

Full Paper

# A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation

Darren Chern Jan Wong<sup>1†</sup>, Rudolf Schlechter<sup>2†</sup>, Alessandro Vannozzi<sup>3</sup>, Janine Höll<sup>4</sup>, Ibrahim Hmam<sup>3</sup>, Jochen Bogs<sup>5,6</sup>, Giovanni Battista Tornielli<sup>7</sup>, Simone Diego Castellarin<sup>1</sup>, and José Tomás Matus<sup>8\*</sup>

<sup>1</sup>Wine Research Centre, University of British Columbia, Vancouver, BC, Canada, <sup>2</sup>Department of Biology, ETH Zurich, 8092 Zurich, Switzerland, <sup>3</sup>Department of Agronomy, Food, Natural Resources, Animals, and Environment, University of Padova, 35020 Legnaro, Padova, Italy, <sup>4</sup>Centre for Organismal Studies Heidelberg, University of Heidelberg, 69120 Heidelberg, Germany, <sup>5</sup>Dienstleistungszentrum Laendlicher Raum Rheinpfalz, Breitenweg 71, Viticulture and Enology Group, 67435 Neustadt/W, Germany, <sup>6</sup>Fachhochschule Bingen, Berlinstr. 109, 55411 Bingen am Rhein, Germany, <sup>7</sup>Department of Biotechnology, University of Verona, 15 - 37134 Verona, Italy, and <sup>8</sup>Center for Research in Agricultural Genomics CSIC-IRTA-UAB-UB, Barcelona 08193, Spain

\*To whom correspondence should be addressed. Tel. +34 935636600 (Ext 3236). Fax. +34 935636601. Email: [tomas.matus@cragenomica.es](mailto:tomas.matus@cragenomica.es)

<sup>†</sup>Equal contribution.

Edited by Prof. Kazuki Saito

Received 8 March 2016; Accepted 17 May 2016

## Abstract

R2R3-MYB transcription factors (TFs) belong to a large and functionally diverse protein superfamily in plants. In this study, we explore the evolution and function of this family in grapevine (*Vitis vinifera* L.), a high-value fruit crop. We identified and manually curated 134 genes using RNA-Seq data, and named them systematically according to the Super-Nomenclature Committee. We identified novel genes, splicing variants and grapevine/woody-specific duplicated subgroups, suggesting possible neo- and sub-functionalization events. Regulatory network analysis ascribed biological functions to uncharacterized genes and validated those of known genes (e.g. secondary cell wall biogenesis and flavonoid biosynthesis). A comprehensive analysis of different MYB binding motifs in the promoters of co-expressed genes predicted grape R2R3-MYB binding preferences and supported evidence for putative downstream targets. Enrichment of *cis*-regulatory motifs for diverse TFs reinforced the notion of transcriptional coordination and interaction between MYBs and other regulators. Analysis of the network of Subgroup 2 showed that the resveratrol-related *VviMYB14* and *VviMYB15* share common co-expressed *STILBENE SYNTHASE* genes with the uncharacterized *VviMYB13*. These regulators have distinct expression patterns within organs and in response to biotic and abiotic stresses, suggesting a pivotal role of *VviMYB13* in regulating stilbene accumulation in vegetative tissues and under biotic stress conditions.

**Key words:** wine, cell wall, ripening, radiation, stress

## 1. Introduction

Grapevine (*Vitis vinifera* L.) is one of the world's most important fruit crops and has been an integral part of human history since it was domesticated 6,000–8,000 yrs ago.<sup>1</sup> Grapes provide a vast number of metabolites with well-known health attributes, and the importance of these metabolites for the human diet underlines the importance of understanding their genetic and physiological basis. The sequence of the grapevine genome<sup>2</sup> offered several lines of evidence for expanding gene families related to wine quality and health attributes. Among these, there is an over-representation of genes involved in the synthesis of secondary metabolites, such as the stilbene (anticarcinogenesis) synthase family.<sup>3,4</sup> Genes related to flavonoid synthesis (nutraceuticals) are also highly duplicated in *Vitis*, including the R2R3-MYB family of transcription factors (TFs),<sup>5</sup> which regulates various aspects of plant metabolism and development.

The R2R3-MYB family was expanded by massive duplication events within plants.<sup>6,7</sup> Genome-wide analyses on eudicotyledonous species show that this family has between 70 and 200 members.<sup>6</sup> The main amplification of this TF family may have occurred before the separation of monocots and eudicots,<sup>8</sup> although some species underwent later duplications associated with their polyploid genome origins (e.g. maize, rice and grapevine<sup>2,8</sup>). In addition to these expansion and diversification episodes, tandem duplication of subgroups have also been predicted,<sup>5,9–11</sup> many of which could have been fixed during the course of domestication.

The first accession of the grapevine genome ( $\times 8.4$ ) allowed the identification of 108 genes in the R2R3-MYB family,<sup>5</sup> although these genes were not assigned to any particular nomenclature. Despite its size, less than 10% of the members of this family have been functionally characterized. With the exception of VviMYB60 (regulating stomata aperture<sup>12</sup>) and VviMYB80 (controlling stamen development,<sup>13</sup> all other R2R3-MYBs characterized to date are involved in regulating the phenylpropanoid pathway. These include activators of the anthocyanin,<sup>14,15</sup> flavonoid,<sup>16–18</sup> proanthocyanidin (PA)<sup>19</sup> and stilbene<sup>20</sup> pathways, and also members of the C2-repressor motif clade.<sup>18,21</sup>

To determine the processes in which other members of the R2R3-MYB family are involved, several network inference approaches can be used to predict gene function and prioritize candidate genes for subsequent characterization (e.g. using reverse genetic tools). Three main network inference methods have been adopted to infer gene function in crop species<sup>22</sup>: (i) transfer of network links between evolutionarily conserved genes (*associalogs*); (ii) genome context similarities (phylogenetic and gene neighbourhood profiling) and (iii) analysis of gene co-expression networks (GCN). GCNs are based on the 'guilt-by-association' principle, where genes involved in biologically related pathways/processes exhibit comparable expression dynamics across a wide range of experimental conditions. GCNs provide a comprehensive overview of gene–gene relationships across multiple experimental conditions (condition-independent GCN) or in determined conditions (condition-specific GCN, e.g. abiotic stress only)<sup>23</sup> in both conserved and species-specific pathways.<sup>24</sup> In addition, analysis of *cis*-regulatory elements/motifs (CRE) in GCN, linking CRE to co-expressed genes and to genes of given biological processes, has also been used to better understand plant transcriptional control.<sup>25,26</sup>

In this study, we propose a revised classification of the grapevine R2R3-MYB family and use omics data to correct gene annotations and study duplications and diversification processes within the family. We used microarray and RNA-Seq data to understand the regulation of expression and alternative splicing. We constructed a global GCN from publicly available data, encompassing a vast collection of tissue/developmental and stress-related conditions. Identification of tightly interconnected relationships was coupled to *cis*-regulatory motif enrichment analysis, carried out in 1 kb upstream regions of all co-expressed genes. These combined analyses allowed us to identify orthologues of known genes, to discover new genes involved in regulating metabolic routes and to suggest common and unique biological roles for members of this family.

## 2. Materials and methods

### 2.1. Model corrections in the grapevine R2R3-MYB family and search of new annotations within the genome

A consensus sequence of the DNA-binding domain (106 residues<sup>5</sup>) was used to run a search of all putative R2R3 MYB genes in the 12Xv1 genome prediction (<http://genomes.cribi.unipd.it/>). The BLASTP and PSI-BLAST suites were used to perform similarity searches in the predicted proteome database. Obtained models were compared to the genes found in Matus *et al.* (2008)<sup>5</sup> and Wilkins *et al.* (2009).<sup>27</sup> Annotated protein sequences were tested for the presence of R2 and R3 repeats using a locally run RPS-BLAST on the Conserved Domain Database (CDD)<sup>28</sup> and a profile Hidden Markov model (HMM) scan from the HMMER suite on the Pfam-A database.<sup>29,30</sup> Annotations were then aligned and compared with previously published RNA-Seq data<sup>31,32</sup> and with a *de novo* assembled grapevine transcriptome (Wong *et al.*, unpublished), using the pairwise or multiple (CLUSTALW2, CLUSTAL OMEGA or MUSCLE) sequence alignment tools available at EMBL-EBI (<http://www.ebi.ac.uk/Tools/msa/>).

### 2.2. Phylogeny reconstructions and identification of conserved protein motifs

Deduced grapevine proteins were aligned against the full predicted amino acid sequences of proteins belonging to the entire *Arabidopsis thaliana* R2R3 MYB-family (126 members). *Arabidopsis* maMYB, an endoplasmic reticulum-anchored R2R3-MYB TF was used as out-group. Multiple sequence alignments (gap open penalty of  $-2.9$ ) were performed using the MUSCLE algorithm-based AlignX module from MEGA5 software.<sup>33</sup> Phylogenetic trees were constructed using the neighbour-joining (NJ) and maximum likelihood (ML) methods [NJ was computed using the no-differences model, ML was computed using the Whelan and Goldman model with Gamma distributed (G+I) rates among sites], both with partial deletion gap treatments. Reliability of tree nodes was evaluated with 1000 bootstraps. Resulting trees were visualized with FigTree (<http://tree.bio.ed.ac.uk/software/figtree>). MYB subgroups were defined both by bootstrap values and by the presence of C-terminal motifs as identified in multiple EM for motif elicitation suite. Protein sequences (excluding DNA binding domains) of R2R3-MYB subgroups were used

to discover motifs with an expected value lower than  $2 \times 10^{-30}$  (<http://meme.nbc.net/>), using different search parameters: (i) mucilage and acidification of vacuoles ‘MAV’ subgroup: 6–13 residues (min–max length), zero or one occurrence per sequence of motifs and four motifs maximum, (ii) MYBPA1-related subgroup ‘PA1’: 6–12 residues (min–max length), zero or one occurrence per sequence and four motifs maximum, (iii): S2Neighbour ‘S2N’: 6–9 residues (min–max length), zero or one occurrence per sequence and four motifs maximum.

### 2.3. Genome structure and duplication analysis

Chromosomal positions and amino acid sequences for all (29,971) protein-coding genes were retrieved from the 12Xv1 prediction of the grapevine genome (<http://genomes.cribi.unipd.it/grape/>). Identification of gene duplication patterns were carried out using the MScanX software suite.<sup>34</sup> Briefly, self-genome all-vs-all BLASTP analysis was first performed according to the recommended settings of MScanX using NCBI-BLAST-2.2.30+.<sup>35</sup> MScanX assigns the various types of duplication event (dispersed, proximal, tandem and segmental) to all the R2R3 MYB genes based on chromosomal positions and the blast alignments.

### 2.4. Expression data and gene co-expression network construction

Publicly available 29K NimbleGen Whole-genome array data from 14 experiments representing 664 arrays were retrieved from Gene Expression Omnibus.<sup>36</sup> Raw intensity data were background-adjusted and summarized with the Robust Multi-Array method using *oligo*<sup>37</sup> and averaged according to the biological replicates. The resulting expression matrix (219 arrays  $\times$  29,000 genes) was used to construct a global highest reciprocal rank (HRR) network.<sup>24</sup> HRR for all genes were calculated in R (<http://www.r-project.org>), by first considering the Pearson correlation coefficient (PCC) ( $r$ ) value between all gene pairs and subsequently ranking them according to the formula;  $HRR(x, y) = [\max(\text{rank}(x, y), \text{rank}(y, x))]$ , whereby  $\text{rank}(x, y)$  is the transformed rank of gene  $y$  according to gene  $x$ 's co-expression list and *vice versa* for  $\text{rank}(y, x)$ . Using a ‘bottom-up’ (targeted) approach, individual MYB-co-expressed gene (MYB-coex) sub-network was constructed by extracting the co-expressed genes within the top 100 HRRs. The MYB-TF gene (MYB-MYB) sub-network was constructed by merging all the individual MYB-co-ex sub-networks and retaining only the MYB genes and their respective co-expression relationships. Visualization of the various sub-networks was carried out in Cytoscape version 3.0.<sup>38</sup>

### 2.5. Over-representation analysis of gene function and promoter regions

Functional annotation and MapMan BINs assignment of transcripts based on the 12Xv1 prediction were performed using Mercator.<sup>39</sup> Significant enrichment of MapMan BIN categories (Adj.  $P$  value  $< 0.05$ ) within MYB co-expressed genes was determined by Fisher's exact test adjusted with Bonferroni correction for multiple testing corrections. For comparison, enrichment (False discovery rate  $< 0.05$ ) of gene ontology (GO) terms based on incremental enrichment analysis was performed using gProfileR.<sup>40</sup>

Using the 12Xv1 assembly, a 1 kb region upstream of the transcriptional start site of all predicted genes (29,971) was retrieved by the Biomart tool available at EnsemblPlants (<http://plants.ensembl.org/index.html>). Promoter sequences containing more than 10% of

ambiguous sequences (indicated as ‘N’) were removed, resulting in a total of 29,839 promoter sequences. A comprehensive and up-to-date list of plant TF binding sites were scanned in promoter regions of MYB-co-expressed genes and the promoter background (29,839 promoters). These sites were inferred from: (i) protein-binding microarrays, representing 76 CRE from 25 TF families,<sup>41</sup> (ii) a selection of consensus sequences from ‘Plant *cis*-acting regulatory DNA elements database’ (PLACE)<sup>42</sup> (50 in total; representing organ-specific and endogenous/environmental/signalling TF binding sites) and (iii) 17 functionally validated R2R3-MYB binding sites (consensus and variant).<sup>43–45</sup> Enrichment for CRE, expressed as  $P$  value (or  $-\log_{10}(P \text{ value}) * 10$ ), was based on the hypergeometric test according to Ma et al.<sup>25</sup> using a custom-made pipeline in R (<http://www.r-project.org>).

### 2.6. Expression analysis of *VviMYB13*, *VviMYB14* and *VviMYB15* by quantitative RT-PCR

Non-stressed grapevine organs (young and old leaves, stems and roots from cv. Chardonnay and cv. Shiraz) and berries from a developmental series (cv. Pinot Noir and cv. Riesling) were collected from a greenhouse and a commercial vineyard, respectively, in Neustadt/W, Germany (49° 22' 9" south, 8° 10' 28" east). Sampling methods, RNA extraction, cDNA synthesis and qPCR reactions were conducted as previously described.<sup>20</sup> Leaf discs from greenhouse-grown *V. vinifera* cv. Shiraz plants were subjected to different biotic and abiotic stresses: wounding, UV-C irradiation and downy mildew infection, as described by Vannozzi et al.<sup>4</sup> Leaf discs of 15 mm of diameter were punched from healthy leaves. The punching of discs was considered as a wounding treatment *per se*, and as a control for other treatments. The UV-C treatment was achieved by exposing the abaxial surface of the discs to 30W UV-C light for 10 min at a distance of 10 cm. Downy mildew (*Plasmopara viticola*) infection was carried out by spraying a solution containing downy mildew sporangia at a concentration of  $10^5$  sporangia  $\text{ml}^{-1}$ . Stressed discs were incubated upside down on moist filter paper in large Petri dishes at 22 °C under 12 h light/12 h dark conditions until sampling (dried with absorbent paper and immediately frozen in liquid nitrogen until RNA extraction).

Expression of grape MYB genes from Subgroup 2 was determined using primers MYB13\_CDS\_qF5'-CGTTTTCATCAGAGCTGGAGG, MYB13\_3'UTR\_qR 5'-TCCCTGTATGATCTCCCCTCT, MYB15\_3'UTR\_qF 5'-CATTGACCAAGAAAGCAAAAA and MYB15\_3'UTR\_qR 5'-AGCAACTTTTCTAAGTCAATTC. MYB14 was amplified with primers described in Holl et al.<sup>20</sup> The cycle threshold values were corrected to *VviUBIQUITIN1* (TC32075), *VviEF1- $\alpha$*  (EC959059) and *VviGAPDH* (CB973647). Normalization against reference genes was conducted as described previously.<sup>46</sup>

### 2.7. Analysis of stilbenes in grapevine organs

Stilbenes were extracted from 50 mg of non-stressed sampled organs. Extraction was done in methanol and measured by reverse-phase High performance liquid chromatography as previously described by Höll et al.<sup>20</sup> Quantification of the glycosylated-resveratrol, piceid, was calculated from calibration curves prepared from commercial stilbene standard *trans*-piceid (PhytoLab GmbH & Co. KG, Germany). Concentrations were calculated as nmol/g FW with error bars indicating the standard error from three biological replicates and two independent experiments.

### 3. Results and discussion

#### 3.1. Novel genes and subgroups in the grapevine R2R3-MYB family: updated nomenclature

The first survey of R2R3-MYB family members in grapevine identified 108 genes in the  $\times 8.4$  genome draft sequence.<sup>5</sup> Subsequent studies reported increased and contradictory numbers of members,<sup>27,47</sup> urging the need to define the true size of this family in the latest genome accession. Here, by inspecting MYB R2R3 DNA-binding signatures (using BLAST-coupled to-HMMER analysis) in the CRIBI 12Xv1 genome accession, we retrieved a total number of 134 sequences (Supplementary Table S1). We compared these to the VCOST gene predictions (annotated on the 12Xv2 of the genome assembly) and uploaded to the Online Resource for Community Annotation of Eukaryotes database (ORCAE, <http://bioinformatics.psb.ugent.be/orcae/>) as part of the grapevine genome curation effort.

We curated gene models by comparing gene structures with known R2R3-MYB genes and by using RNA-Seq data, which also validated the expression of these annotations (Supplementary Dataset 1). The search for mRNA sequences in comprehensive RNA-Seq datasets of *V. vinifera*<sup>31,32</sup> (and Wong *et al.*, unpublished) allowed us to overcome annotation errors due to the constraints of the *ab initio* prediction algorithm. Inspection of these sequences supported the existence 196 transcribed units, including several splicing forms, for a total of 115 MYB genes (Supplementary Table S1). The use of RNA-Seq data has largely been effective for improving gene family characterizations.<sup>48,49</sup> Here, we corrected several models in terms of the length and presence of their DNA-binding and/or C-terminal domains (Supplementary Dataset 1), such as MYB135 (VIT\_14s0036g00460), where the RNA-Seq data showed that the gene encodes an integral R3 repeat and a larger C-terminal domain, or MYB82A (VIT\_11s0016g05650) and MYB186 (VIT\_13s0064g00960), corrected in the complete extension of their DNA binding domains. The MYB13 (VIT\_05s0049g01010) gene possesses three exons instead of four (in the 12Xv1 model a portion of the third exon was mistakenly predicted to be an intron). Aside of the 134 genes, we were able to amend two incorrectly annotated R2R3-MYB genes, which actually correspond to R1R2R3R4-MYB genes (VIT\_17s0000g02730 and VIT\_16s0039g01750). The model VIT\_02s0025g02210, which we originally identified,<sup>5</sup> was not found in this re-analysis, as conserved domains were no longer present (NCBI Conserved Domain Database).

A small fraction of the models identified (e.g. those not found in RNA-Seq data) may represent pseudogenes that have lost their expression or protein-coding ability. This is the case of MYBA4 (VIT\_02s0033g00370), which lacks part of the DNA binding domain due to a frameshift mutation. MYBA4 is located in a cluster of MYB genes in chromosome 2 (chr2), next to the berry colour locus, but it has no expression.<sup>15</sup> It may have arisen from a tandem duplication of MYBA1 or MYBA2, which are true regulators of anthocyanin biosynthesis in the berry.<sup>14,15</sup> In addition to MYBA4, other predictions from the same chr2 cluster could represent pseudogenes, particularly MYBA8 (VIT\_02s0033g00380), which it is expressed (corroborated by RNA-Seq data) but has an incomplete DNA-binding domain (63 residues, Supplementary Dataset 1).

We generated phylogenetic trees of the grape and Arabidopsis R2R3-MYB families using the confirmed or modified full-predicted proteins derived from the RNA-Seq data (Supplementary Fig. S1). Where no data were available to confirm the model, we used the annotated sequence from the Pinot Noir reference genome (CRIBI annotation). Grape proteins with a complete R2R3 domain (excluding

MYBA4 and MYBA8) were compared with 126 Arabidopsis proteins to assess cases of one-to-one orthology. As in our previous work,<sup>5</sup> we observed an expansion of subgroups related to phenylpropanoid/flavonoid biosynthesis (e.g. anthocyanin and PA). These subgroups are either activators (Subgroups 5 and 6) or repressors (Subgroup 4, also known as C2 repressor clade) of the pathway. Grape genotypes with expanded MYB genes related to flavonoid synthesis may have been selected during grapevine natural selection and also later throughout domestication. However, this may have resulted from branch-specific selections, since not all flavonoid-related groups are expanded. For example, while Arabidopsis has three members in flavonol-related Subgroup 7, grapevine has only two, which may be because grapes synthesize other molecules related to biotic and abiotic stress in addition to flavonols such as stilbenes, which are not present in Arabidopsis.

Based on our comparison of phylogenetic trees and according to the guidelines from the Super-Nomenclature Committee for Grape Gene Annotation,<sup>50</sup> we propose a nomenclature for the grapevine R2R3-MYB family based on sequence similarities with the Arabidopsis genes (retaining the same name in cases of close orthology), but also maintaining the original names for R2R3-MYB TFs that have already been characterized (e.g. MYBA and MYBPA for all anthocyanin and PA-related genes, respectively). Genes with no clear orthology with a single Arabidopsis MYB member were named MYB $x$ , where  $x$  is a consecutive number starting from 135 (Supplementary Table S1). Genes at the same phylogenetic distance from a single Arabidopsis gene were differentiated by a letter, as in the case of MYB5A and MYB5B. Additional genes identified in future re-annotations or novel genes arising in specific cultivars should be numbered from MYB197 onwards.

We compared the MYB subgroups found in the grapevine to those previously described in Arabidopsis.<sup>45,51</sup> We also compared the grapevine/Arabidopsis MYB tree with other trees obtained by combining Arabidopsis (individually) with the apple, tomato, poplar and Eucalyptus families (Supplementary Table S2, Supplementary Fig. S2). We generally observed conserved topologies, except for the glucosinolate clade (Subgroup 12, present in Arabidopsis but absent in the other species analysed), which regulates the biosynthesis of secondary metabolites found exclusively in the Brassicaceae family.<sup>52</sup> We observed some differences in subgroup composition with respect to other plant species (i.e. number of members). Some subgroups were larger in some species than in Arabidopsis, such as Subgroups 5 and 14, while in others they were smaller (e.g. Subgroup 19).

Phylogenetic comparison showed that AtMYB5 homologues cluster in a different clade than AtMYB123 homologues, in contrast to previous works who generally group them in Subgroup 5.<sup>45,51</sup> We named this independent Subgroup 'MAV' in reference to the roles of AtMYB5,<sup>53</sup> PhPH4<sup>54</sup> and VviMYB5A/5B.<sup>18</sup> We also identified several subgroups that were missing in Arabidopsis and present in other plant species, including the flavonoid-related PA1 and the S2 neighbour (S2N) Subgroups, which correspond to the woody preferential groups WPS-II and WPS-IV/V in Eucalyptus, respectively.<sup>47</sup> The PA1 subgroup is the closest to MAV, and VviMYBPA1 is its only characterized member.<sup>19</sup> Protein motif discovery analyses showed that this group has two conserved motifs (PA1-1 and PA1-2) that are not shared with the MAV Subgroup (Supplementary Fig. S3). The S2N Subgroup shows a close relationship to Subgroup 2 due to high conservation of the DNA-binding domain. Moreover, S2N is divided into two clades, where S2Na presents one of the two phenylalanine-tryptophan (FW) conserved motifs found in Subgroup 2 (both form part of a single motif).<sup>51</sup> Despite not having an FW motif, MYB135

lies very close to Subgroup 2 because of greater conservation in the DNA-binding domain.

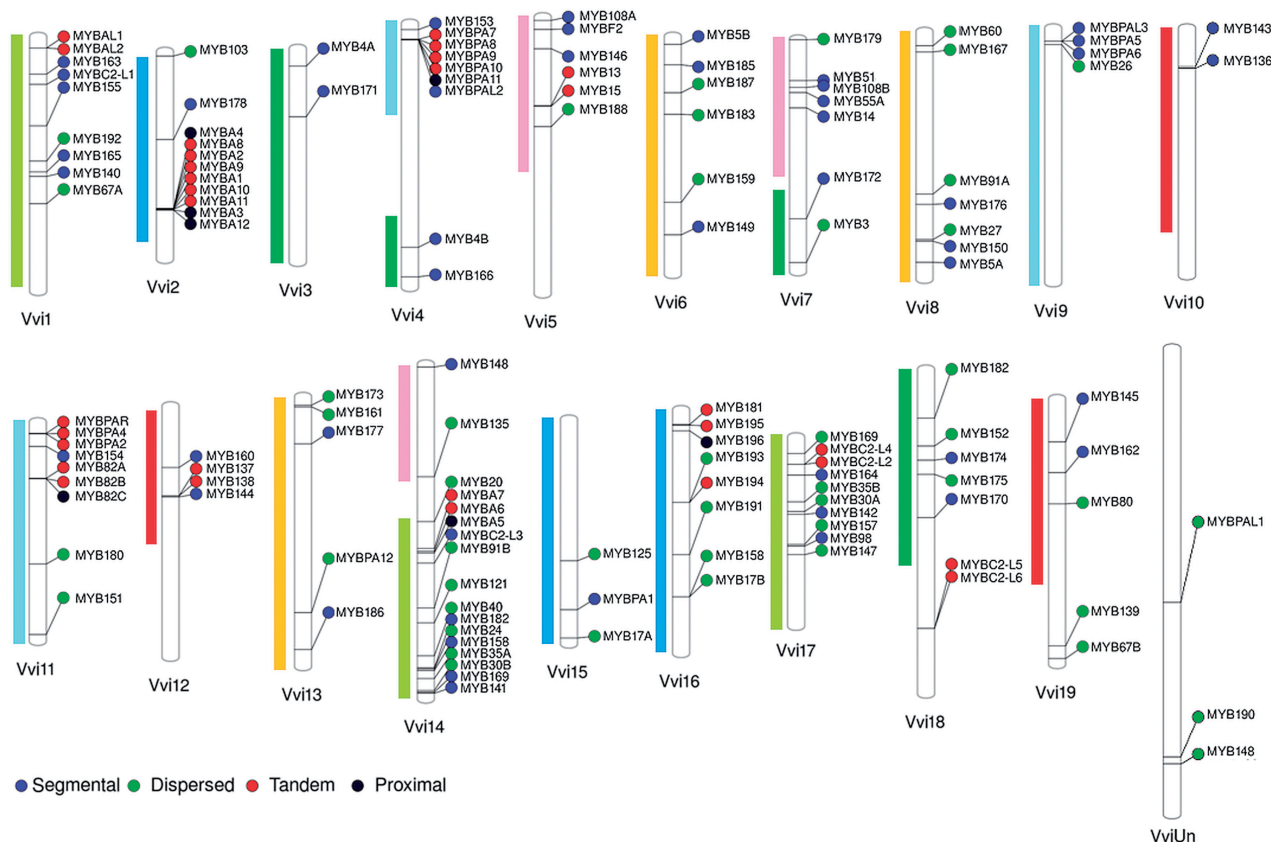
### 3.2. Tandem duplications specifically contributed to the expansion of many grape R2R3-MYB genes linked to wine quality traits

We successfully mapped a total of 131 grapevine R2R3-MYB genes to the 19 *V. vinifera* chromosomes; while *MYBPAL1*, *MYB190* and *MYB148* are still unallocated (uncharacterized chromosome). The locations of these genes are visualized in the chromosome based on the 12Xv1 assembly (Fig. 1). We used MCScanX software to further determine the duplication types that might have given rise to the expansion of this large gene family, and found that segmental (36.7%; 50 of 134) and dispersed (36.7%; 50 of 134) duplications appear to be the most prominent contributors (Supplementary Table S1). Many segmental and dispersed duplicated MYBs are located in collinear regions on chr1/14/17 (green), chr6/8/13 (orange) and chr5/7 (pink) (Fig. 1). A recent study showed that syntenic regions of related chromosomes in land plants such as maize, tomato, Arabidopsis and poplar, have large proportions of R2R3-MYB paralogue correspondence (25%, 49%, 48% and 38%, respectively).<sup>6</sup> This reinforces the idea that R2R3-MYB expansion in land plants, including grapevine (37%, reported here), primarily occurs through segmental duplication. In addition, dispersed duplications, defined as paralogues that do not share synteny and/or are not in close proximity in the

corresponding chromosome, may also contribute to expansion of gene families in plants.<sup>55</sup>

We observed a strong influence of dispersed mechanisms on R2R3-MYB expansion (37%), in concordance with wide-spread repetitive/transposable elements within the grapevine genome (approximately 40% of the total genome).<sup>2</sup> Indeed, many of these elements have been linked to grapevine's genomic variability.<sup>56</sup> Tandem duplication (21.3%; 29 of 134) also contributed to family expansion (Fig. 1). We observed a dense map of MYB genes in tandem on chromosomes 2, 4 and 11, such as for *MYBA* (9 out of 14 genes) and *MYBPA* (5 out of 15 genes). Expansion of R2R3-MYB families related to anthocyanin and PA synthesis have been related to artificial selection (i.e. breeding), possibly because of higher flavonoid concentration in grapes and wines.<sup>5</sup> In addition, our analysis also revealed that tandem duplication might have contributed to the expansion of the C2-repressor clade, albeit on a smaller scale. Expansion of this subgroup might also be linked to selection for wine quality traits, as recent studies show that MYB C2-repressors have a central role in regulating the phenylpropanoid and flavonoid biosynthetic pathways.<sup>18,21</sup> C2-repressors may go through transcriptional sub-specialization after duplication, as suggested for grapevine<sup>18</sup> and other species such as clover, where a high concentration of flavonoids is also a desirable trait.<sup>57</sup> Finally, we assigned expansion by proximal duplication to 7 of 134 genes (5.14%), although these might also be part of tandem duplications.

Gene duplications related to segmental and tandem duplications reported here are consistent with the results of a survey of segmental



**Figure 1.** Distribution of R2R3-MYB genes in the grapevine chromosomes. Parologue regions of the grapevine genome were coloured according to Jaillon et al.<sup>2</sup> Genes predicted to arise from dispersed, proximal, tandem or segmental duplications are shown. Details related to potential paralogues generated by tandem duplication or collinear pairs within syntenic regions are summarized in Supplementary Table S1.

and tandem duplications of R2R3-MYB genes across land plants,<sup>6</sup> with minor differences attributable to the updated/improved annotation of *Vitis* family. These duplication patterns indicated that rapid expansion of R2R3-MYB genes in grapevine might be strongly linked to adaptation strategies to challenging environments in order to accomplish species-specific functions.

### 3.3. Intron retention as a conserved mechanism of alternative splicing in the grape R2R3-MYB family

Alternative splicing is a tightly regulated mechanism that contributes to proteome diversity, and can be studied using transcriptomic approaches. In grapevine, the NimbleGen microarray platform does not assay splice variants, and the Affymetrix system detects only some mRNA alternative isoforms (with differences between the 16 k and custom arrays). Deep sequencing can overcome these limitations by providing details of all splice forms. Inspection of the RNA-Seq data used in this study, which comes from a vast collection of tissues, organs and stages of berry development, allowed us to detect more than two mRNA isoforms for ~35% of the family, of which a quarter of cases were represented by three or more splice forms (Supplementary Dataset 1). These results are consistent with estimates of overall alternative splicing events in *Vitis*.<sup>58</sup> The most common type of splicing event we observed (>90%) was intron retention, generating premature termination codons. We observed a relatively similar distribution of retention for introns 1 and 2, although the latter was more common in some subgroups (e.g. Subgroup 4; Supplementary Fig. S4). Since the DNA-binding domain of grape R2R3-MYB genes is generally split between Exons E1, E2 and E3 (for class I) or E1 and E2 (for classes II and III),<sup>5</sup> the truncated protein sequences derived from intron-retaining isoforms have partial DNA-binding domains (e.g. MYB-like, with one of the two R repeats) and incomplete or abolished C-terminal regions. We observed few cases of exon skipping, such as *MYB147* (*VIT\_17s0000g09080*), where its alternative spliced isoform (*MYB147.2*) lacks half of the R2 repeat but maintains the R3 repeat and its C-terminal region (Supplementary Dataset 1).

Intron retention accounts for a large proportion of alternative splicing events in Arabidopsis, rice and grapevine, and can generate relatively abundant isoforms with premature termination codons.<sup>58,59</sup> We observed three spliced isoforms of the stilbene-related *MYB15* gene (*VIT\_05s0049g01020*; Supplementary Dataset 1). The retention of intron 2 in *MYB15.2* generates two potential open reading frames (ORFs), each encoding only a segment of the complete *MYB15* protein (ORF1 encodes R2 and part of R3, while ORF2 encodes the rest of the protein). In contrast, *MYB15.3* retains intron 1 and encodes a single R2 protein without any C-terminal region. As reported by Li *et al.*<sup>60</sup>, *AtMYB59* and *AtMYB48* undergo alternative splicing and generate MYB-like transcripts whose 5' untranslated regions (UTR) are insufficient to initiate translation. Further studies will determine if truncated isoforms of *VviMYB15* can be translated and if those proteins have a regulatory role in stilbene biosynthesis.

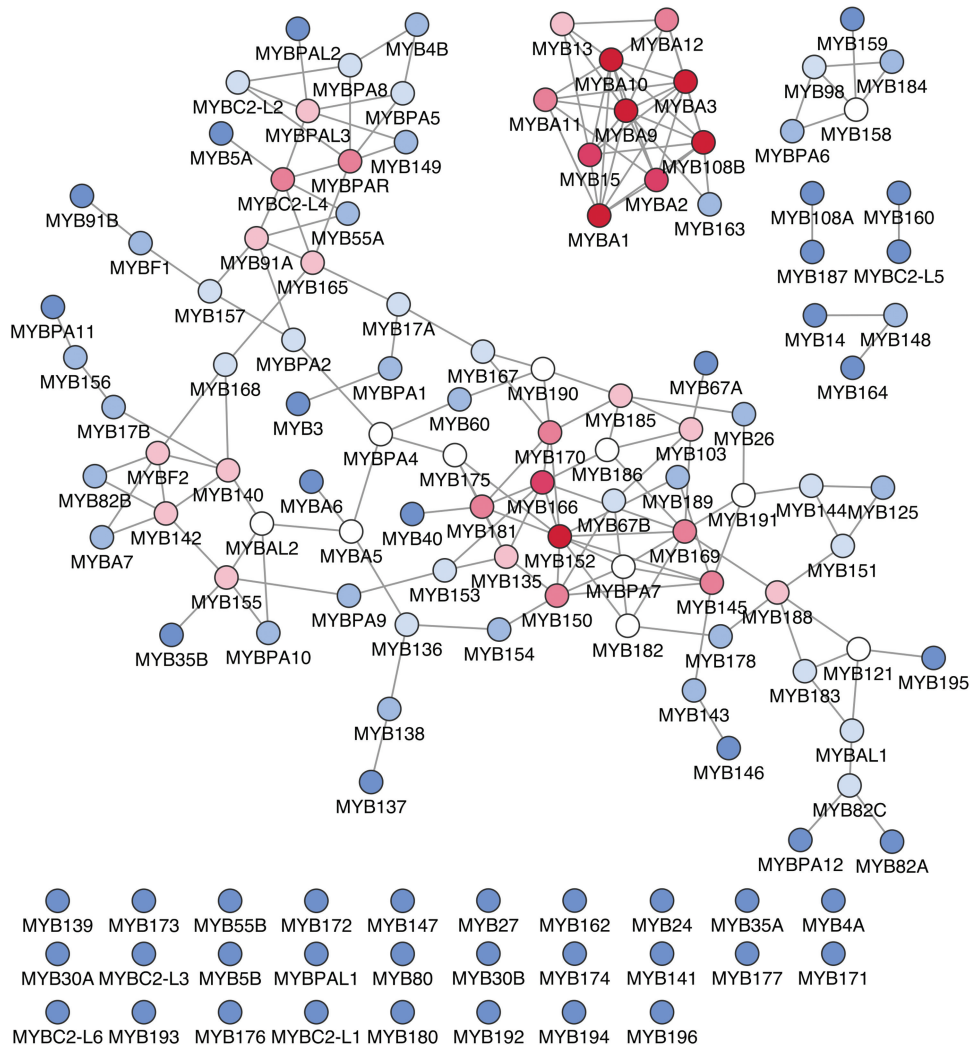
Transcription and its regulation are two processes with high-energy demands. Alternative splicing should, therefore, be a precisely regulated event since the production of futile products can be detrimental to plant fitness. This comes in part from the observation that alternative splicing is widespread in plants and plays an important role in regulating gene expression (reviewed by Staiger and Brown<sup>61</sup>). *VviMYB24* (*VIT\_14s0066g01090*) is an interesting case as its long intron 2 (2.9 kb) is retained in one of the two mRNA isoforms. This gene has a tissue-specific expression in reproductive

organs,<sup>5</sup> and is highly induced by UV-B in ripening berries.<sup>62</sup> Thus, alternative splicing of its intron 2 may be important for its regulation.

### 3.4. Expression modules potentially regulated by grapevine R2R3-MYBs are primarily enriched with transcription factors and secondary metabolic pathway genes

GCN analysis is not yet as widely adopted in crop species as in model plants. In grapevine, very few studies have adopted these approaches. Candidate genes involved in regulating the accumulation of anthocyanin<sup>63</sup> and organic acid<sup>64</sup> in the berry have been prioritized based on a previous GCN resource developed by Wong *et al.*<sup>65</sup> Another study constructed a berry-specific GCN that prioritized master regulators of phase transition in berry ripening.<sup>66</sup> To reliably predict the function and potential downstream targets of grapevine R2R3-MYB TFs, we constructed a MYB GCN using measures of HRR. Rank of correlations generates more robust results<sup>24,67</sup> than correlation-based metrics, such as the PCC. We performed a new analysis increasing the number of dataset considered (15 experiments, 219 conditions averaged from 664 NimbleGen arrays; Supplementary Table S3), and obtained better representation of development and stress than the current *Vitis* Co-expression database (VTCdb, 480 NimbleGen arrays, eight experiments).<sup>65</sup> Among the additional experiments considered, we included transcriptional studies of berry circadian oscillation and day/night transcriptional program, berry response to *Botrytis cinerea* infection, heat, UV-B and drought/water-deficit stress (Supplementary Tables S3 and S4, Supplementary Fig. S5). To assess MYB co-expression, we chose the top 100 HRR-ranked co-expressed genes for 132 grape R2R3-MYBs (excluding *MYBA4* and *MYBA8*; Supplementary Table S5). This threshold provided a reasonable limit for further functional studies and a manageable list of co-expressed genes while providing statistical power for inferring co-expression relationships.<sup>23,68</sup>

Initial inspection of the exclusive MYB-MYB GCNs showed a strong correlation between 102 annotated MYB gene pairs. Seventy-nine genes and 127 edges formed a central cluster while a second cluster contained 11 MYB genes connected by 35 edges (Fig. 2; Supplementary Table S5). The fact that a large proportion of MYB genes are reciprocally correlated suggests different types of relations: (i) role redundancy, (ii) partnership in the regulatory mechanisms and (iii) shared response or signalling pathways, in which an MYB may mediate neighbouring MYB-expression (i.e. acting as orchestrator). Within the central cluster, two hub-like regions were prevalent (top 5% of highest degree nodes; node degree > 6). Genes such as *MYB152* (*VIT\_18s0001g08470*) and *MYB166* (*VIT\_04s0044g01380*) showed a high degree of relationship in hub 1, while *MYBPAR* (*VIT\_11s0016g01300*) and *MYBC2-L4* (*VIT\_17s0000g02650*) represent high-degree of relationship in hub 2. MYBA genes predominate in a smaller second cluster in addition to *MYB13* (*VIT\_05s0049g01010*) and *MYB15* (*VIT\_05s0049g01020*) from Subgroup 2, and *MYB108B* (*VIT\_07s0005g01950*) from Subgroup 20. Surprisingly, five MYBA genes (*MYBA5*, *MYBA6*, *MYBA7*, *MYBAL1* and *MYBAL2*) from the expanded anthocyanin subclade (Subgroup 6, Supplementary Fig. S1) were not connected to the main MYBA cluster within the MYB-MYB GCN. This is due to clear divergence among these genes in expression patterns across various tissues/organs and stress conditions (Supplementary Table S4), which may indicate sub- and/or neo-functionalization after duplication, as commonly observed in regulatory genes from Arabidopsis.<sup>69</sup>



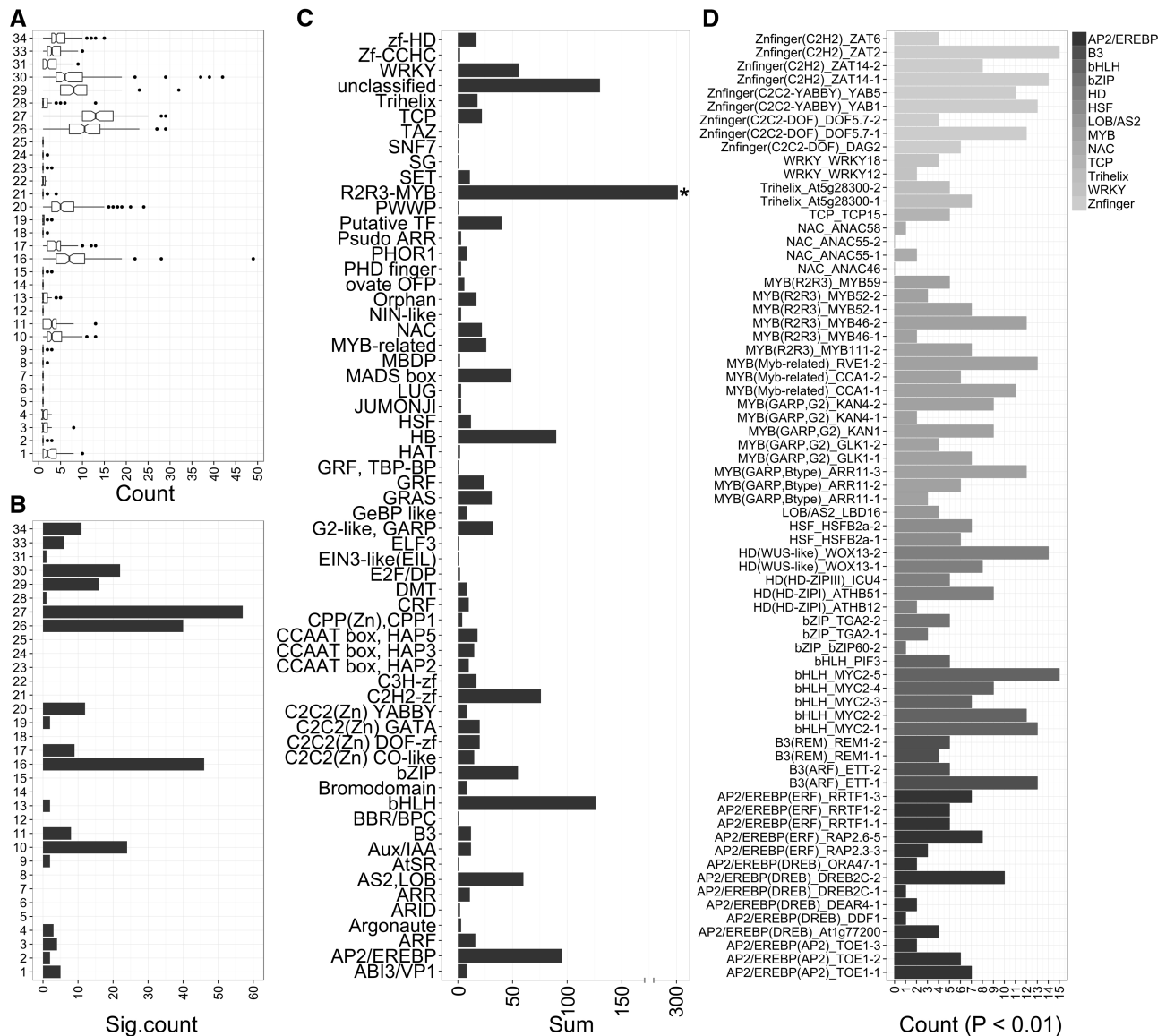
**Figure 2.** The MYB-MYB GCN based on the top 100 HRR. Nodes represent genes and edges indicate significant co-expressions between *R2R3-MYB* genes. The coloured nodes indicate the degree of connections of each node with its neighbours (red/dark indicates a degree  $\geq 5$ , white indicates a degree of 4 and blue/light grey indicates a degree  $\leq 3$ ).

Despite this diversification in expression, the group of *MYBA5*, *MYBA6* and *MYBA7* genes located in chr14, is indeed related to the regulation of anthocyanin synthesis (Matus et al., unpublished results).

To determine the biological processes that each *R2R3-MYB* gene might be involved in, we performed over-representation analysis of the MapMan BIN terms<sup>70</sup> for the top 100 co-expressed genes of all MYB GCNs (Fig. 3A and B; Supplementary Table S6). This analysis revealed that RNA regulation (BIN27) was the most commonly enriched term (Adj. *P* value  $< 0.05$ , for 57 out of the 130 MYB networks), suggesting that *R2R3-MYBs* may orchestrate plant responses by regulating or interacting with proteins related to transcription, RNA stability, etc. Terms related to metabolism, especially secondary metabolism (BIN16), hormone metabolism (BIN17) and stress (BIN20) were also commonly presented in the top 100 co-expressed genes. Processes related to the cell wall (BIN10), lipid (BIN11) and photosynthesis (BIN1) were also represented, albeit to a lesser extent (Fig. 3A and B; Supplementary Table S6). Among the most commonly associated ‘regulation by RNA’ BIN term, 35 MYB GCNs were also enriched with ‘MYB domain TF family’ (BIN27.3.

25), followed by ‘secondary metabolism’ (BIN16) with 44 MYB GCNs (Supplementary Table S6). Among secondary metabolic pathways, the most enriched terms were related to phenylpropanoid derivatives and terpenes (Supplementary Fig. S6). We conducted GO enrichment analysis using the R ‘gProfileR’ package and found similar results to the MapMan analysis (Supplementary Fig. S7; Supplementary Table S7). Together, these results highlight that *R2R3-MYB* TFs primarily participate in secondary metabolic processes and in transcriptome modulation.

Analysis of terms related to secondary metabolism showed that many co-expressed MYB genes were associated with the flavonoid pathway (Supplementary Tables S6 and S7). Grapevine *MYBA1-A2* regulate anthocyanin accumulation by modulating *UDP-GLUCOSE:FLAVONOID 3-O-GLUCOSYLTRANSFERASE (UFGT)* expression.<sup>13</sup> We were able to associate eight *MYBA* genes (*MYBA1*, *MYBA2*, *MYBA3*, *MYBA7*, *MYBA9*, *MYBA10*, *MYBA11* and *MYBA12*) and four additional MYB TFs (*MYB108B*, *MYB15*, *MYBPAL10* and *MYB142*) with anthocyanin-related genes. For example, *MYBA1-2* were co-expressed with *UFGT* (*VIT\_16s0039g02230*) while *MYB108B* was co-expressed with *ANTHOCYANIN*



**Figure 3.** Distribution of MapMan BIN terms and enriched CRE within the MYB GCN. (A) Boxplot represents the number of co-expressed genes associated with a BIN term. (B) Bar graph depicting the number of MYB GCN which are enriched (Adj.  $P$  value  $< 0.05$ ) with at least one BIN term. Only high-level MapMan BIN terms were considered to provide an overview. BIN1, PS; BIN2, major CHO metabolism; BIN3, minor CHO metabolism; BIN4, glycolysis; BIN5, fermentation; BIN6, gluconeogenesis/glyoxylate cycle; BIN7, OPP; BIN8, TCA/org transformation; BIN 9, mitochondrial electron transport/ATP synthesis; BIN10, cell wall; BIN11, lipid metabolism; BIN12, N-metabolism; BIN13, amino acid metabolism; BIN13, amino acid metabolism; BIN14, S-assimilation; BIN15, metal handling; BIN16, secondary metabolism; BIN17, hormone metabolism; BIN18, co-factor and vitamin metabolism; BIN19, tetrapyrrole synthesis; BIN20, stress; BIN21, redox; BIN22, polyamine metabolism; BIN23, nucleotide metabolism; BIN24, biodegradation of xenobiotics; BIN25, C1-metabolism; BIN26, misc; BIN27, RNA; BIN28, DNA; BIN29, protein; BIN30, signalling; BIN31, cell; BIN33, development; BIN34, transport. (C) Distribution of TF family terms (BIN27.3) within the MYB GCN. Bar graph depicts the total number of genes within each TF family (sub-BIN) from the sum of MYB co-expressed genes. (D) Distribution of enriched CRE, shown as the number of MYB GCNs enriched with the respective CRE at  $P < 0.01$ . Different grey tones are depicted for each TF family to aid visualization.

*GLUTATHIONE S-TRANSFERASE 4 (GST4, VIT\_04s0079g00690)*, the gene involved in anthocyanin transport into vacuoles.<sup>71</sup> In addition, *MYB15* was associated with anthocyanin pathway through co-expression with several *MYBA* genes. As *MYB15* has only been related to the stilbene pathway,<sup>18</sup> this observation suggests that *MYB15* and *MYBA* genes are induced by similar conditions (e.g. stress) and not necessarily involved in regulating the same branch of the pathway. On the other hand, *MYB30A*, which was not co-expressed with a large set of anthocyanin-related genes, is co-expressed with *ANTHOCYANIN O-METHYLTRANSFERASE 1*

and 2 (*AOMT1-2, VIT\_01s0010g03510* and *VIT\_01s0010g03490*), the genes responsible for anthocyanin methylation.<sup>72</sup>

The robustness of the networks generated was also evident from inspecting regulators of the PA pathway. *MYBPA1*, a positive regulator of PA biosynthesis in grapevine,<sup>19</sup> was co-expressed with structural genes of the flavonoid pathway, including a *FLAVONOID 3',5'-HYDROXYLASE (VIT\_08s0007g05160)*, a *LEUCOANTHOCYANIDIN REDUCTASE (VIT\_01s0011g02960)* and a *ANTHOCYANIDIN REDUCTASE (ANR, VIT\_00s0361g00040)*, which corroborates the activation of PA-



flavonoid pathway gene promoters seen by Bogs et al.<sup>19</sup> The role of MYBPAR<sup>73</sup> was corroborated by its co-expression with ANR, a co-expressed target shared also with MYBPA1. In the case of MYBPA2<sup>74</sup> co-expression list, we found a grape homologue of the Petunia AN11 gene (*VIT\_08s0007g02920*). This gene was recently proposed as candidate for the regulation of flavonoid synthesis in the grape berry skin.<sup>63</sup>

We also tested additional biochemical pathways of biotechnological interest for association with R2R3-MYB proteins. For example, six MYB GCNs (those of *MYB26*, *MYB103*, *MYB185*, *MYB186*, *MYB190* and *MYB191*) were enriched with terms related to lignification/secondary cell wall formation (Supplementary Fig. S8, Supplementary Tables S6 and S7). These six MYB GCNs have two genes in common, a cellulose biogenesis-related gene (*COBRA*, *VIT\_17s0000g05060*) and an NAC secondary wall thickening-promoting factor (*NST1*, *VIT\_02s0025g02710*). Arabidopsis homologues of these genes have been shown to be part of a large suite of secondary cell wall regulators (e.g. MYBs) and master switches (e.g. *NST1*).<sup>75,76</sup> This supports the robustness of our GCN and points to the existence of a comparable transcriptional network in grapevine. Terms related to wax metabolism were significantly enriched in six MYB GCNs, suggesting that R2R3-MYB TFs are also involved in this pathway (e.g. by regulating plant cuticle-structural genes). This may be relevant for grapevine MIXTA-like MYBs, *VviMYB140* and *VviMYB17A*, as MIXTA-like TFs regulate both epidermal cell morphology and the biosynthesis of waxes, very long-chain fatty acids and cutins in Arabidopsis and crown violets.<sup>77</sup> Interestingly, *MYBA5* and *MYBAL2* GCNs were enriched with terms related to wax metabolism and anthocyanin metabolism, which supports sub-/neo-functionalization, as described earlier (Supplementary Tables S6 and S7).

Inspection of TF families in MYB-GCNs showed that, in addition to members of the same R2R3-MYB family, there was a strong representation of the basic helix-loop-helix (bHLH, 126 occurrences) and APETALA2/ethylene-responsive binding protein (AP2/EREBP, 95 occurrences) families (Fig. 3C). Protein-protein interaction and cross-regulation is common between MYB and bHLH TFs. For instance, MYB-bHLH protein complexes are essential for regulating epidermal cell fate, including flavonoid synthesis,<sup>78</sup> trichome development<sup>79</sup> and also in flower and seed development.<sup>80</sup> In grapevine, several studies using *LUCIFERASE* reporters have demonstrated that the activation of flavonoid structural genes only occurs when MYB and bHLH TFs are co-transformed.<sup>15,17,19</sup> In addition to the well-known MYB-bHLH interactions, a recent study provided evidence of a new transcriptional complex of AP2/EREBP-R2R3-MYB proteins involved in regulating lignification.<sup>81</sup> These examples support the frequent co-expression of R2R3-MYBs with other TF families such as bHLH and AP2/EREBP.

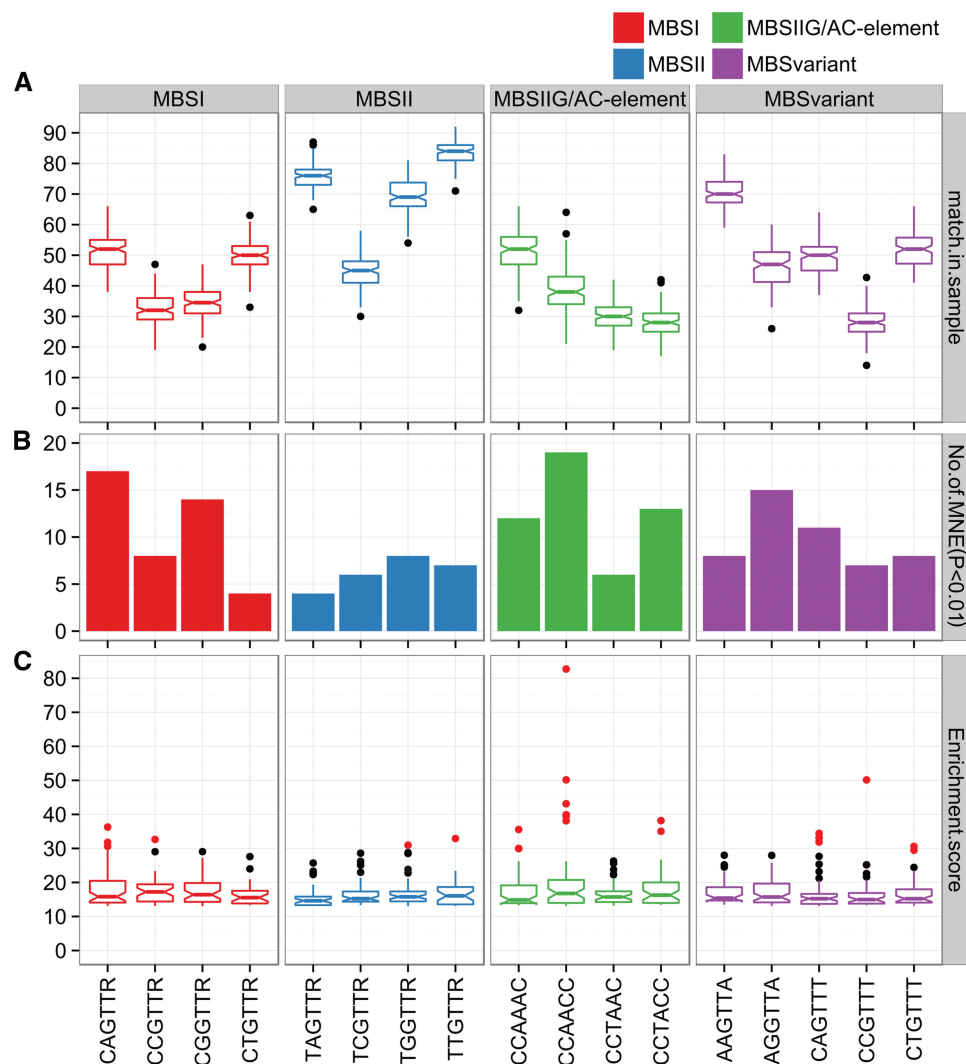
### 3.5. Analysis of *cis*-regulatory elements in promoter regions of R2R3-MYB co-expressed genes supports coordinated regulation with other transcription factors, defines DNA-binding specificity and prioritizes co-expressed target genes

In an attempt to resolve how the presence of CRE affect the transcriptional control of biological processes, several studies have coupled enrichment of regulatory motifs in promoters of co-expressed genes with enrichment of gene function.<sup>25,26</sup> We scanned the promoters of MYB co-expressed genes using a comprehensive and up-to-date plant CRE list (See Materials and methods). Our results provide some clues to explain the observed co-regulation of different TF

families within the MYB GCNs (Fig. 3D; Supplementary Fig. S9; Supplementary Table S8). The observed enrichment (at  $P$  value < 0.01) of several CREs such as bHLH (MYC2-, 3- and 4-), homeobox (WUS), zinc finger (C2H2) and AP2/EREBP binding sites generally coincided with the prevalence of associated TF co-regulation within each MYB GCN (Fig. 3C). This indicates that TFs from other families may coordinate biological responses via a synchronic and cooperative mechanism together with MYB proteins, as previously shown for some characterized MYB TFs.<sup>78,80,81</sup>

To dissect the organization of CRE motifs in the promoter regions of co-expressed genes, we considered three main MYB binding types (type I—CNGTTR-, II—TNGTTR- and IIG/AC-elements—CCWAMC-),<sup>43,44</sup> as well as several other reported variants<sup>45</sup> (Fig. 4). Some motifs were more prevalent in the promoters of co-expressed genes, such as CAGTTR ( $\bar{x}$ : 52) for type I, TTGTTR ( $\bar{x}$ : 84) for type II, CCAACC ( $\bar{x}$ : 38) for type IIG/AC-elements, and AGGTTA ( $\bar{x}$ : 47) for MYB variants (Fig. 4A; Supplementary Tables S9 and S10). In addition, we observed that the number of MYB GCNs enriched with these motifs was generally between 7 and 20, with a  $P$  value < 0.01 (Fig. 4B). Deviations in the number of MYB enriched GCNs with each CRE ( $P$  < 0.01) were prominent for types 'I' (17 for CAGTTR versus 4 for CTGTTR) or 'IIG/AC-element' (19 for CCAACC versus 6 for CCTACC), which may reflect the general preference for binding specificity within each class. Considering these binding motifs as a complete set, 72 MYB GCNs were enriched ( $P$  < 0.01) with at least one main MYB consensus group, with type IIG/AC-elements generally being the most over-represented (39 MYB enriched GCNs), followed by type I and type II (32 and 23 GCNs, respectively). As for MYB variants such as AGGTTA, this motif was also enriched in 15 MYB GCNs ( $P$  < 0.01). In several cases, more than one consensus was enriched (e.g. MBSI/MBSIIG or MBSIII) (Supplementary Fig. S10, Supplementary Table S11). This *in silico* analysis suggests that grapevine MYB TFs may bind to more than one MYB consensus CRE (up to all three and their variants) in the promoters of co-expressed genes. This has been demonstrated experimentally for many Arabidopsis R2R3-MYB proteins, which can bind a variety of MYB sites,<sup>45</sup> consistent with our observations.

Enrichment score plots show that some MYB GCNs are highly enriched ( $P$  < 0.001) with various MYB-binding CREs (Fig. 4C), suggesting a strong relationship between binding elements and certain biological roles. For example, the GCN surrounding grapevine homologues of *AtMYB46* and *AtMYB83* (e.g. *VviMYB185*; *VIT\_06s0004g02110*) were highly enriched with the CCAACC element ( $P$  < 1.50E-09) present in 64 co-expressed genes, many of which are involved in secondary cell wall biosynthesis (Supplementary Tables S5 and S8). Interestingly, the MYB GCN with the highest enrichment of the TGGTTR element (7.99E-04) was *VviMYB185*. Many of the TGGTTR-containing promoters also correspond to secondary cell wall-related genes. The prevalence of these two elements may provide dual regulation for the majority of the GCN. This result is consistent with the function of *AtMYB46* and *AtMYB83* (which directly bind type IIG/AC-elements) and in regulating an extensive suite of downstream target genes involved in secondary cell wall metabolism.<sup>76</sup> As for type I motifs such as CAGTTR, we observed a strong enrichment ( $P$  < 4.62E-05, 64 CAGTTR-containing co-expressed gene promoters) in the *VviMYB30A* GCN (64; Supplementary Table S8). Many CAGTTR-containing co-expressed gene promoters were highly enriched in lipid, sugar-derivative metabolism and disease-related signalling genes presumably involved in the hypersensitive cell death response. This is consistent with the role of MYB30 in Arabidopsis and its



**Figure 4.** Distribution profiles of various cognate R2R3-MYB TF binding sites in the promoter regions of co-expressed genes across the MYB GCN. The figure illustrates the (A) median number of co-expressed genes (*match\_in\_sample*) containing the respective R2R3 MYB binding sites, (B) number of MYB GCNs enriched (MNE) with the associated MYB binding sites at  $P < 0.01$  and (C) general distribution of enrichment scores expressed as  $[-\log_{10}(P \text{ value}) * 10]$ . Outlier dots in red/grey represent extreme enrichment score ( $>30$ ) for an associated MYB GCN. N = A/C/G/T, R = A/G, W = A/T, M = A/C.

ability to bind the promoters of lipid-related genes.<sup>82</sup> These observations suggest preferential binding of type I motifs (i.e. CAGTTR) by grapevine MYB30A and support its role in stress (i.e. osmotic) regulation in grapevine, as shown previously.<sup>12</sup> In this sense, VviMYB30A may respond to salt stress by regulating many of the genes identified in our co-expression analysis.

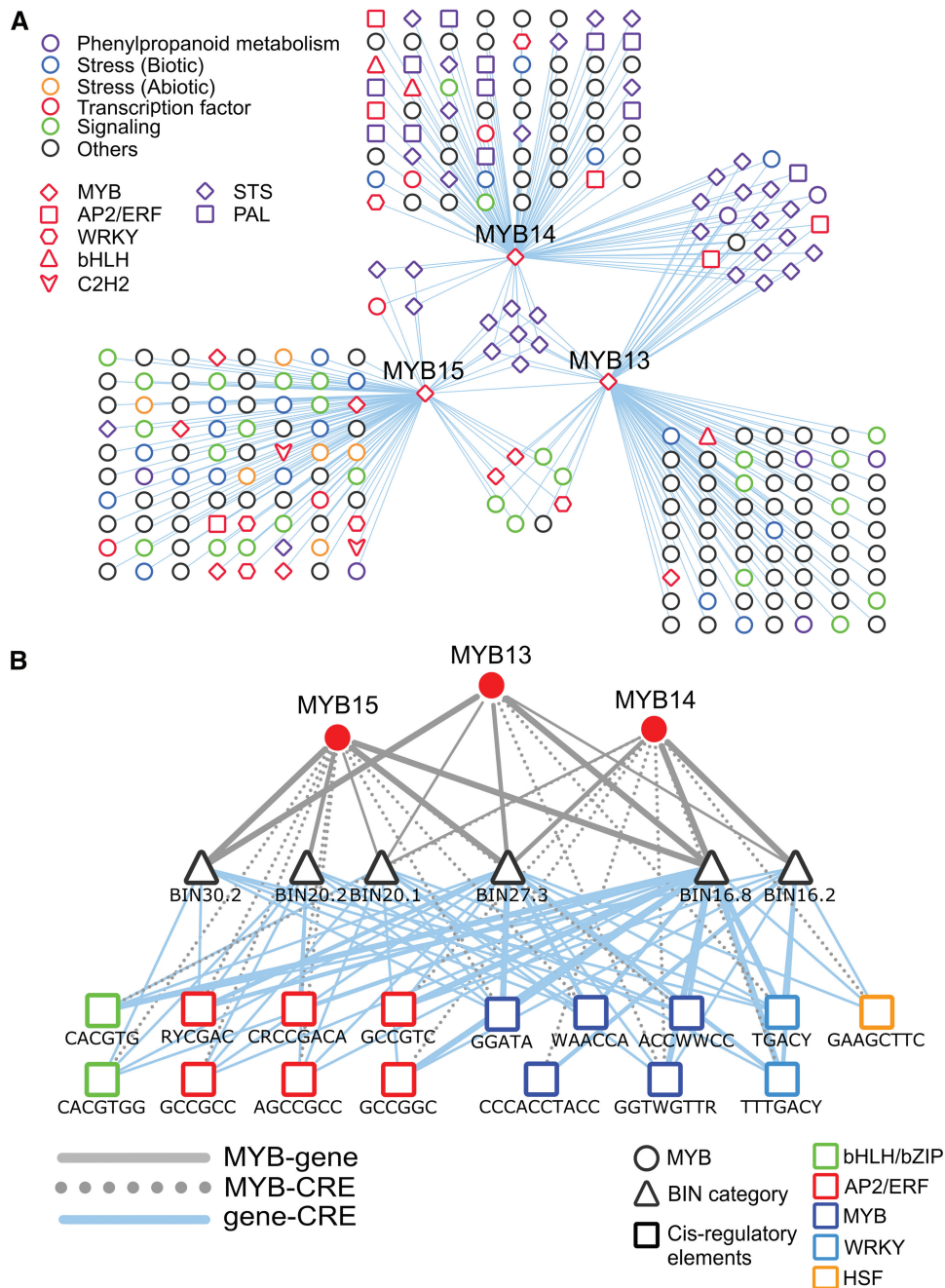
The cases described above are some examples of how analysis of promoter binding sites supports the functional implications of R2R3-MYB factors (e.g. secondary metabolism, stress). Our analysis also highlights that many of the co-expressed genes might actually be targets of MYB TFs within each GCN, and proposes binding preferences for each grape R2R3-MYB.

### 3.6. Co-expression and *cis*-regulatory sub-network analysis of Subgroup 2 reveals a new putative regulator of stilbene biosynthesis

By incorporating multiple levels of co-expressed genes, functional evidence and promoter information, we can robustly infer the complex

regulation exerted by known and novel grapevine R2R3-MYBs. This is the case with Subgroup S2 members VviMYB14 and VviMYB15. While S2 members in *Arabidopsis* have been linked to the control of inflorescence architecture, cold stress tolerance and the shikimate pathway,<sup>83,84</sup> their orthologues in the Vitaceae family form part of the stilbene-regulating clade. Stilbenes, which are absent in *Arabidopsis*, are specifically produced in grape organs during development (e.g. in roots,<sup>20,21</sup> and in berries,<sup>20,62</sup> and in response to abiotic and biotic stresses<sup>85–87</sup>). These phytoalexins have a defensive role and are produced by stilbene synthases (STS), the key structural enzymes that compete with chalcone synthases for a *p*-coumaroyl phenylpropanoid intermediate in the production of *cis-trans*-resveratrol. As recently demonstrated, the grapevine MYB14 and MYB15 TFs are involved in transcriptional regulation of *STILBENE SYNTHASES* *STS41* and *STS29*.<sup>20</sup>

To understand large-scale transcriptional control of stilbene-related MYBs, we combined the GCN of MYB14 (*VIT\_07s0005g03340*) and MYB15 (*VIT\_05s0049g01020*) with that of MYB13 (*VIT\_05s0049g01010*), the uncharacterized closest



**Figure 5.** Integrated gene co-expression and *cis*-regulatory network centred on the grapevine MYB13, 14, and 15 sub-networks (Subgroup 2). (A) Nodes in open forms represent genes and edges indicate significant co-expression between each MYB and its neighbours. The border of nodes represents the assigned MapMan function indicating phenylpropanoid metabolism, biotic stress, abiotic stress, transcriptional regulation and signalling. (B) The network depicts the connections between MYB13/14/15 (circles), the associated function of the co-expressed genes, collapsed into respective sub-BINs (triangles), connected by edges and the highly enriched ( $P$  value < 0.001) CRE within the promoter regions of co-expression genes (squares). Gene-CRE edges indicate the relative number of co-expressed genes containing the various enriched CRE sites within their promoter regions (thick, medium and thin edges indicate > 10, between 5 and 10, and < 4 co-expressed genes, respectively), while dashed lines indicate the various CRE-binding sites enriched within each MYB network. BIN20.1, biotic stress; BIN20.2, abiotic stress; BIN30.2, signalling by receptor kinases; BIN27.3, TF regulation; BIN16.8, flavonoid metabolism (STSs, etc.); BIN16.2, phenylpropanoid (PALs, etc.) metabolism.

homologue of *MYB15* (Fig. 5A). We observed that a core set of seven *STS* transcripts were co-expressed with all three MYBs while 21 *STS* genes were shared between MYB13 and MYB14 (Supplementary Table S5). Three *STS*s were co-expressed with MYB15 and MYB14 (including *STS41*) while 10 were exclusively co-expressed with MYB14. Eleven (out of 12) *PHENYLALANINE AMMONIA-*

*LYASE* (*PAL*) genes, constituting the first step of the phenylpropanoid pathway, were present in the MYB14 GCN (BIN 16.2.1,  $P < 3.87E-24$ ) while only one *PAL* gene was shared between MYB14 and MYB13. Moreover, transcripts involved in transcriptional regulation, such as a WRKY factors (*VIT\_01s0010g03930*) and seven serine/threonine protein kinases that are likely involved in signalling

(e.g. *VIT\_10s0071g00440*), were co-expressed with *MYB13* and *MYB15*. Genes from the *MYB15* GCN encode for many pathogenesis-related (4 thaumatin, 2 osmotin), nitrilase (11 transcripts, BIN26.8,  $P < 2.75E-11$ ), and receptor kinase (15 transcripts, BIN30.2,  $P < 9.96E-05$ ) transcripts. Inducible defence-related proteins such as thaumatin, osmotin and nitrilases are commonly elicited during pathogen infection and in situations of osmotic stress, cold, wounding and development.<sup>88</sup> These results suggest that S2 members are involved in defence responses not only by regulating *STS* transcripts but also by regulating other stress-related transcripts.

We used a comprehensive compendium of CRE (based in protein-binding microarrays, see Materials and methods) to analyse the promoter regions of *STS* transcripts that were co-expressed with Subgroup 2 members. This analysis showed that these *STS* promoters were highly over-represented ( $P < 0.001$ ) with CRE-regulated by the R2R3-MYBs (i.e. GGTWGTR, ACCWWCC), WRKY (i.e. TTTGACY) and AP2/ERF (i.e. RYCGAC, GCCGGC) TFs (Fig. 5B, Supplementary Table S8). CREs within co-expression modules (in combination with many other CREs) provide adequate fine control of genes during stress and development.<sup>25,26</sup> Thus, our analysis reinforces the notion that Subgroup 2 MYBs are likely to bind a larger set of *STS* promoter regions than previously shown.<sup>20</sup> These findings also highlight the fact that different upstream stress response pathways, related to jasmonic acid and ethylene signalling pathways through involvement of the WRKY (i.e. WRKY03, *VIT\_01s0010g03930*; WRKY24, *VIT\_08s0058g00690* and WRKY29, *VIT\_10s0116g01200*) and AP2/ERF (i.e. *VIT\_16s0013g00890*, *VIT\_18s0072g00260*) regulators, respectively, may be part of a larger grapevine *STS* regulatory network. We also observed significant enrichment of MYB and WRKY binding sites in the promoter regions of biotic stress-related (BIN 20.1) co-expressed genes. Moreover, MYB and AP2/ERF binding sites were enriched in the promoter regions of abiotic stress (BIN 20.2) and nitrilase (BIN26.8) transcripts that are unique to the *MYB15* GCN. Taken together, this supports the idea that many stress-related genes are also part of the *MYB13/14/15* regulatory network. While many co-expressed genes are shared between the *MYB13/14/15* GCNs, some WRKY (*VIT\_01s0010g03930*, *VIT\_10s0116g01200*) and AP2/ERF (*VIT\_16s0013g01120*) TFs may be regulated in a coordinated way in individual MYB networks. The major TF families involved in the transcriptional networks of plant immunity often comprise R2R3-MYB, WRKY and AP2/ERFs TFs, and their binding motifs are enriched in the promoters of co-expressed genes.<sup>25,89</sup> These observations may explain the enrichment of stress-related BIN terms, specific TFs and CRE and their combinatorial associations as reported here. In summary, our multifaceted analyses indicate that *MYB13* may have functional similarities with *MYB14/15* and that the complexity in *STS* regulation might require co-operation of other TF families.

### 3.7. A sub-specialized expression of *MYB13* correlates with the accumulation of stilbenes under non-stressed conditions

To understand the possible role of *MYB13* in regulating stilbene biosynthesis, we investigated its expression under different conditions, including vegetative organs, two berry developmental series and biotic/abiotic stress treatments (Fig. 6). Due to the high homology (80%) of the coding sequence of *MYB15* and *MYB13*, we designed gene-specific primers in the 3'UTR of both genes. We also measured stilbene levels to see if these were correlated with *MYB13* expression. *MYB13* was expressed in non-stressed leaves (young and old), roots

and stems of Chardonnay and Shiraz cultivars, consistent with the piceid accumulation found in these organs (Fig. 6A). The high stilbene content in the roots was comparable with strong expression of *MYB13*, *MYB14*, *MYB15* and several *STS*s (*STS25/27/29* and *STS41/45*). However, where there was almost no expression of *MYB14* and *MYB15*, we found high expression of *MYB13*, correlated with the detection of piceids, particularly in leaves and stems. These results suggest a role of *MYB13* in regulating stilbene biosynthesis in these organs under non-stressed conditions. They also suggest that the expression of *STS* genes other than *STS25/27/29* and *STS41/48* might be regulated in these organs by *MYB13*.

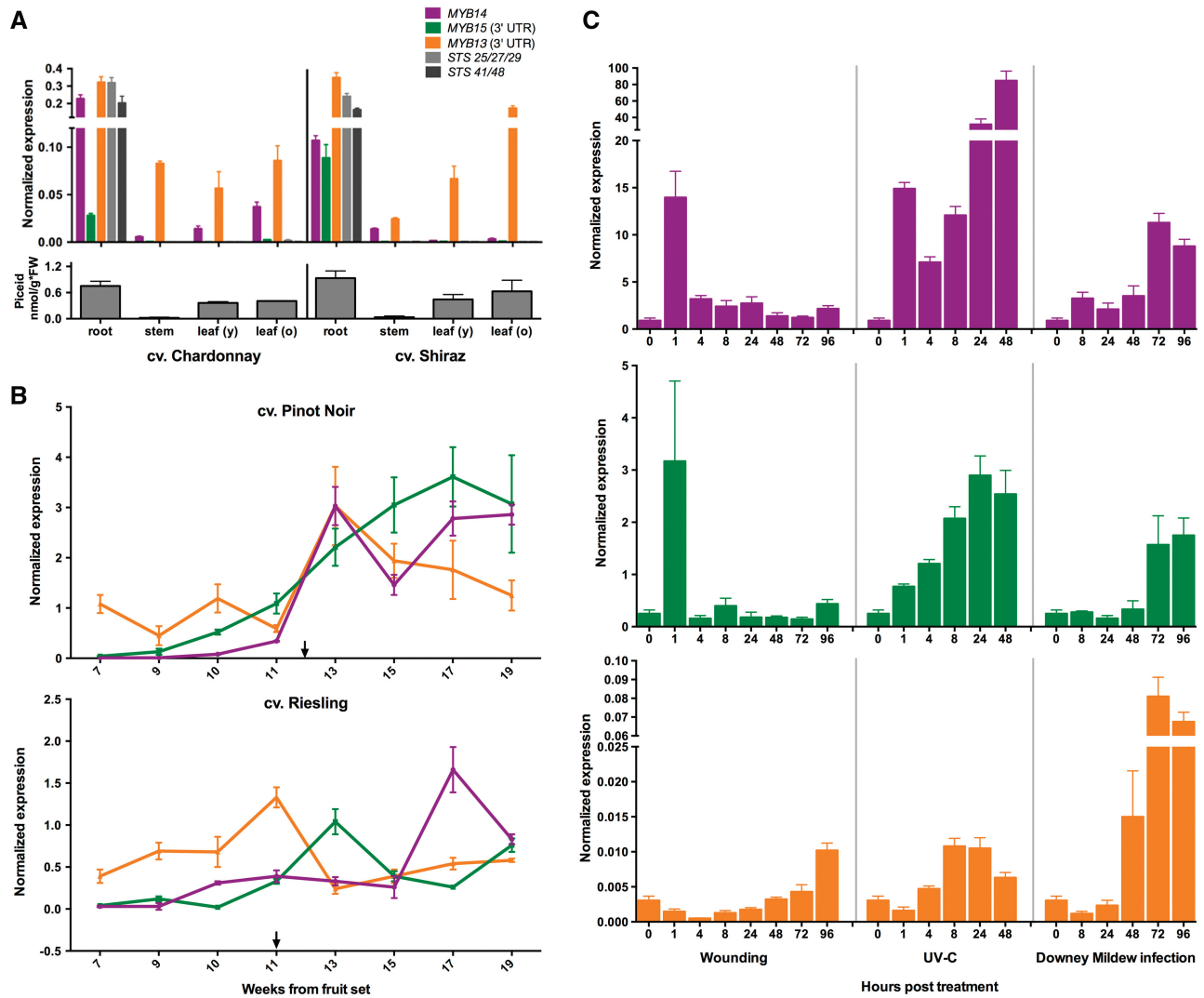
Analysis of *MYB13* transcript levels during development in the black-skinned 'Pinot Noir' berry and the white-skinned 'Riesling' berry showed that *MYB13* expression peaks around veraison. Expression of *MYB14* is only concomitant with *MYB13* in 'Pinot Noir', as in 'Riesling' both *MYB14* and *MYB15* peak at later ripening stages. This observation suggests that *MYB13* may have a critical role in regulating stilbenes at the onset of ripening in berries (preceding the expression of its paralogues in some cultivars). These profiles are consistent with previous data from 'Pinot Noir',<sup>20</sup> where *MYB14*, *MYB15* and *STS* (*STS25/27/29* and *STS41/48*) expression were correlated with the accumulation of glycosylated resveratrol (*trans*-piceid) during the late stages of berry ripening. Furthermore, stilbene accumulation has also been found to increase during early stages, when *MYB14* and *MYB15* transcripts are not detected. This suggests that another regulator affects stilbene biosynthesis, possibly *MYB13*.

The concentration of stilbenes produced in ripening berries is only 1–5% of those in berries or leaves subject to biotic and abiotic stresses.<sup>85–87</sup> Thus, we further tested the expression of *MYB13* in response to conditions that significantly increase *STS* expression, namely wounding, UV-C radiation and downy mildew infection,<sup>4</sup> and observed several differences between expression of the three genes. In general, *MYB13* showed much lower transcript accumulation under stress conditions than *MYB14* and *MYB15*. However, on examining fold induction between treated/untreated conditions, *MYB13* appears to have a particular stress-response with different expression kinetics. Upon wounding, *MYB13* expression peaked 96 hours post-treatment (hpt), in contrast to that of *MYB14* and *MYB15*, which had a subtler peak at 1 hpt. The behaviour of *MYB13* following UV-C treatment was more similar to that observed for *MYB14* and *MYB15*, but with lower intensity. Finally, downy mildew infection produced the most striking response, with the *MYB13* transcript beginning to accumulate at very early stages (48 hpt) and reaching much higher fold induction levels than *MYB14* and *MYB15*.

Compared to *MYB14* and *MYB15*, *MYB13* appears to be prominently expressed in roots and leaves, where stilbenes are accumulated. During fruit development *MYB13* seem to precede the expression of its paralogues, although this appears to be cultivar-specific. By including the stress-responsive experiments, we suggest an *MYB13* expression sub-specialization related to an *STS* basal expression in vegetative tissues but also to a specific induction to biotic stress. The Gene Network Analysis conducted here for the three Subgroup 2 MYBs will help to direct subsequent characterization approaches *in planta* to demonstrate *MYB13*'s ability to induce stilbene accumulation through the activation of *STS* genes.

## 4. Concluding remarks

Gene regulatory networks, controlled in part by TFs, have essential roles in all processes of life, from metabolism to cell cycle control,



**Figure 6.** qPCR expression profiles of *MYB14*, *MYB15* and *MYB13* in (A) grapevine vegetative organs, (B) berry developmental stages and (C) in response to abiotic and biotic stress. (A) Normalized expression in 'Shiraz' and 'Chardonnay' cultivars, as mean values of three replicates. Accumulation of *trans*-piceid in the same sampled organs was measured as methanol extracts using reverse-phase HPLC analysis. Columns represent *trans*-piceid in nmol/g\*FW with error bars indicating SEM from three biological replicates and two independent experiments. Data points in (B) are given as weeks during berry ripening (veraison; labelled with an arrow) starting at week seven.

differentiation and response to the environment. A clearer understanding of the roles of TFs may allow us to predict and alter the behaviour of gene networks in important processes such as disease resistance, or to use them for crop trait improvement programs.

Following the sequencing of the *V. vinifera* genome, progress in transcriptomic data acquisition has facilitated the collection of large-scale data with enormous potential for gene discovery. In this context, our study aimed to find the biological processes regulated by grapevine R2R3-MYB TFs by combining gene expression data, phylogenetic information, DNA-binding motif discovery and experimental evidence. Our results suggest that these regulators are involved in a plethora of processes, possibly in coordination with other specific TFs. We observed an over-representation of genes co-expressed with MYBs that were related to transcriptional regulation, cell wall biogenesis and secondary metabolism, the latter playing a significant role in fruit quality, health-promotion and plant disease prevention. Our study also provides evidence on the evolution of this family through tandem duplications, probably retained throughout

domestication, that have contributed significantly to the expansion of genes linked to these quality traits. One of the main outcomes of this study is the large collection of candidate genes and biological processes to be tested in future gene characterizations. Validation of these candidates will require studying transcriptional control exerted by R2R3-MYBs using a combination of chromatin immunoprecipitation techniques and expression profiling experiments (e.g. by ChIP-Seq). Further DNA-binding experiments will benefit from this study as discovery of over-represented motifs in the genomic regions of co-expressed genes may direct protein-DNA interactions assays (e.g. by electrophoretic mobility shift assays).

## Acknowledgements

We would specially like to thank Dr Ian B. Dry (CSIRO Agriculture, SA) for access to grapevine material treated under stress conditions and to Dr Jerome Grimplet (Research Centre of Vine and Wine-related Science, ICVV) for his guidance in gene nomenclature and annotation in ORCAE.

## Conflict of interest

None declared.

## Supplementary data

Supplementary data are available at [www.dnaresearch.oxfordjournals.org](http://www.dnaresearch.oxfordjournals.org).

## Funding

This study was partially funded by the University of British Columbia (10R18459) and Genome British Columbia (10R21188).

## References

1. This, P., Lacombe, T. and Thomas, M.R. 2006, Historical origins and genetic diversity of wine grapes, *Trends Genet.*, **22**, 511–9.
2. Jaillon, O., Aury, J.-M., Noel, B., et al. 2007, The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla, *Nature*, **449**, 463–7.
3. Parage, C., Tavares, R., Réty, S., et al. 2012, Structural, functional, and evolutionary analysis of the unusually large stilbene synthase gene family in grapevine, *Plant Physiol.*, **160**, 1407–19.
4. Vannozzi, A., Dry, I.B., Fasoli, M., Zenoni, S. and Lucchin, M. 2012, Genome-wide analysis of the grapevine stilbene synthase multigenic family: genomic organization and expression profiles upon biotic and abiotic stresses, *BMC Plant Biol.*, **12**, 130.
5. Matus, J.T., Aquea, F. and Arce-Johnson, P. 2008, Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes, *BMC Plant Biol.*, **8**, 83.
6. Du, H., Liang, Z., Zhao, S., et al. 2015, The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: new insights into subfamily classification and expansion, *Sci. Rep.*, **5**, 11037.
7. Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C. and Lepiniec, L. 2010, MYB transcription factors in *Arabidopsis*, *Trends Plant Sci.*, **15**, 573–81.
8. Rabinowicz, P.D., Braun, E.L., Wolfe, A.D., Bowen, B. and Grotewold, E. 1999, Maize R2R3 Myb genes: sequence analysis reveals amplification in the higher plants, *Genetics*, **153**, 427–44.
9. Dias, A.P., Braun, E.L., McMullen, M.D. and Grotewold, E. 2003, Recently duplicated maize R2R3 Myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication, *Plant Physiol.*, **131**, 610–20.
10. Yasui Y., Hirakawa H., Ueno M., et al. 2016, Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes, *DNA Res.*, doi:10.1093/dnares/dsw012.
11. Chen, B., Niu, F., Liu, W.-Z., et al. 2016, Identification, cloning and characterization of R2R3-MYB gene family in canola (*Brassica napus* L.) identify a novel member modulating ROS accumulation and hypersensitive-like cell death, *DNA Res.*, **23**, 101–14.
12. Galbiati, M., Matus, J.T., Francia, P., et al. 2011, The grapevine guard cell-related VvMYB60 transcription factor is involved in the regulation of stomatal activity and is differentially expressed in response to ABA and osmotic stress, *BMC Plant Biol.*, **11**, 142.
13. Zheng, H., Yu, X., Yuan, Y., et al. 2016, The VviMYB80 gene is abnormally expressed in *Vitis vinifera* L. cv. ‘Zhong Shan Hong’ and its expression in tobacco driven by the 35S promoter causes male sterility, *Plant Cell Physiol.*, **57**, 540–57.
14. Kobayashi, S., Ishimaru, M., Hiraoka, K. and Honda, C. 2002, Myb-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis, *Planta*, **215**, 924–33.
15. Walker, A.R., Lee, E., Bogs, J., McDavid, D.A., Thomas, M.R. and Robinson, S.P. 2007, White grapes arose through the mutation of two similar and adjacent regulatory genes, *Plant J.*, **49**, 772–85.
16. Deluc, L., Barrieu, F.C., Marchive, C., et al. 2006, Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway, *Plant Physiol.*, **140**, 499.
17. Deluc, L., Bogs, J., Walker, A.R., et al. 2008, The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries, *Plant Physiol.*, **147**, 2041–53.
18. Cavallini, E., Matus, J.T., Finezzo, L., et al. 2015, The phenylpropanoid pathway is controlled at different branches by a set of R2R3-MYB C2 repressors in grapevine, *Plant Physiol.*, **167**, 1448–70.
19. Bogs, J., Jaffé, F.W., Takos, A.M., Walker, A.R. and Robinson, S.P. 2007, The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development, *Plant Physiol.*, **143**, 1347–61.
20. Höll, J., Vannozzi, A., Czemplak, S., et al. 2013, The R2R3-MYB transcription factors MYB14 and MYB15 regulate stilbene biosynthesis in *Vitis vinifera*, *Plant Cell*, **25**, 4135–49.
21. Huang, Y.F., Vialet, S., Guiraud, J.L., et al. 2014, A negative MYB regulator of proanthocyanidin accumulation, identified through expression quantitative locus mapping in the grape berry, *New Phytol.*, **201**, 795–809.
22. Lee, T., Kim, H. and Lee, I. 2015, Network-assisted crop systems genetics: network inference and integrative analysis, *Curr. Opin. Plant Biol.*, **24**, 61–70.
23. Obayashi, T., Okamura, Y., Ito, S., et al. 2014, ATTED-II in 2014: evaluation of gene coexpression in agriculturally important plants, *Plant Cell Physiol.*, **55**, e6.
24. Mutwil, M., Klie, S., Tohge, T., et al. 2011, PlaNet: combined sequence and expression comparisons across plant networks derived from seven species, *Plant Cell*, **23**, 895–910.
25. Ma, S., Shah, S., Bohnert, H.J., Snyder, M. and Dinesh-Kumar, S.P. 2013, Incorporating motif analysis into gene co-expression networks reveals novel modular expression pattern and new signaling pathways, *PLoS Genet.*, **9**, e1003840.
26. Zou, C., Sun, K., Mackaluso, J.D., et al. 2011, Cis-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 14992–7.
27. Wilkins, O., Nahal, H., Foong, J., Provart, N.J. and Campbell, M.M. 2009, Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors, *Plant Physiol.*, **149**, 981–93.
28. Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., et al. 2009, CDD: specific functional annotation with the Conserved Domain Database, *Nucleic Acids Res.*, **37**, 205–10.
29. Eddy, S.R. 2008, A probabilistic model of local sequence alignment that simplifies statistical significance estimation, *PLoS Comput. Biol.*, **4**, e1000069.
30. Finn, R.D., Bateman, A., Clements, J., et al. 2014, Pfam: the protein families database, *Nucleic Acids Res.*, **42**, 222–30.
31. Zenoni, S., Ferrarini, A., Giacomelli, E., et al. 2010, Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq, *Plant Physiol.*, **152**, 1787–95.
32. Venturini, L., Ferrarini, A., Zenoni, S., et al. 2013, *De novo* transcriptome characterization of *Vitis vinifera* cv. Corvina unveils varietal diversity, *BMC Genomics*, **14**, 41.
33. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol. Biol. Evol.*, **28**, 2731–9.
34. Wang, Y., Tang, H., DeBarry, J.D., et al. 2012, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, *Nucleic Acids Res.*, **40**, e49.
35. Camacho, C., Coulouris, G., Avagyan, V., et al. 2009, BLAST plus: architecture and applications, *BMC Bioinformatics*, **10**, 421.
36. Barrett, T., Wilhite, S.E., Ledoux, P., et al. 2013, NCBI GEO: archive for functional genomics data sets – update, *Nucleic Acids Res.*, **41**, D991–5.
37. Carvalho, B.S. and Irizarry, R.A. 2010, A framework for oligonucleotide microarray preprocessing, *Bioinformatics*, **26**, 2363–7.
38. Shannon, P., Markiel, A., Ozier, O., et al. 2003, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.*, **13**, 2498–504.

39. Lohse, M., Nagel, A., Herter, T., et al. 2014, Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data, *Plant Cell Environ.*, **37**, 1250–8.
40. Reimand, J., Kull, M., Peterson, H., Hansen, J. and Vilo, J. 2007, g:Profiler – a web-based toolset for functional profiling of gene lists from large-scale experiments, *Nucleic Acids Res.*, **35**, 193–200.
41. Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P. and Solano, R. 2014, DNA-binding specificities of plant transcription factors and their potential to define target genes, *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 2367–72.
42. Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T. 1999, Plant cis-acting regulatory DNA elements (PLACE) database: 1999, *Nucleic Acids Res.*, **27**, 297–300.
43. Prouse, M.B. and Campbell, M.M. 2012, The interaction between MYB proteins and their target DNA binding sites, *Biochim. Biophys. Acta.*, **1819**, 67–77.
44. Romero, I., Fuertes, A., Benito, M.J., Malpica, J.M., Leyva, A. and Paz-Ares, J. 1998, More than 80R2R3-MYB regulatory genes in the genome of *Arabidopsis thaliana*, *Plant J.*, **14**, 273–84.
45. Kelemen, Z., Sebastian, A., Xu, W., et al. 2015, Analysis of the DNA-binding activities of the Arabidopsis R2R3-MYB transcription factor family by one-hybrid experiments in Yeast, *PLoS One*, **10**, e0141044.
46. Pfaffl, M.W., Horgan, G.W. and Dempfle, L. 2002, Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR, *Nucleic Acids Res.*, **30**, e36.
47. Soler, M., Camargo, E.L.O., Carocha, V., et al. 2015, The Eucalyptus grandis R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function, *New Phytol.*, **206**, 1364–77.
48. Matus, J.T., Aquea, F., Espinoza, C., et al. 2014, Inspection of the grapevine BURP superfamily highlights an expansion of RD22 genes with distinctive expression features in berry development and ABA-mediated stress responses, *PLoS One*, **9**, e110372.
49. Grimplet, J., Martínez-Zapater, J.M. and Carmona, M.J. 2016, Structural and functional annotation of the MADS-box transcription factor family in grapevine, *BMC Genomics*, **17**, 80.
50. Grimplet, J., Adam-Blondon, A.-F., Bert, P.-F., et al. 2014, The grapevine gene nomenclature system, *BMC Genomics*, **15**, 1077.
51. Stracke, R., Werber, M. and Weisshaar, B. 2001, The R2R3-MYB gene family in *Arabidopsis thaliana*, *Curr. Opin. Plant Biol.*, **4**, 447–56.
52. Barah, P., Winge, P., Kusnierczyk, A., Tran, D.H. and Bones, A.M. 2013, Molecular signatures in *Arabidopsis thaliana* in response to insect attack and bacterial infection, *PLoS One*, **8**, e58987.
53. Li, S.F., Milliken, O.N., Pham, H., et al. 2009, The Arabidopsis MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis, *Plant Cell*, **21**, 72–89.
54. Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J. and Koes, R. 2006, PH4 of Petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway, *Plant Cell*, **18**, 1274–91.
55. Freeling, M. 2009, Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition, *Annu. Rev. Plant Biol.*, **60**, 433–53.
56. Benjak, A., Forneck, A. and Casacuberta, J.M. 2008, Genome-wide analysis of the ‘cut-and-paste’ transposons of grapevine, *PLoS One*, **3**, e3107.
57. Albert, N.W. 2015, Subspecialization of R2R3-MYB repressors for anthocyanin and proanthocyanidin regulation in forage legumes, *Front. Plant Sci.*, **6**, 1–13.
58. Potenza, E., Racchi, M.L., Sterck, L., et al. 2015, Exploration of alternative splicing events in ten different grapevine cultivars, *BMC Genomics*, **16**, 706.
59. Filichkin, S.A., Priest, H.D., Givan, S.A., et al. 2010, Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*, *Genome Res.*, **20**, 45–58.
60. Li, J., Li, X., Guo, L., et al. 2006, A subgroup of MYB transcription factor genes undergoes highly conserved alternative splicing in Arabidopsis and rice, *J. Exp. Bot.*, **57**, 1263–73.
61. Staiger, D. and Brown, J.W.S. 2013, Alternative splicing at the intersection of biological timing, development, and stress responses, *Plant Cell*, **25**, 3640–56.
62. Carbonell-Bejerano, P., Diago, M.-P., Martínez-Abaigar, J., Martínez-Zapater, J.M., Tardaguila, J. and Núñez-Olivera, E. 2014, Solar ultraviolet radiation is necessary to enhance grapevine fruit ripening transcriptional and phenolic responses, *BMC Plant Biol.*, **14**, 183.
63. Costantini, L., Malacarne, G., Lorenzi, S., et al. 2015, New candidate genes for the fine regulation of the colour of grapes, *J. Exp. Bot.*, **66**, 4427–40.
64. Sweetman, C., Sadras, V.O., Hancock, R.D., Soole, K.L. and Ford, C.M. 2014, Metabolic effects of elevated temperature on organic acid degradation in ripening *Vitis vinifera* fruit, *J. Exp. Bot.*, **65**, 5975–88.
65. Wong, D.C.J., Sweetman, C., Drew, D.P. and Ford, C.M. 2013, VTCdb: a gene co-expression database for the crop species *Vitis vinifera* (grapevine), *BMC Genomics*, **14**, 882.
66. Palumbo, M.C., Zenoni, S., Fasoli, M., et al. 2014, Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development, *Plant Cell*, **26**, 4617–35.
67. Obayashi, T. and Kinoshita, K. 2009, Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression, *DNA Res.*, **16**, 249–60.
68. Wong, D.C.J., Sweetman, C. and Ford, C.M. 2014, Annotation of gene function in citrus using gene expression information and co-expression networks, *BMC Plant Biol.*, **14**, 186.
69. Duarte, J.M., Cui, L., Wall, P.K., et al. 2006, Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis, *Mol. Biol. Evol.*, **23**, 469–78.
70. Thimm, O., Bläsing, O., Gibon, Y., et al. 2004, MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes, *Plant J.*, **37**, 914–39.
71. Conn, S., Curtin, C., Bézier, A., Franco, C. and Zhang, W. 2008, Purification, molecular cloning, and characterization of glutathione S-transferases (GSTs) from pigmented *Vitis vinifera* L. cell suspension cultures as putative anthocyanin transport proteins, *J. Exp. Bot.*, **59**, 3621–34.
72. Huguency, P., Provenzano, S., Verries, C., et al. 2009, A novel cation-dependent O-methyltransferase involved in anthocyanin methylation in grapevine, *Plant Physiol.*, **150**, 2057–70.
73. Koyama, K., Numata, M., Nakajima, I., Goto-Yamamoto, N., Matsumura, H. and Tanaka, N. 2014, Functional characterization of a new grapevine MYB transcription factor and regulation of proanthocyanidin biosynthesis in grapes, *J. Exp. Bot.*, **65**, 4433–49.
74. Terrier, N., Torregrosa, L., Ageorges, A., et al. 2009, Ectopic expression of VvMybPA2 promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway, *Plant Physiol.*, **149**, 1028–41.
75. Zhong, R., Lee, C., Zhou, J., McCarthy, R.L. and Ye, Z.-H. 2008, A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis, *Plant Cell*, **20**, 2763–82.
76. Zhong, R. and Ye, Z.-H. 2012, MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes, *Plant Cell Physiol.*, **53**, 368–80.
77. Oshima, Y., Shikata, M., Koyama, T., Ohtsubo, N., Mitsuda, N. and Ohme-Takagi, M. 2013, MIXTA-like transcription factors and WAX INDUCER1/SHINE1 coordinately regulate cuticle development in Arabidopsis and *Torenia fournieri*, *Plant Cell*, **25**, 1609–24.
78. Pireyre, M. and Burrow, M. 2015, Regulation of MYB and bHLH transcription factors: a glance at the protein level, *Mol. Plant*, **8**, 378–88.
79. Qi, T., Huang, H., Wu, D., et al. 2014, Arabidopsis DELLA and JAZ proteins bind the WD-repeat/bHLH/MYB complex to modulate gibberellin and jasmonate signaling synergy, *Plant Cell*, **26**, 1118–33.

80. Qi, T., Huang, H., Song, S. and Xie, D. 2015, Regulation of jasmonate-mediated stamen development and seed production by a bHLH-MYB complex in *Arabidopsis*, *Plant Cell*, **27**, 1620–33.
81. Zeng, J., Li, X., Xu, Q., et al. 2015, EjAP2-1, an AP2/ERF gene, is a novel regulator of fruit lignification induced by chilling injury, via interaction with EjMYB transcription factors, *Plant Biotechnol. J.*, **13**, 1325–34.
82. Raffaele, S., Vaillau, F., Léger, A., et al. 2008, A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in *Arabidopsis*, *Plant Cell*, **20**, 752–67.
83. Agarwal, M., Hao, Y., Kapoor, A., et al. 2006, A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance, *J. Biol. Chem.*, **281**, 37636–45.
84. Chen, Y., Chen, Z., Kang, J., Kang, D., Gu, H. and Qin, G. 2013, AtMYB14 regulates cold tolerance in *Arabidopsis*, *Plant Mol. Biol. Report.*, **31**, 87–97.
85. Bais, A.J., Murphy, P.J. and Dry, I.B. 2000, The molecular regulation of stilbene phytoalexin biosynthesis in *Vitis vinifera* during grape berry development, *Aust. J. Plant Physiol.*, **27**, 425–33.
86. Douillet-Breuil, A.C., Jeandet, P., Adrian, M. and Bessis, R. 1999, Changes in the phytoalexin content of various *Vitis* spp. in response to ultraviolet C elicitation, *J. Agric. Food Chem.*, **47**, 4456–61.
87. Langcake, P. 1981, Disease resistance of *Vitis* spp. and the production of the stress metabolites resveratrol,  $\epsilon$ -viniferin,  $\alpha$ -viniferin and pterostilbene, *Physiol. Plant Pathol.*, **18**, 213–26.
88. Van Loon, L.C., Rep, M. and Pieterse, C.M.J. 2006, Significance of inducible defense-related proteins in infected plants, *Annu. Rev. Phytopathol.*, **44**, 135–62.
89. Tsuda, K. and Somssich, I.E. 2015, Transcriptional networks in plant immunity, *New Phytol.*, **206**, 932–47.