Singer S.D., Wang H., Mao L., Fei Z., Wang X. Genomic organization, phylogenetic comparison and differential expression of the SBP-box family genes in grape. PLoS ONE 8:e59358 (2013)
47. Schwarz S., Grande A.V., Bujdoso N., Saedler H., Huijser P. The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis. Plant Mol. Biol. 67:183-195 (2008)
48. Tesniere C., Pradal M., El-Kereamy A., Torregrosa L., Chatelet P., Roustan J.P., Chervin C. Involvement of ethylene signalling in a non-climacteric fruit: new elements regarding the regulation of ADH expression in grapevine. J. Exp. Bot. 55:2235-2240 (2004)
49. Young P.R., Lashbrooke J.G., Alexandersson E., Jacobson D., Moser C., Velasco R., Vivier M.A. The genes and enzymes of the carotenoid metabolic pathway in *Vitis vinifera* L. BMC Genomics 13:243 (2012)
50. Cakir B., Agasse A., Gaillard C., Saumonneau A., Delrot S., Atanassova R. A grape ASR protein involved in sugar and abscisic acid signaling. Plant Cell 15:2165-2180 (2003)

51. Davies C. & Robinson S.P. Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterization of cDNAs encoding putative cell wall and stress response proteins. Plant Physiol. 122:803-812 (2000)
52. Achard P., Herr A., Baulcombe D.C., Harberd N.P. Modulation of floral development by a gibberellin-regulated microRNA. Development 131:3357-3365 (2004)
53. Jones-Rhoades M.W. & Bartel D.P. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol. Cell 14:787-799 (2004)
54. Sunkar R. & Zhu J.K. Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell 16:2001-2019 (2004)
55. Llave C. MicroRNAs: more than a role in plant development? Mol. Plant Pathol. 5:361-366. (2004).
56. Llave C., Kasschau K.D., Rector M.A., Carrington J.C. Endogenous and silencing-associated small RNAs in plants. Plant Cell 14:1605-1619 (2002)
57. Park W., Li J., Song R., Messing J., Chen X. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Curr. Biol. 12:1484-1495 (2002)
58. Reinhart B.J., Weinstein E.G., Rhoades M.W., Bartel B., Bartel D.P. MicroRNAs in plants. Genes Dev. 16:1616-1626 (2002)
59. Aukerman M.J. & Sakai H. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. Plant Cell 15:2730-2741 (2003)
60. Carrington J.C. & Ambros V. Role of microRNAs in plant and animal development. Science 301:336-338 (2003)
61. Hunter C. & Poethig R.S. miSSING LINKS: miRNAs and plant development. Curr. Opin. Genet. Dev. 13:372-378 (2003)
62. Palatnik J.F., Allen E., Wu X., Schommer C., Schwab R., Carrington J.C., Weigel D. Control of leaf morphogenesis by microRNAs. Nature 425:257-263 (2003)
63. Chen X. A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 303:2022-2025 (2004)
64. Juarez M.T., Kui J.S., Thomas J., Heller B.A., Timmermans M.C.P. MicroRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. Nature 428:84-88 (2004)
65. Mallory A.C. & Vaucheret H. MicroRNAs: something important between the genes. Curr. Opin. Plant Biol. 7:120-125 (2004)
66. Jones-Rhoades M.W., Bartel D.P., Bartel B. MicroRNAs and their regulatory roles in plants. Annu. Rev. Plant Biol. 57:19-53 (2006)
67. Mette M.F., van der Winden J., Matzke M., Matzke A.J. Short RNAs can identify new candidate transposable element families in Arabidopsis. Plant Physiol. 130:6-9 (2002)
68. Emery J.F., Floyd S.K., Alvarez J., Eshed Y., Hawker N.P., Izhaki A., Baum S.F., Bowman J.L. Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. Curr. Biol. 13:1768-1774 (2003)
69. Bao N., Lye K.W., Barton M.K. MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. Dev. Cell 7:653-662 (2004)

70. Floyd S.K. & Bowman J.L. Gene regulation: ancient microRNA target sequences in plants. Nature 428:485-486 (2004)
71. Laufs P., Peaucelle A., Morin H., Traas J. MicroRNA regulation of the CUC genes is required for boundary size control in Arabidopsis meristems. Development 131:4311-4322 (2004)
72. Baker C.C., Sieber P., Wellmer F., Meyerowitz E.M. The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in Arabidopsis. Curr. Biol. 15:303-315 (2005)
73. Williams L., Grigg S.P., Xie M., Christensen S., Fletcher J.C. Regulation of Arabidopsis shoot apical meristem and lateral organ formation by microRNA miR166g and its AtHD-ZIP target genes. Development 132:3657-3668 (2005)
74. Axtell M.J., Westholm J.O., Lai E.C. Vive la différence: biogenesis and evolution of microRNAs in plants and animals. Genome Biol. 12:221-234 (2011)
75. Aida M., Ishida T., Fukaki H., Fujisawa H., Tasaka M. Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. Plant Cell 9:841-857 (1997)
76. Takada S., Hibara K., Ishida T., Tasaka M. The CUPSHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation. Development 128:1127-1135 (2001)
77. Blein T., Pulido A., Vialette-Guiraud A., Nikovics K., Morin H., Hay A., Johansen I.E., Tsiantis M., Laufs P. A conserved molecular framework for compound leaf development. Science 322:1835-1839 (2008)
78. Berger Y., Harpaz-Saad S., Brand A., Melnik H., Sirding N., Alvarez J.P., Zinder M., Samach A., Eshed Y., Ori N. The NAC-domain transcription factor GOBLET specifies leaflet boundaries in compound tomato leaves. Development 136:823-832 (2009)
79. Wang C., Leng X., Zhang Y., Kayesh E., Zhang Y., Sun X., Fang J. Transcriptome-wide analysis of dynamic variations in regulation modes of grapevine microRNAs on their target genes during grapevine development. Plant Mol. Biol. 84:269-285 (2014)
80. Muñoz C., Gomez-Talquenca S., Chialva C., Ibáñez J., Martinez-Zapater J.M., Peña-Neira Á., Lijavetzky D. Relationships among gene expression and anthocyanin composition of Malbec grapevine clones. J. Agric. Food Chem. 62:6716-6725 (2014)
81. Castellarin S.D., Pfeiffer A., Sivilotti P., Degan M., Peterlunger E., Di Gaspero G. Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. Plant Cell Environ. 30:1381-1399 (2007)
82. Powell A.L., Kalamaki M.S., Kurien P.A., Gurrieri S., Bennett A.B. Simultaneous transgenic suppression of LePG and LeExp1 influences fruit texture and juice viscosity in a fresh market tomato variety. J. Agric. Food Chem. 51:7450-7455 (2003)
83. Atkinson R.G., Sutherland P.W., Johnston S.L., Gunaseelan K., Hallett I.C. Mitra, D., Brummell D.A., Schröder R., Johnston J.W., Schaffer R.J. Down-regulation of POLYGALACTURONASE1 alters firmness, tensile strength and water loss in apple (Malus x domestica) fruit. BMC Plant Biol.12:129 (2012)
84. Deluc L.G., Grimplet J., Wheatley M.D., Tillett R.L., Quilici D.R., Osborne C., Schooley D.A., Schlauch K.A., Cushman J.C., Cramer G.R. Transcriptomic and metabolite analyses of Cabernet Sauvignon grape berry development. BMC Genomics 8:429 (2007)

85. Pilati S., Perazzolli M., Malossini A., Cestaro A., Demattè L., Fontana P., Dal Ri A., Viola R., Velasco R., Moser C. Genome-wide transcriptional analysis of grapevine berry ripening reveals a set of genes similarly modulated during three seasons and the occurrence of an oxidative burst at vèraison. BMC Genomics 8:428 (2007)
86. Lund S.T., Peng F.Y., Nayar T., Reid K.E., Schlosser J. Gene expression analyses in individual grape (*Vitis vinifera* L.) berries during ripening initiation reveal that pigmentation intensity is a valid indicator of developmental staging within the cluster. Plant Mol. Biol. 68:301-315 (2008)
87. Agudelo-Romero P., Erban A., Sousa L., Pais M.S., Kopka J., Fortes A.M. Search for transcriptional and metabolic markers of grape pre-ripening and ripening and insights into specific aroma development in three Portuguese cultivars. PLoS ONE 8:e60422 (2013)
88. Fujisawa M., Nakano T., Shima Y., Ito Y. A large-scale identification of direct targets of the tomato MADS box transcription factor RIPENING INHIBITOR reveals the regulation of fruit ripening. Plant Cell 25:371-386 (2013)
89. Bemer M., Karlova R., Ballester A.R., Tikunov Y.M., Bovy A.G., WoltersArts M., Rossetto Pde.B., Angenent G.C., de Maagd R.A. The tomato FRUITFULL homologs TDR4/FUL1 and MBP7/FUL2 regulate ethylene-independent aspects of fruit ripening. Plant Cell 24:4437-4451 (2012)
90. Fujisawa M., Shima Y., Nakagawa H., Kitagawa M., Kimbara J., Nakano T., Kasumi, T., Ito Y. Transcriptional regulation of fruit ripening by tomato FRUITFULL homologs and associated MADS box proteins. Plant Cell 26:89-101 (2014)
91. Alexander L. & Grierson D. Ethylene biosynthesis and action in tomato: a model for climacteric fruit ripening. J. Exp. Bot. 53:2039-2055 (2002)
92. Zhao D., Shen L., Fan B., Yu M., Zheng Y., Lv S., Sheng J. Ethylene and cold participate in the regulation of LeCBF1 gene expression in postharvest tomato fruits. FEBS Lett. 583:3329-3334 (2009)
93. Martel C., Vrebalov J., Tafelmeyer P., Giovannoni J.J. The tomato MADS-box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening processes in a COLORLESS NONRIPENING-dependent manner. Plant Physiol. 157:1568-1579 (2011)
94. Klipper-Aurbach Y., Wasserman M., Braunspiegel-Weintrob N., Borstein D., Peleg S., Assa S., Karp M., Benjamini Y., Hochberg Y., Laron Z. Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. Med. Hypotheses 45:486-490 (1995)
95. Jaillon O., et al. French-Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-467 (2007)
96. Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., Salzberg S.L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14:R36 (2013)

Chapter 12

Concluding remarks

SUMMARY

Grape berry development and ripening have been intensively studied at a transcriptomic level for the last decade[1] in order to characterize this complex and dynamic phenomenon but also to better understand the effect of environmental changes or agricultural practises on berry yield and quality[2-5]. The most part of these studies were realized using microarray platforms that only allow an indirect measurement of gene expression[4-10]. On the other hand, since 2010, the application of high-throughput sequencing technology on gene expression studies – namely, RNA-sequencing - has arisen a new method to obtain quantitative transcriptomic results[11-14]. Nevertheless, most of these transcriptomic studies focused on only one variety at a time and thus preventing any comparison among varieties. In this thesis, a novel study in grape berry development was reported. A RNA-Seq assay was performed on 120 berry samples, corresponding to the 10-variety berries – five red-skinned and five white-skinned grape cultivars – collected at four phenological growth stages in biological triplication (**chapter 2** and **3**).

A main part of this work was focused on discovering common transcriptomic traits during berry development, in particular:

- the genes that are expressed in all varieties during berry development (**chapter 4**);
- the genes that are commonly modulated between each subsequent pair of growth stages (**chapter 6**);
- the discovery of putative stage- and phase-specific positive biomarker transcripts (**chapter 5**);
- the establishment of the first berry development transcriptomic route (**chapter 7**);
- the identification of switch genes that induce major transcriptome reprogramming during grapevine development[15] and more particularly during grape berry ripening (**chapter 11**);
- the genes that have a very low expression throughout berry development (**chapter 10**).

Moreover, this thesis allowed unravelling transcriptomic traits specific to each skin colour group or each variety as:

- the genes that are expressed only in each colour group or specifically expressed in one variety (**chapter 4**);
- the genes that are modulated only in either red- or white-skinned berries between each subsequent pair of phenological stages (**chapter 6**);
- the genes which exhibited a greater expression in either red- or white-coloured grapes (**chapter 8**);
- the 6,003 genes which expression level is responsible for the particular harvest stage transcriptomic pattern of Barbera, Negro amaro and Refosco full-ripe berries maybe due to a high anthocyanin accumulation in their skin (**chapter 8**).

Finally, **chapter 9** reports the first exhaustive RNA-profiling of the phenylpropanoid, flavonoid and stilbene biosynthesis pathway-related genes in ten varieties, allowing correlating each gene related to phenolic compound biosynthesis with the already-known biosynthesis timing, and to identify phenylpropanoid and flavonoid biosynthesis pathway-related genes participating in the downstream pathways. This example and the provided data set could be a support to expand further gene expression studies.

BERRY TRANSCRIPTOME COMPARISON OF TEN GRAPEVINE VARIETIES

The first aim of this work was to find out the transcriptomic traits that were common in a large number of grape varieties, independently on skin colour, in order to increase understanding of the molecular basis of berry development. Consequently, ten grapevine cultivars were selected: five red-skinned – Sangiovese, Barbera, Negro amaro, Refosco, Primitivo – and five white-skinned grapes – Vermentino, Garganega, Glera, Moscato bianco and Passerina. These cultivars were chosen among several hundreds of Italian varieties for their diversity in agronomical and oenological traits, and for adaptation to different growing locations (http://catalogoviti.politicheagricole.it). Furthermore, the ten grapevine varieties were drafted on the same rootstock and were cultivated in the same experimental vineyard (Conegliano, Veneto, Italy) in order to avoid the effects of the rootstock and environmental conditions on berry phenotype. The selection of these ten cultivars allowed the detection of the expression of ~89.4% of the V1 predicted grapevine genes[16] in at least one berry sample, which corresponds to almost 9,500 genes more than the total number of expressed genes in Corvina variety[11] and ~6,000 ones more than in Shiraz cultivar[12]. Hence, carrying out RNA-Seq

assay with more varieties and more biological replicates allowed covering much more transcriptome and variety-specific patterns.

The second task of this work involved the observation of the differences between red- and white-skinned berries throughout development at a transcriptomic level. Indeed, the most evident phenotypic difference between berries is the skin colour due to anthocyanin accumulation in the skin of red grapes.

A deep green-to-maturation transcriptome shift occurs at véraison independently on skin colour and variety

This RNA-Seq analysis revealed that samples tend to cluster more according to their phase and stage of development rather than skin colour group or variety identity, thus suggesting a deep immature-to-mature transcriptome shift occurring at véraison independently on skin colour and variety. This transcriptome shift involves the suppression of diverse metabolic processes related to vegetative growth, and the induction of only a few pathways as secondary metabolic processes, and responses to abiotic and biotic stimuli. This aspect was confirmed and highlighted by distinct approaches: 1) identification of the transcriptional relationships among samples (Pearson's correlation distance and Principal Component Analysis (PCA) approaches), 2) discovery of putative stage- and phase-specific positive biomarkers (O2PLS-DA approach[9]), 3) identification of differentially expressed genes between each subsequent pair of growth stages (differential gene expression analysis approach), 4) establishment of the first grape berry development transcriptomic route – corresponding to the genes having similar patterns of expression during whole development independently on the variety – (clustering analysis approach) and 5) identification of switch genes that are key genes regulating the metabolic shift between immature and maturation phases (correlation network approach).

Establishing the grape berry development transcriptomic route permitted to group 913 differentially expressed transcripts into six clusters. The first three clusters were composed of 542, 120 and 94 transcripts respectively having a higher expression intensity during green phase, whereas the three last clusters containing 134, 3, 20 transcripts respectively exhibiting a peak of expression during the maturation phase. This approach allowed identifying genes involved in the main biological processes occurring during berry development including the shutdown of the photosynthesis-

related genes either at pea-sized berry stage or at the end of the green phase, and the modulation of genes associated with sugars metabolism and transport such as two sucrose synthase genes *VvSUS3* and *VvSUS5*[17] and a tonoplast monosaccharide transporter *VvTMT3*[18] – both employed during berry formation – and the hexose transporter gene *VvHT6/TMT2*[18] and a SWEET gene[19], which seems to be rather involved in sugar accumulation during maturation phase. Moreover, the water inflow occurring during maturation phase, responsible for berry enlargement and softening, could be linked to the up-regulation of the aquaporin *TIP1-3* gene[20] just after véraison when berry starts enlarging again after the lag phase. Auxin-related genes were found down-regulated after véraison, confirming the role of auxins as enhancers of berry formation and inhibitors of ripening[21]. Furthermore, each growth event – berry formation by rapid cell division and cell enlargement, berry expansion and softening during maturation phase – requests a distinct set of cell wall-related genes and also different genes belonging to large gene family as the expansin gene family. Interestingly, only transcription of few genes related to secondary metabolism was found.

As a part of this thesis, the red-skinned berry transcriptomes were also analysed to unravel the correlation network arising from the dataset using a novel approach developed by Palumbo and colleagues[15]. By studying grapevine (*Vitis vinifera*) and tomato (*Solanum lycopersicum*) gene expression atlases and the red-grape transcriptomic data set during the transition from immature to mature growth, a category named “fight-club hubs” was characterized by a marked negative correlation with the expression profiles of neighbouring genes in the network. A special subset named “switch genes” was identified, with the additional property of many significant negative correlations outside their own group in the network. Switch genes showed a significant increase in their expression in ripening berries and were found involved in multiple processes, including both transcription factors and enzymes with fundamental roles in berry development at the onset of ripening. This suggests that switch genes represent master regulators of development that – acting in concert with their direct targets – determine a core function that allows plant organs to complete developmental transition. In particular, switch genes direct targets were 1,266 genes – called neighbouring genes – down-regulated between the immature and maturation phases, showing again that immature-to-mature transition involves mostly the suppression of diverse metabolic processes related to vegetative growth. Therefore, the identified

putative master regulators supposedly act by inhibiting – directly or indirectly – the transcription of genes expressed in the preceding developmental phase in order to promote berry maturity.

High anthocyanin content seems to influence transcriptome behaviours of red grape varieties during maturation phase and more particularly at harvest stage

Another investigation carried out within this work was the first study of transcriptomic differences between red- and white-skinned berries. Red skin colour is caused by the biosynthesis of anthocyanins in the skin of red grapes. Anthocyanin production is regulated by the additive activity of two MybA-type transcription factors: VvMybA1 and VvMybA2. In white cultivars, the insertion of the *Gret1* Gypsy-type retrotransposon into the promoter of *VvMybA1* prevents *VvMybA1* transcription[22], while a mutation of *VvMybA2* results in a double STOP codon responsible for the truncation of MybA2 protein, rendering it inactive[23]. Thus, the goal of this work was to study the transcriptomic traits specific to each skin colour group – other than the differences regarding the above-stated anthocyanin-related genes.

Few genes were identified as only expressed in either red or white grapes throughout development. Moreover, the major part of them were devoid of a predictive functional annotation. The comparison of the differentially expressed genes (DEGs) between each subsequent pair of growth stages allowed the discovery of genes that were statistically up- and down-regulated only in red- or white-skinned berries. Red-skinned berries, besides their skin colour, had more specific DEGs than white-skinned berries throughout development. In addition, skin colour-specific DEGs between pea-sized berry and prevéraison stages were very few, indicating that skin colour seemed to have a greatest influence on the transcriptome during the maturation phase.

After véraison, skin colour seemed to influence Refosco, Barbera and Negro amaro berry transcriptomes that clustered separately from the other varieties at harvest, suggesting that these three red grape varieties might have specific transcriptional traits at maturity. This aspect was highlighted by Person's correlation distance and PCA approaches. Furthermore, red- and white-coloured berries described different patterns during maturation phase. In fact, performing two distinct PCA analyses with either the 20 red-coloured samples or the 20 white-coloured

samples showed that white-coloured berry transcriptomes grouped together, whereas red-skinned berry transcriptomes were separated, suggesting that specific transcriptomic traits were present only among red varieties samples after véraison.

Differentially expressed genes between red- and white-skinned berries at the two ripening stages were selected using t-test analyses. Thus, 443 DEGs were identified at the end of véraison, representing 2% of the expressed genes, corresponding to 382 and 61 genes with a higher expression level in red- and white-skinned samples respectively. Concerning the harvest stage, we found 837 DEGs between red- and white-skinned samples, i.e. 4% of the data set, including 536 and 289 genes that were highly expressed in red- and white-skinned samples respectively. Nevertheless, removing of these differentially expressed genes was not sufficient to explain red grape transcriptome patterns.

Identification of almost 6,000 transcripts specifically involved in the maturation process of Barbera, Negro amaro and Refosco red-coloured berries was allowed by a novel analysis, consisting in the extraction of transcript loadings from PCA outcome and the detection of the so-called loading threshold-breaking point, at which red-variety samples separation in the PCA was not no longer reproducible. Moreover, anthocyanin content measurements confirmed that these three red grapes are characterized by a high anthocyanin concentration at full-maturity, as Mattivi and colleagues[24] already showed. Consequently, although differences in anthocyanin accumulation are present, gene expression and modulation related to phenylpropanoid biosynthetic pathway only was found to be insufficient to explain the differences between red- and white-grape transcriptomes. However, it is possible that the high anthocyanin accumulation in the skin, forming a dark screen on berry surface, could influence the expression of genes belonging to other biological processes and then be responsible of transcriptome pattern differences in full-ripe red-skinned grapes.

The higher accumulation of photosynthesis-related transcripts in full-ripe white-skinned berries and Sangiovese and Primitivo berries in comparison to the other red grapes (Barbera, Negro amaro and Refosco), suggests that anthocyanin content could influence the shutdown of this process. In fact, anthocyanin accumulation in the skin, forming a sort of dark screen, could diminish photosynthetic processes or accelerate their turning off already started at the onset of ripening. Furthermore, ripening red grapes expressed preferentially the hexose transporter gene *VvHT5* and the cell wall invertase gene *VvINV4*[25] whereas *VvHT2* exhibited a higher expression in white

grapes. In addition, Barbera, Negro amaro and Refosco full-ripe berries accumulated less *VvHT2*, *VvHT3/HT7* [25] and *VvHT6/TMT2*[18] transcripts than the other varieties. Nevertheless, high sugar content of Barbera, Negro amaro and Refosco full-ripe berries could be explained by a greater expression of *VvINV4*. Moreover, influence of sugar supply on anthocyanin accumulation has been demonstrated by several studies[26-29] but little is known about the effect of anthocyanin production on sugar accumulation. In this work, anthocyanin production and sugar accumulation genes seemed to be co-regulated at harvest in Barbera, Negro amaro and Refosco berries, and thus influencing together their maturity. Moreover, many abiotic and biotic stress-related genes had a greater expression in red grapes and more particularly in Barbera, Negro amaro and Refosco berries compared to white grapes. Considering that Barbera, Negro amaro and Refosco berries did not reach full-maturity at the same time – Barbera full-ripe berries were collected early September as the white-skinned berries, and Negro amaro and Refosco berries reached full-maturity two weeks later like Sangiovese and Primitivo – they did not underwent the same environmental conditions at the end of ripening. Therefore, we can wonder whether this phenomenon was related to anthocyanin accumulation. Finally, further metabolomic investigations should be necessary to confirm the stated hypothesis but could also indicate which of anthocyanin content or composition predominantly influenced transcriptome behaviours throughout maturation phase.

PERSPECTIVES

This work represents the first RNA-Seq assay done with such a large number of varieties, developmental stages and replicates. The generated data – 100-nucleotide reads – could be used to improve the annotation of the reference genome or to discriminate gene family members. In fact, the grapevine genome sequence revealed several examples of expanding gene families[30-31]. Among them, some may have an impact on berry yield and quality, and then on the organoleptic properties of wine. The provided data set could be used to choose candidate gene for stable transformation. For example, genes that were found as specifically expressed in only one variety but with no hit or an unknown function could represent good candidates for further functional characterization analysis in order to discover if their expression is responsible for varietal physiological and morphological characteristics, as berry shape

and size or flavour production and composition. Moreover, several transcription factor genes have been identified as either biomarker transcripts, or members of the berry development transcriptomic route or switch genes. Genes belonging to the same category represent good candidates as direct targets of these transcription factor genes.

Finally, the availability of such a RNA-Seq data set could pave the way to the creation of genomic tools supporting vineyard management, as the detection of molecular changes that may affect grape yield and quality but also the response to the environment.

REFERENCES

1. Tornielli G.B, Zamboni A., Zenoni S., Delledonne M, Pezzotti M. Transcriptomics and metabolomics for the analysis of grape berry development. In The Biochemistry of the Grape Berry, vol. 1. Edited by Gerós H., Chaves M.M., Delrot S. Sharjah: Bentham Science pp. 218-240 (2012)
2. Mori K., Goto-Yamamoto N., Kitayama M., Hashizume K. Loss of anthocyanins in red-wine grape under high temperature. J Exp Bot 58(8): 1935-45 (2007)
3. Deluc L.G., Quilici D.R., Decendit A., et al. Water deficit alters differentially metabolic pathways important flavor and quality traits in grape berries of Cabernet Sauvignon and Chardonnay. BMC Genomics 10: 212 (2009)
4. Dal Santo S., Tornielli G.B., Zenoni S., Fasoli M., Farina L., Anesi A., Guzzo F., Delledonne M. and Pezzotti M. The plasticity of the grapevine berry transcriptome. Genome Biol. 14: r54 (2013)
5. Pastore C., et al. Increasing the source/sink ratio in Vitis vinifera (cv Sangiovese) induces extensive transcriptome reprogramming and modifies berry ripening. BMC genomics 12:631 (2011)
6. Grimplet J., Deluc L.G., Tillet L.R., et al. Tissue-specific mRNA expression profiling in grape berry tissues. BMC Genomics 8: 187 (2007)
7. Deluc L.G., Grimplet J., Wheatley M.D., et al. Transcriptomic and metabololite analyses of Cabernet Sauvignon grape berry development. BMC Genomics 8:429 (2007)
8. Pilati S., Perazzolli M., Malossini A., et al. Genome-wide transcriptional analysis of grapevine berry ripening reveals a set of genes similarly modulated during three seasons and the occurrence of an oxidative burst at véraison. BMC Genomics 8: 428 (2007)
9. Zamboni A., Di Carli M., Guzzo F., et al. Identification of putative stage-specific berry biomarkers and omics data integration into networks. Plant Physiol 154(3): 1439-59 (2010)
10. Fasoli M., Dal Santo S., Zenoni S., Tornielli G.B., Farina L., Zamboni A., Porceddu A., Venturini L., Bicego M., Murino V., Ferrarini A., Delledonne M., Pezzotti M. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. Plant Cell 24:3489-3505 (2012)
11. Zenoni S., Ferrarini A., Giacomelli E., et al. Characterization of transcriptional complexity during berry development of Vitis vinifera using RNA-seq. Plant Physiol 152(4): 1787-95 (2010)
12. Sweetman C., Wong D.C.J., Ford C.M., Drew D.P. Transcriptome analysis at four developmental stages of grape berry (Vitis vinifera cv. Shiraz) provides insights into regulated and coordinated gene expression. BMC Genomics 13:691 (2012)
13. Da Silva C., Zamperin G., Ferrarini A., et al. 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(12):4777-4788 (2013)
14. Degu A., Hochberg U., Sikron N., et al. Metabolite and transcript profiling of berry skin during fruit development elucidates differential regulation between Cabernet Sauvignon and Shiraz cultivars at branching points in the polyphenol pathway. BMC Plant Biology 14:188 (2014)
15. Palumbo M.C., Zenoni S., Fasoli M., Massonnet M., Farina L., Castiglione F., Pezzotti M., Paci P. Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative

switch genes that induce major transcriptome reprogramming during grapevine development. Plant Cell 26(12):4617-4635 (2014)

16. Forcato C. Gene prediction and functional annotation in the Vitis vinifera genome. PhD Thesis, Universita' Degli Studi Di Padova (2010)
17. Xin H., Zhang J., Zhu W., Wang N., Fang P., Han Y., Ming R., Li S. The effects of artificial selection on sugar metabolism and transporter genes in grape. Tree Genetics and Genomes 9, 1343–1349 (2013)
18. Wormit A., Trentmann O., Feifer I., et al. Molecular identification and physiological characterization of a novel monosaccharide transporter from Arabidopsis involved in vacuolar sugar transport. Plant Cell 18:3476-3490 (2006)
19. Chen L.Q., Qu X.Q., Hou B.H., Sosso D., Osorio S., Fernie A.R., Frommer W.B. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science 335:207-211 (2012)
20. Maurel C., Verdoucq L., Luu D.T., Santoni V. Plant Aquaporins: Membrane Channels with Multiple Integrated Functions. Annual Review of Plant Biology 59:595-624 (2008)
21. Böttcher C., Keyzers R.A., Boss P.K., Davies C. Sequestration of auxin by the indole-3-acetic acid-amido synthetase GH3-1 in grape berry (Vitis vinifera L.) and the proposed role of auxin conjugation during ripening. J Exp Bot. 61:3615-3625 (2010)
22. Kobayashi S., Goto-Yamamoto N., Hirochika H. Retrotransposon-induced mutations in grape skin color. Science 304(5673):982 (2004)
23. Walker A.R., et al., White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant J. 49:772 (2007)
24. Mattivi F., Guzzon R., Vrhovsek U., Stefanini M., Velasco R. Metabolite profiling of grape: Flavonols and anthocyanins. J Agric Food Chem. 54(20):7692-7702 (2006)
25. Hayes M.A., Davies C., Dry I.B. Isolation, functional characterization, and expression analysis of grapevine (Vitis vinifera L.) hexose transporters: differential roles in sink and source tissues. J Exp Bot. 58:1985-1997 (2007)
26. Gollop R., Farhi S., Perl A. Regulation of the leucoanthocyanidin dioxygenase gene expression. Plant Science 161:579-588 (2001)
27. Gollop R., Even S., Colova-Tsolova V., Perl A. Expression of the grape dihydroflavonol reductase gene and analysis of its promoter region. Trends in Plant Science 53:1397-1409 (2002)
28. Zheng Y., Tian L., Liu H., Pan Q., Zhan J., Huang W. Sugars induce anthocyanin accumulation and flavanone 3-hydroxylase expression in grape berries. Plant Growth Regulators 58:251-260 (2009)
29. Dai Z.W., Meddar M., Renaud C., Merlin I., Hilbert G., Delrot S., Gomès E. Long-term in vitro culture of grape berries and its application to assess the effects of sugar supply on anthocyanin accumulation. J Exp Bot. 65(16):4665-4677 (2014)
30. Jaillon O., Aury J.M., Noel B., et al. French-Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161): 463-7 (2007)
31. Velasco R., Zharkikh A., Troggio M., et al. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. Plos One 2(12): e136 (2007)

ACKNOWLEDGEMENTS

A PhD thesis is tricky to write but I think that acknowledgements are even worst! I would like to start by saying that these three years were amazing and I would like to thank all people that supported me during this nice experience.

First, I would like to thank my PhD director, Prof. Mario Pezzotti, for having believed in me and entrusted this huge project to me. You made me feel welcome into your lab, but also sometimes into your life, more particularly the first year with my crazy eyes… You continuously pushed me to my very limits, allowing me finishing this project (ok, don't worry, we will write the article asap!!!^^).

I would like to thank all the people of my lab: my labmates: Marianna, Silvia, Anne-Marie, Eric(hett)a; the Giambas: EriKa, Laura, Alessandra; the GADs: Linda, Matilde, Elisa, Roby; the Polverari's: Arianna and Peter, oups Pietro!^^ But also my "co-directors" (not written in the paper, but you deserve it) Giamba and Sara. You welcomed me despite my bad English and my nonexistent Italian. The beginning was a bit complicated but you were all very kind and patient with me, particularly Anne-Marie and Laura for dealing with all the paper work.

Thank you, my labmates, you brightened up these three years! Anne-Marie, being your labmate was an honour. I learnt a lot about life seated right at your right. You helped me to improve my English, and you learned me to be STRONGLY POLITE!!!! :D So much laughs with you, thank you. Thanks Silvia for helping me throughout these three years, even if there were just small helpings, I always knew that I could count on you. Thanks Erichetta for… being you! ☺ You are a good girl and I'm sure you will succeed to transform your plants. I don't know how or when, but I know that you are enough persevere to do it! ;) The last labmate, Mari. What can I say? You were always here for me, you always took the time for helping me, despite your crazy over-booked schedule and your part-time in US. You are the kindest person I know and I would like to thank you from the bottom of my heart.

Thank you, Sara and Giamba! You followed me and my project during all this PhD. Thanks Sara for you kindness, you will stay Mamma Sara forever! Thanks Giamba for your questions, you always wanted to go deeper (sometimes too much…).

But, by the way, you pushed me to reflect and to reconsider my goals (sometimes too much…^^). I will never forget your whistling!

Grazie a tutti! Era una straordinaria esperienza! Ci vediamo tra poco! ☺

Now, let switch English with French!

Avant tout chose, j'aimerais remercier le professeur Serge Delrot de m'avoir mise en contact avec Mario pour cette thèse. Même si je serais bien restée à Bordeaux, ce doctorat à l'étranger, et qui plus est en Italie, m'a faite mûrir (si on se réfère à une baie de raisin, je pense avoir passé la véraison ! Il me reste encore toute la maturation à achever, mais laissons le temps au temps).

Maintenant, je propose de sortir les mouchoirs : séquence émotion !

Merci à mes parents. Papa, Maman (Maman, Papa, vous choisissez !^^), sans vous, je n'aurais jamais pu faire ce long parcours. Cette thèse fut longue (Putain, trois ans !) mais vous m'avez toujours soutenue et vous avez toujours cru en moi. J'ai toujours fait en sorte que vous soyez fiers de moi et j'espère du fond du cœur que c'est le cas aujourd'hui. Cette thèse, elle est pour vous ! (Enfin, un peu pour moi aussi !!!^^)

Merci à toute ma petite famille (mon cocon, comme j'aime bien l'appeler). Ces moments en famille m'ont toujours boostée, et bien souvent m'ont permis de remonter dans l'avion ! Merci à ma sœur Stéph, Ludo, et à Loulou bien sûr, d'être venu me voir. Ce sont des moments que je n'oublierais jamais ! Comment oublier qu'on a fait le tour de Vérone pour trouver une trancheuse ??? XD Merci à mon frère, Benoit, pour m'avoir fait découvrir la bio. Même si je n'ai pas choisi la même voie (J'aime pô les bêtes !), tu as toujours respecté mon choix (plus ou moins…). Tu as été mon exemple et tu le resteras. Merci à Emma et Roudoudou d'être tout simplement là ! Je vous aime tous ! (Ok, là on sort les mouchoirs…!!!)

Merci aussi à mes ami(e)s ! Merci de m'avoir suivie durant ces trois années. Vive Facebook et Skype ! Merci à mes sœurs de cœur, ma Nis et ma poulette Audrey, d'être venue me voir et d'avoir toujours œuvré pour que l'on puisse se voir à chaque fois que je rentrais en France. Vous voir était une vraie bouffée d'air !

Enfin, le meilleur pour la fin… Merci à mon amour, mon meilleur ami, l'homme de ma vie, Julien. Ces trois années loin de toi furent longues et courtes à la fois. Tu as été mon pilier. Tu m'as toujours encouragée, soutenue, épaulée…
Durant ces trois années, nous avons sillonné l'Italie ensemble. Je te propose à présent une nouvelle aventure : la vie ensemble.