




Direct regulation of shikimate, early phenylpropanoid, and stilbenoid pathways by Subgroup 2 R2R3-MYBs in grapevine

Luis Orduña¹ , Miaomiao Li², David Navarro-Payá¹, Chen Zhang¹, Antonio Santiago¹, Pablo Romero¹, Živa Ramsak³, Gabriele Magon⁴, Janine Höll⁵, Patrick Merz⁵, Kristina Gruden³, Alessandro Vannozzi⁴, Dario Cantu⁶, Jochen Bogs⁵, Darren C. J. Wong⁷, Shao-shan Carol Huang²  and José Tomás Matus^{1*} 

¹Institute for Integrative Systems Biology (I2SysBio), Universitat de València-CSIC, Paterna, 46908, Valencia, Spain,

²Center for Genomics and Systems Biology, Department of Biology, New York University, USA,

³Department of Biotechnology and Systems Biology, National Institute of Biology, Večna pot 111, 1000, Ljubljana, Slovenia,

⁴Department of Agronomy, Food, Natural resources, Animals, and Environment (DAFNAE), University of Padova, Legnaro 35020, Italy,

⁵Dienstleistungszentrum Ländlicher Raum Rheinpfalz, Viticulture and Enology Group, Neustadt/W, Germany,

⁶Department of Viticulture and Enology, University of California Davis, Davis, California, USA, and

⁷Ecology and Evolution, Research School of Biology, The Australian National University, Acton, Australia

Received 23 September 2021; revised 17 January 2022; accepted 19 January 2022; published online 29 January 2022.

*For correspondence (e-mail tomas.matus@gmail.com, tomas.matus@uv.es).

SUMMARY

The stilbenoid pathway is responsible for the production of resveratrol in grapevine (*Vitis vinifera* L.). A few transcription factors (TFs) have been identified as regulators of this pathway but the extent of this control has not been deeply studied. Here we show how DNA affinity purification sequencing (DAP-Seq) allows for the genome-wide TF-binding site interrogation in grape. We obtained 5190 and 4443 binding events assigned to 4041 and 3626 genes for MYB14 and MYB15, respectively (approximately 40% of peaks located within –10 kb of transcription start sites). DAP-Seq of MYB14/MYB15 was combined with aggregate gene co-expression networks (GCNs) built from more than 1400 transcriptomic datasets from leaves, fruits, and flowers to narrow down bound genes to a set of high confidence targets. The analysis of MYB14, MYB15, and MYB13, a third uncharacterized member of Subgroup 2 (S2), showed that in addition to the few previously known stilbene synthase (*STS*) targets, these regulators bind to 30 of 47 *STS* family genes. Moreover, all three MYBs bind to several *PAL*, *C4H*, and *4CL* genes, in addition to shikimate pathway genes, the *WRKY03* stilbenoid co-regulator and resveratrol-modifying gene candidates among which *ROMT2-3* were validated enzymatically. A high proportion of DAP-Seq bound genes were induced in the activated transcriptomes of transient *MYB15*-overexpressing grapevine leaves, validating our methodological approach for delimiting TF targets. Overall, Subgroup 2 R2R3-MYBs appear to play a key role in binding and directly regulating several primary and secondary metabolic steps leading to an increased flux towards stilbenoid production. The integration of DAP-Seq and reciprocal GCNs offers a rapid framework for gene function characterization using genome-wide approaches in the context of non-model plant species and stands up as a valid first approach for identifying gene regulatory networks of specialized metabolism.

Keywords: secondary metabolism, regulatory networks, transcription factors, transcriptional regulation, DNA affinity purification sequencing.

INTRODUCTION

The evolved complexity of plant specialized metabolism can be traced back to the colonization of dry land by green algal-derived ancestors (Waters, 2003). Since their origin, land plants have been accompanied by a persistent range of abiotic and biotic stresses such as ultraviolet (UV)

radiation, desiccation, or unfavourable microbial communities that have led to the selective emergence of novel protective metabolites derived from primary metabolism (Kenrick & Crane, 1997). A clear example of this is the phenylpropanoid pathway (PPP), a major source of aromatic secondary metabolites in plants, with many of their

roles related to tolerance and adaptation to the above-mentioned stresses (Davies et al., 2020). While some branches of the pathway are ubiquitous in the plant kingdom (such as those producing flavonoids), others such as the stilbene pathway are restricted to a small number of species across at least 10 unrelated families, including Vitaceae (e.g. *Vitis vinifera* L.) and Moraceae (e.g. *Morus alba*) (Dubrovina & Kiselev, 2017). Grapevine is not only an important crop species but an interesting model to study the complexity of the stilbene pathway given the remarkable expansion of the stilbene synthase (*STS*) family in its genome through segmental and tandem gene duplications, reaching up to, in total, 47 genes (Parage et al., 2012; Vannozzi et al., 2012), albeit 12 of these are considered pseudogenes.

Stilbenes are phytoalexins, which are small, lipophilic compounds with key roles in plant defence, and they accumulate in response to a range of abiotic and biotic stresses. In recent years, the grapevine *STS* gene family has been studied to reveal a high degree of responsiveness and effectiveness against different biotic or abiotic stresses. For instance, ectopic expression of *VqSTS36* from the Chinese wild species *Vitis quinquangularis*, in both *Arabidopsis* and tomato, enhanced resistance to powdery mildew and osmotic stress (Huang et al., 2018), while the expression of *VqSTS29* in *Arabidopsis* also led to powdery mildew resistance (Xu et al., 2019). Moreover, the induction of *STS* gene expression in different *V. vinifera* tissues has been observed in response to fungal infection, UVC, or heat treatments (Lecourieux et al., 2017; Vannozzi et al., 2012; Yin et al., 2016).

STS enzymes are direct competitors of chalcone synthases for pathway precursors. Both proteins are closely related type III polyketide synthases, generating a tetraketide intermediate from the condensation of p-coumaroyl-coenzyme A with three molecules of malonyl-coenzyme A, which depending on *STS*/chalcone synthase activity, will generate resveratrol or naringenin chalcone, respectively, thus defining the entry point of the stilbene and flavonoid branches.

The main and first stilbene produced in grape tissues is resveratrol, a well-known nutraceutical with many characterized properties ranging from antioxidant to antiviral activities (e.g. it has been recently shown to inhibit SARS-CoV-2 *in vitro* replication in human lung cells; Pasquereau et al., 2021). Resveratrol is derivatized into a broad range of stilbenes such as pterostilbene, viniferins, piceid, and piceatannol, which involve methoxylation, oligomerization, glucosylation, and hydroxylation, respectively. The enzymes catalysing these reactions are mostly uncharacterized in grapevine except for a resveratrol *O*-methyltransferase (ROMT1) responsible for the production of pterostilbene (Schmidlin et al., 2008) and a resveratrol glucosyl transferase in *Vitis labrusca*, leading to the

production of piceid (Hall & De Luca, 2007). Hydroxylation of resveratrol into piceatannol could be carried out by cytochrome P450 oxidoreductases but no candidates have yet been identified in grape.

The stilbene pathway in grapevine is mainly regulated at the transcriptional level through transcription factors (TFs). In particular, the R2R3-type MYB14 and MYB15 [members from Subgroup 2 (S2)] have been shown to activate a few *STS* promoters specifically (*STS29* and *STS41*) in transient reporter assays (Höll et al., 2013), but modulation of other stilbenoid branch enzymes remains unexplored. Furthermore, gene co-expression networks (GCNs) and further correlation with stilbene accumulation have pointed out MYB13 as an additional putative regulator of stilbene accumulation (Wong et al., 2016) although it has not yet been validated *in planta*. We have initially explored TF regulatory networks interrogated by the use of GCNs, leading to the identification of members of the *AP2/ERF*, *bZIP*, and *WRKY* gene families as potential regulators of *STS* expression (Wong & Matus, 2017); however, experimental evidence of binding is necessary to prove regulatory causality. Nevertheless, systems biology approaches initially conducted in Wong et al. (2016), applied to the regulation of transcription in grapevine, have paved the way for the functional characterization of additional stilbene pathway regulators such as bZIP1, ERF114, MYB35A, and WRKY53 in recent years (Vannozzi et al., 2018; Wang et al., 2019; Wang & Wang, 2019), proving their efficacy in hypothesis-driven research for TF discovery. Interestingly, one of the newly identified stilbene regulators, WRKY53, binds to a subset of *STS* genes (*STS32* and *STS41*) and it is thought to form a regulatory complex with MYB14 and MYB15, probably increasing its activity (Wang et al., 2020). Moreover, WRKY03 has also been shown to work in synergy with MYB14 in the upregulation of *STS29* expression (Vannozzi et al., 2018).

Among the identified *STS* regulators, MYB14 and MYB15 have been proposed as upstream TFs in the regulatory cascade governing stilbenoid accumulation mainly due to their rapid activation response. However, this high hierarchy has no experimental validation to date. *MYB14* and *MYB15* have been shown to be differentially expressed under biotic and abiotic stresses in different *Vitis* species and cultivars with varying stilbene accumulation levels. For instance, the higher stilbene content of *V. labrusca* cv. 'Concord', in response to UVC and Al³⁺, compared with *V. vinifera* cv. 'Cabernet Sauvignon' has been directly linked to a greater *MYB14* promoter activity (Bai et al., 2019). The detailed characterization of these two potential regulators is hence of great importance. In addition, no other processes controlled by these TFs have been identified. This study combines genome-wide TF binding-site interrogation using DNA affinity purification sequencing (DAP-Seq), and aggregate whole genome co-expression networks to

lay out MYB14 and MYB15 cistrome landscapes and identify their complete repertoire of target genes. We have also analysed the yet uncharacterized MYB13. Our results suggest that these three MYBs bind to regulatory elements in most members of the *STS* gene family, other shikimate and early phenylpropanoid genes as well as bind to *WRKY* regulators of stilbene synthesis, representing major regulators of this specialized metabolic pathway.

RESULTS

MYB14 and MYB15 bind proximal upstream regions of a large set of *STS* genes

Our DAP-Seq analysis reported 5190 and 4443 TF-binding events, i.e. peaks (Figure 1a and Dataset S1), which were assigned to 4041 and 3626 different genes for MYB14 and MYB15, respectively. An initial inspection of all binding events showed that 73% and 75% of the peaks are present between 10 kb upstream of transcription start sites (TSSs) and 2 kb downstream of annotated gene ends for MYB14 and MYB15, respectively. In total, 30% and 32% of MYB14 and MYB15 peaks, respectively, are found within 5 kb upstream of TSSs. MYB14 and MYB15 shared, in total, 2709 bound genes as well as an almost identical DNA-binding motif obtained from the enrichment analysis of the top 600 most significant peaks sequences (Figure 1b). The cistromes of MYB14 and MYB15 Arabidopsis orthologues have not been previously studied.

A closer look at the *STS* gene family revealed that 22 of the 47 family members have MYB14/MYB15 DAP-Seq peaks associated to them (including five of 13 pseudogenes). Examining the *STS* genomic regions revealed clear DNA-binding signals upstream of TSSs for both MYB14 and MYB15 TFs, which was not observed in the pIX-HALO non-specific DNA binding control (Figure 2). A selection of housekeeping genes was used as a negative control with no specific DNA binding of MYB14 or MYB15 TFs around their TSSs. Interestingly, many DNA bound genes belong to the shikimate pathway or even to the first committed steps of the PPP. In addition to *STS*, other enzyme categories such as *O*-methyltransferases and glucosyltransferases are also represented among bound genes.

Inspection of MYB14 and MYB15 gene-centred co-expression networks

We generated condition-dependent leaf and fruit aggregate whole genome co-expression networks using public data uploaded in Sequence Read Archive (SRA), composed of >1400 runs belonging to >70 independent experiments (Dataset S2). Network performance was assessed through the area under the receiver operator characteristic curve (AUROC) measurement of the final aggregate network, obtaining AUROC values of approximately 0.75. The final fruit/flower network consisted of 31 723 genes as

individual nodes and 13 323 660 co-expression connections as directed edges (weighted by frequency of co-expression across the selected SRA studies). The leaf network consisted of 31 759 genes and 13 338 780 edges. The two networks share, in total, 1 795 970 common interactions, showing a great number of tissue-specific co-expression connections. GCNs centred on individual genes were extracted from the final aggregate leaf and fruit/flower networks in the form of the top 420 interactions for each particular gene (corresponding to the top 1% of all grapevine genes according to the V.Cost annotation). All network data as well as a range of visualization tools are publicly available at VitViz (<https://tomsbiolab.com/vitviz>). As *MYB14* and *MYB15* have very low expression in flower tissues, any detected co-expression relationship with other genes in the whole genome network are most surely attributable to fruit samples, therefore, fruit/flower GCNs are referred from here on as fruit GCNs.

Individual *MYB14*- and *MYB15*-gene centred fruit GCNs revealed that both TFs shared 146 genes of the 420 GCN members. In addition, *MYB14* is present in the GCN of *MYB15* and vice versa. We found 13 *STS* genes present in both fruit GCNs. In addition, 28 *STS* genes are exclusively present in the fruit GCN of *MYB14* meaning that almost all known *STS* genes in the grapevine genome (41 of 47) are present in any one of the fruit GCNs. The six *STS* genes, which are not present in either TF-centred GCN, are *STS4/11/33/34/44*, all of which are putative pseudogenes. Interestingly, *MYB15* does not have any exclusive *STS*s in its fruit GCN, suggesting a more direct relationship of *MYB14* with this gene family. Other notable genes that are common to both GCNs are *PAL*s, *WRKY*s, and other genes related to secondary metabolism. There are five *PAL*s shared by both fruit GCNs, while the *MYB14* and *MYB15* fruit GCNs have two and one exclusive *PAL* gene, respectively. A comparison of *MYB14* and *MYB15* leaf GCNs revealed 180 common genes of 420, with *MYB14* and *MYB15* again being present in each other's GCNs. Regarding *STS* genes, 28 of 47 are present in both MYB GCNs, while four and eight are exclusive for *MYB14* and *MYB15*, respectively. The presence of other notable genes of the secondary metabolism pathways in both GCNs, such as *PAL*s and *WRKY*s, is also remarkable.

Results in both leaf and fruit GCNs greatly overlap with DAP-Seq data, providing transcriptional regulation evidence for many co-expression relationships (Figure 3). In total, the number of co-expressed genes in both fruit GCNs is of 694, of which 23% were identified as MYB-bound genes by DAP-Seq. A similar tendency is observed for the comparison between leaf GCNs and DAP-Seq data. Moreover, among these are genes of interest coding for TFs such as *WRKY*s, *NAC*s, or enzymes within the shikimate, early phenylpropanoid or stilbenoid pathways. Most of the genes identified both by DAP-Seq and co-expression

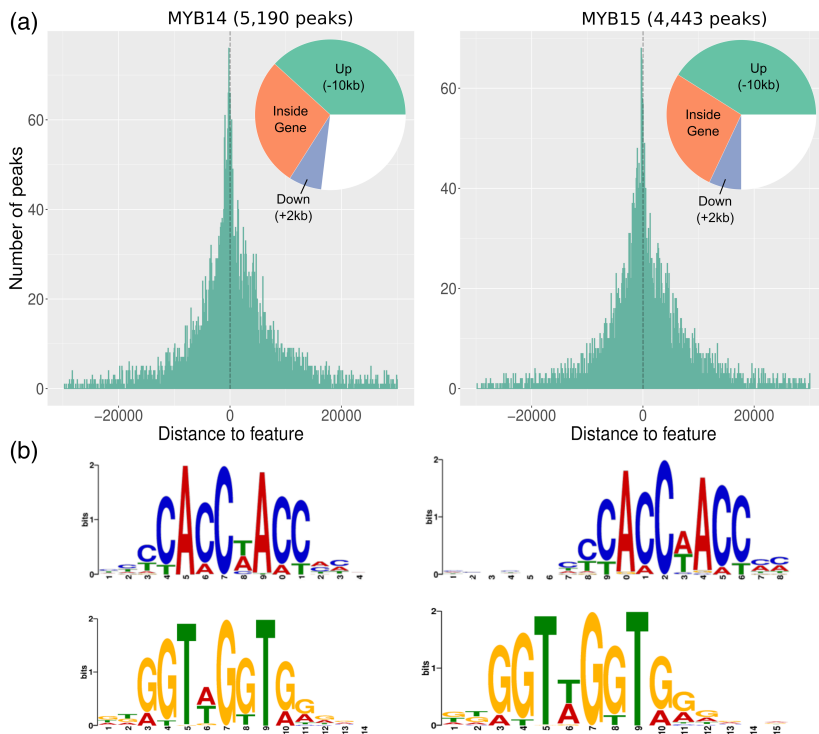


Figure 1. MYB14 and MYB15 DAP-Seq derived cis-trome landscapes in *Vitis vinifera* cultivar 'PN40024'.

(a) DNA-binding events with respect to all transcription start sites of assigned genes. The proportion of binding peaks 10 kb upstream of transcription start sites, inside genes or 2 kb downstream of genes are represented within the pie-charts in green, orange, and blue, respectively.

(b) *De novo* binding motifs, forward and reverse, obtained from the top 600 scoring peaks of MYB14 and MYB15 using MEME suite.

analyses had the MYB binding motifs identified within 5 kb upstream of TSSs. An overlap of these data with the grapevine reference gene catalogue v1.1 (Navarro-Payá et al., 2022) available at Integrape (<http://www.integrape.eu/index.php/resources/genomes/>) revealed a considerable representation of secondary metabolism genes (all the provided gene symbols in this study are in accordance with the reference gene catalogue, v1.1).

Integrating GCNs and DAP-Seq data to identify high confidence targets

Bearing in mind that MYB14 and MYB15 may be major regulators and hence be co-expressed with a large number of genes, a fraction of biologically relevant co-expressed genes is expected to be absent in their GCNs, which have a 420-gene cut-off. Therefore, in addition to the *MYB14* and *MYB15* GCNs, the GCNs of MYB-bound genes were also inspected to interrogate the presence of either *MYB14* or *MYB15*. These 'reciprocal' gene co-expression relationships allowed further prediction of a reliable list of targets.

Gene-centred GCNs were extracted for each DAP-Seq identified gene from the whole genome co-expression network (4041 and 3626 for *MYB14* and *MYB15* respectively). This allowed for further integration of co-expression and TF-binding results by overlapping DAP-Seq results with the reciprocal GCNs. For this overlap, only MYB-bound genes with peaks within 10 kb upstream of the TSS and 2 kb downstream of the end of their gene feature were considered.

Bound genes with a co-expression relationship present in at least one of the two GCNs (i.e. *MYB14* being present in a MYB14-bound gene GCN and/or vice versa) were considered high confidence targets (HCTs). This integration of gene-centred GCNs increased the number of MYB-bound genes supported by network data. For instance, in fruit HCTs an increase from 76 to 145 and from 56 to 127 was observed for *MYB14* and *MYB15*, respectively. The different tissues used for the aggregate networks did not affect the total number of MYB HCTs, obtaining 146 and 145 *MYB14* HCTs in leaf and fruit, respectively. Gene enrichment analysis conducted for each MYB HCT list, using either Gene Ontology (GO) or Kyoto Encyclopedia of Genes and Genomes (KEGG), show trihydroxystilbene synthase activity as the most significant term (Figure S1a and Dataset S2). Moreover, there are a number of interesting terms such as shikimate 3-dehydrogenase activity, phenylalanine ammonia-lyase activity, and other shikimate/early phenylpropanoid related terms. Many of the 57 common fruit HCTs in fact correspond to *PAL* and *STS* genes. Although gene set enrichment analyses offer similar results for both tissue-specific HCTs, the defence response term was exclusive to leaf.

Common MYB14/MYB15 fruit HCTs show the presence of at least one known co-regulator of the stilbenoid pathway, i.e. *WRKY03*, as well as specific shikimate pathway enzymes, such as *DAHPS3* and *SDH4*, and key early phenylpropanoid enzymes such as *4CL8* and *C4H1-3*. Common fruit HCTs also contain 12 *STS* genes while individual

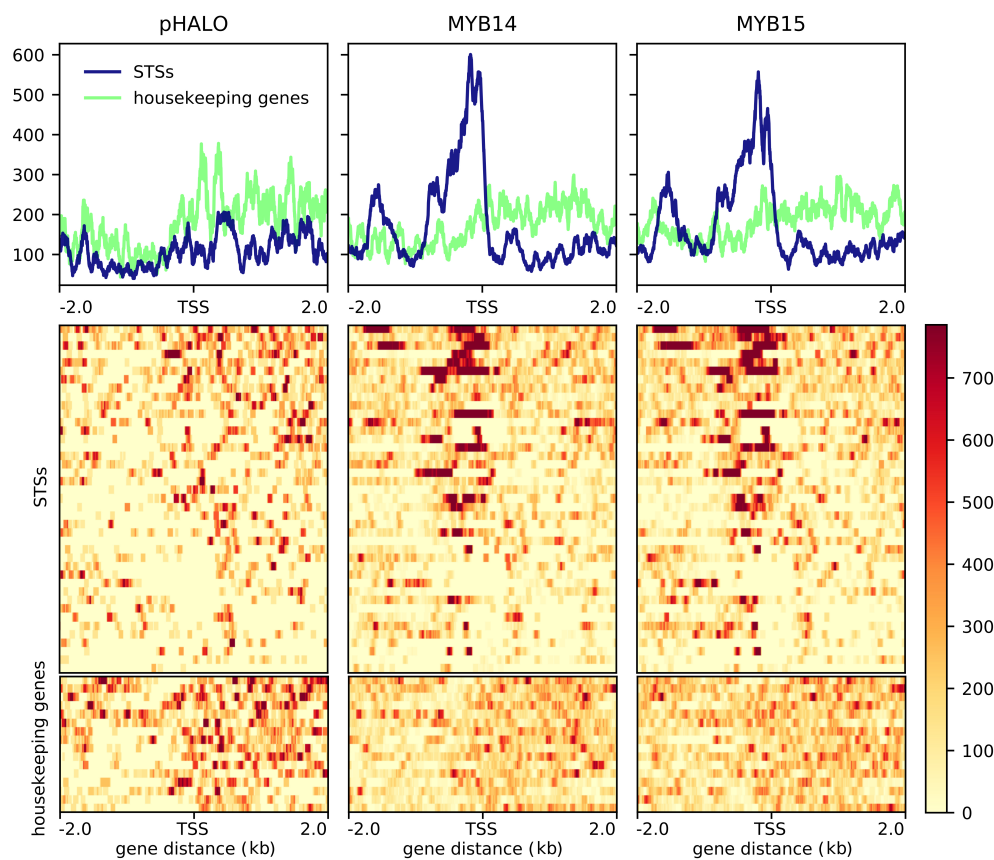


Figure 2. MYB14 and MYB15 DNA-binding events in the promoter regions of stilbene synthase (*STS*) genes.

DAP-Seq binding signal for MYB14 and MYB15 at -2 kb and $+2$ kb from the transcription start site (TSS) of *STS* genes is high when compared with background housekeeping genes both in the density plots, which show the average binding signal, and the heatmaps showing individual gene profiles. *STS* and housekeeping genes used for this figure are listed in Dataset S1. The figures were generated using the Deeptools suite v.3.3.2, computing and normalizing the coverage for each BAM file with a BinSize = 10 and RPKM normalization. RPKM value for each bin is the average between the 10 positions that define each bin. BigWig files for the individual replicates for each transcription factor were merged using bigWigMerge v.2 and bedGraphToBigWig v.4.

fruit HCTs contain seven exclusive *STS* genes in the case of *MYB14* and one in the case of *MYB15*. Common MYB14/15 leaf HCTs show similar composition to fruit HCTs, with the presence of *WRKY03*, *4CL8*, *C4H2*, and 17 different *STS* genes. Eight *STS* genes appear as HCTs for fruit and leaf in both MYB14 and MYB15. Only one *STS* bound gene, the pseudogene *STS3*, is not an HCT (Figure S1b). MYB-bound genes are distributed among the three PPP branches (stilbenoid, lignin, and flavonoid); however, a lack of HCTs is observed among the lignin or flavonoid branches in both fruit and leaf HCTs (Figure S2 and Dataset S2). To corroborate MYB targets further, we overexpressed *MYB15* in grapevine leaves.

Transient *MYB15* overexpression in grapevine confirms many fruit and leaf predicted targets

We validated MYB15 HCTs by transiently overexpressing it in grapevine leaves from 10-week-old plants and monitoring gene expression changes through microarrays. Endogenous and total *MYB15* expression showed a short-

term increase at the initial time-points probably due to the agroinfiltration *per se* (i.e. wounding stress). Nonetheless, a greater and longer-lasting increase in expression was observed thereafter in the *MYB15*-agroinfiltrated samples, which can be fully attributed to the overexpression of the transgene (Figures 4a and S3). WGCNA analysis resulted in 18 modules (Figure 4b, left panel). We found the *MYB15* probe in module eigengene 5 (ME5), which is nearly identical to ME14, ME2, and ME4, all showing a higher expression in 35S:*MYB15* leaves compared with controls at all time-points. These four modules also clustered closely with ME12 and ME16, which still showed similar Z-score patterns across the different samples. We determined that these four modules (ME5, ME2, ME4, and ME14) hold most of *MYB15* targets (Figure 4b). The average gene expression of these four modules is indeed higher in *MYB15*-overexpressed samples with respect to control time-points (Figure 4c).

GO and KEGG enrichment analyses on each module (Figure S4) revealed interesting enrichment terms such as

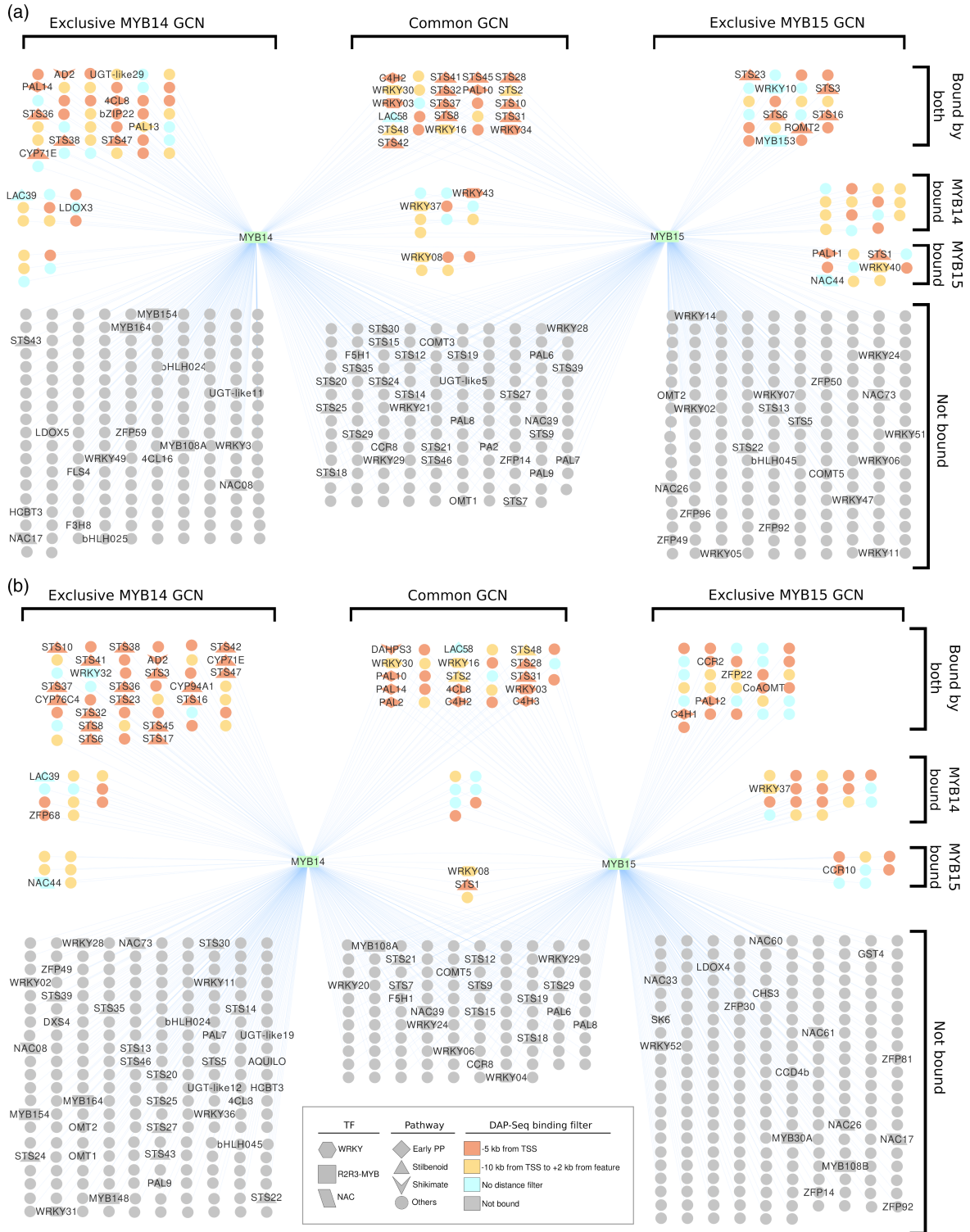


Figure 3. MYB14 and MYB15 leaf and fruit gene co-expression networks (GCNs) share many co-expressed genes, many of which are also detected with DNA affinity purification sequencing (DAP-Seq).

MYB14 and MYB15 GCNs consist of their top 420 co-expressed genes. The colour of each node (i.e. gene) depicts the different distance filters met by the closest DAP-Seq peaks. When genes are bound by both MYB14 and MYB15 the distance filters are only indicated if both transcription factors (TFs) have peaks within them. The shape of each node determines the metabolic pathway or TF family it belongs to.

(a) Leaf GCNs.

(b) Fruit GCNs.

the expected trihydroxystilbene synthase term (ME5). In addition, other terms were *O*-methyltransferase activity (ME5), other shikimate/early phenylpropanoid enzymatic terms, and defence response terms such as 'response to stress', 'response to fungus', and 'chitin metabolic process'. Probe-associated genes with positive fold changes cluster together and they mostly belong to the four modules of interest. Moreover, the probe representing *ROMT1-6* shows a strong fold induction across all time-points (Figure 4d). A similar induction dynamic is observed for other phenylpropanoid-related enzymes (e.g. *4CL8*) and TFs (e.g. *WRKY03*, *WRKY08*, or *WRKY34*).

VviMYB13 shares a high proportion of bound genes with its two S2 co-members

R2R3-MYB S2 is composed of MYB13/14/15 in grapevine and Arabidopsis. Alignments and phylogenetic analyses of this subfamily in these and other plant species show that this close evolutionary relationship is in part explained by the conservation of the R2/R3 repeats and their C-terminal FW1 and FW2 domains (Figure S5). The closest grape MYBs to S2 have lost at least one of these domains (i.e. *VviMYB135-138*). Thus, we additionally checked the potential contribution of *VviMYB13* to the regulation of secondary metabolism in grapevine. Our *VviMYB13* DAP-Seq analysis reported 17 019 binding events assigned to 10 624 different genes (Figure S6 and Dataset S1). Most *VviMYB14/15* bound genes are contained within *VviMYB13* bound genes (Figure S6b). Approximately 74% of peaks are present between 10 kb upstream of TSSs and 2 kb downstream of the annotated gene ends. *VviMYB13* binds to *STS* promoter regions in a similar manner to *VviMYB14/15* (Figures 5b and S7). The DNA-binding motif obtained for *VviMYB13* is practically identical to that of *VviMYB14/15* (Figure 1a) and even to the binding motif identified for *AtMYB13* by (O'Malley et al., 2016) (Figure 5a). Interestingly, *VviMYB13* was observed to bind at -2.3 kb from the TSSs of both *VviMYB13* itself and *VviMYB14*. Gene set enrichment analyses for *AtMYB13* and *VviMYB13/14/15* bound genes illustrate the functional relatedness of grape S2 MYBs with an Arabidopsis homologue (Figure 5c and Dataset S1). Surprisingly, *AtMYB13* also presents shikimate- and early phenylpropanoid-related terms as its *V. vinifera* homologues. As *Arabidopsis thaliana* does not produce stilbenes, the trihydroxystilbene synthase activity term is only significantly enriched for *VviMYB13/14/15*. An interesting term that is only present for *AtMYB13* is lignin biosynthesis suggesting a potential regulatory diversification between Arabidopsis and grape homologues.

Despite the similarities of *VviMYB13/14/15* through DAP-Seq analysis, an overall lack of expression of *VviMYB13* across the SRA experiments used in the aggregate whole genome co-expression networks meant that no valuable

co-expression data could be extracted for *VviMYB13* (Figure 5d). Transient overexpression of *VviMYB13* in grapevine, or further inspection of datasets where *MYB13* is induced could be used to narrow down its bound genes to a list of HCTs.

As shown previously, *VviMYB13/14/15* bound genes include many members of the shikimate pathway, early PPP, and stilbenoid pathways (Figure S8). Surprisingly, we also found bound genes among the lignin and flavonoid branches, many of which are known to be oppositely expressed compared with *STS* and *MYB14/15* genes; as an example, chalcone synthase genes have been shown to be strongly repressed under UVC or fungal stress conditions, which in turn largely activate *MYB14/15* and *STS* gene expression (Blanco-Ulate et al., 2015; Vannozzi et al., 2012). This observation and the fact that these additional bound genes are not HCTs (Figure S1) may suggest that they are negatively regulated by the interaction of other proteins with R2R3-MYBs from S2.

Characterization of new stilbenoid-pathway genes identified as S2 MYB targets

Several MYB14-15 HCTs and MYB13-bound genes encode different types of functionally uncharacterized enzymes. Among those potentially producing stilbene derivatives, we found several *O*-methyltransferases, laccases, hydroxylases (cytochrome P450s), and glycosyltransferases, all representing interesting cases for further functional validation, which would position DAP-Seq as a tool for novel enzyme identification.

As a first approach to study these potentially new enzymes of the stilbenoid pathway, we conducted a 7-day time-course experiment of grapevine (cv. 'Gamay Fréaux') cell cultures, elicited with methyl-jasmonate and cyclodextrins (MeJA + CD) that are known to activate, largely and specifically, the expression of *MYB14/15* (Almagro et al., 2014). As cell suspensions from this *teinturier* (i.e. red fleshed) cultivar accumulate anthocyanins ectopically in the presence of light, we grew and subcultured the cells for several passes in dark conditions to avoid the premature commitment of the PPP for flavonoid production of flavonoids in detriment of stilbenoids. Anthocyanin-devoid (i.e. white) cells were elicited for the quantification of secondary metabolites at 4 and 7 days. An accumulation of resveratrol was observed only inside MeJA-treated cells at both time-points while anthocyanin content was greatly reduced compared with the control at day 7 (Figures 6 and S9). Anthocyanins show an increasing tendency only in the control due to the effect of sugar replenishment (conducted at the beginning of the time-course) while in the elicited cells jasmonate seems to direct preferentially the flux of the PPP for stilbene accumulation rather than flavonoids despite the effect of the sugars. Moreover, there was an increase of piceatannol and viniferin (particularly on day

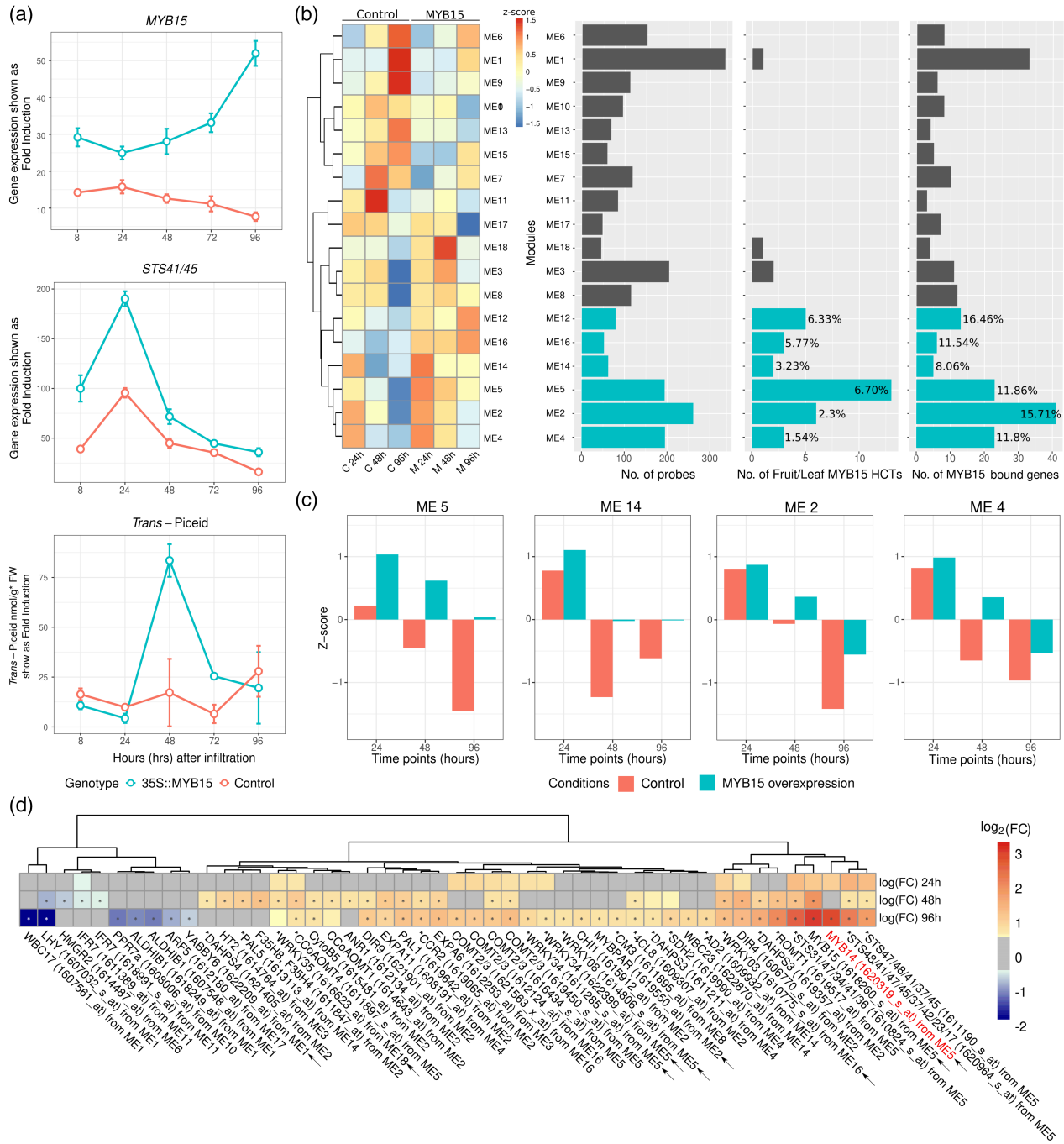


Figure 4. Secondary metabolism genes are induced upon overexpression of *MYB15* in grapevine leaves. (a) Increase in *MYB15* and *STS* expression as well as in *trans-piceid* content expressed as fold induction compared with non-infiltrated leaves for both 35S::Vvi-MYB15 and empty vector control transformations (blue and red, respectively). (b) Clustered mean Z-scores across each time-point for modules obtained by WGCNA from $\log_2(\text{RMA} + 1)$ expression data. Bar plots for every module show: (i) number of probes per module, (ii) number of probes with at least one MYB15 leaf or fruit high confidence target (HCT) assigned, and (iii) number of probes with at least one MYB15 DNA affinity purification sequencing bound gene assigned. Percentages are provided for the four cluster modules of interest (marked in blue). ME, module eigengene. (c) Mean Z-scores across different leaf samples grouped by agroinfiltration treatment and time-point for the four modules of interest. (d) Temporal gene expression changes in response to *MYB15* overexpression for MYB15-bound genes (as well as *MYB14*) present in the grapevine reference gene catalogue (v1.1) with \log_2 FC greater than 0.53 or smaller than -0.53 at at least one time-point. \log_2 FC values between -0.53 and 0.53 are greyed out. FC, fold change.

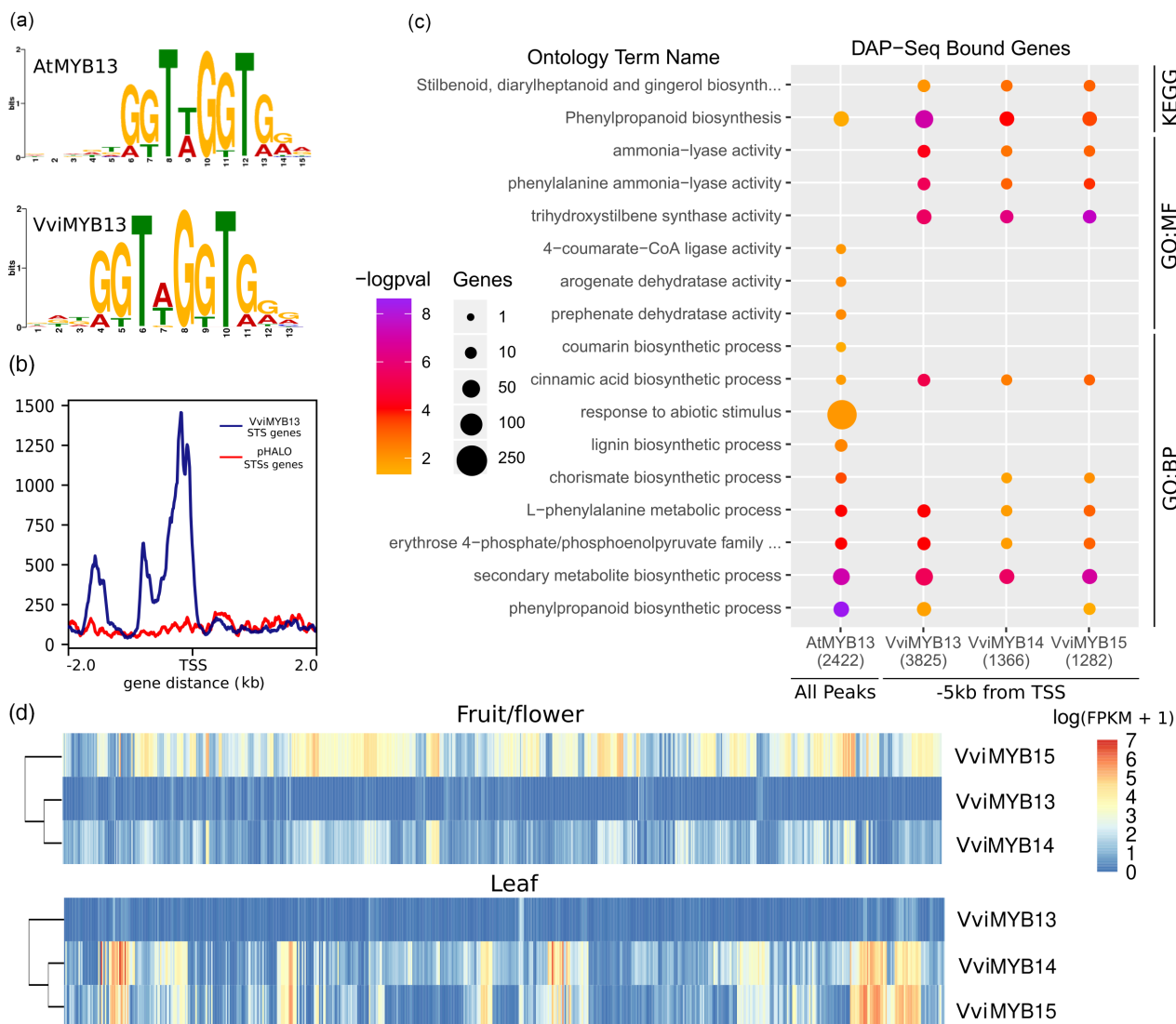


Figure 5. VviMYB13-binding features are similar to both VviMYB14 and VviMYB15 despite its different tissue-specific gene expression patterns.

(a) AtMYB13 (O'Malley et al., 2016) and VviMYB13 DNA-binding motifs identified using MEME suite.

(b) VviMYB13 DNA affinity purification sequencing (DAP-Seq) binding peaks events with respect to 2 kb upstream and downstream of stilbene synthase (STS) transcription start sites. The figure was generated using the DeepTools suite v.3.3.2, computing and normalizing the coverage for each BAM file with a BinSize = 10 and RPKM normalization. RPKM value for each bin is the average between the 10 positions that define each bin. BigWig files for the individual replicates for the transcription factors were merged using bigWigMerge v.2 and bedGraphToBigWig v.4. TSS, transcription start site.

(c) Selection of significantly enriched terms from Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) ontologies is shown for AtMYB13 and VviMYB13/14/15 bound genes. A $-\log_{10}p$ value scale is provided where a higher value represents a greater statistical significance on a continuous colour scale from orange to purple. Number of bound genes intersecting with each ontology term is represented by point size. GO levels are indicated as GO:MF (molecular function) and GO:BP (biological process).

(d) A continuous colour scale from blue to red represents the $\ln(x)$ of FPKM + 1 for VviMYB13/14/15 across different SRA runs for every experiment used to build the fruit/flower and leaf aggregate networks.

7), while piceid accumulation did not present significant changes across treatment and time-points. We compared the accumulation of metabolites with the expression of the shikimate, early phenylpropanoid, and stilbene pathway genes (including the potentially new pathway genes) by reanalysing the previously published microarray study of Almagro et al., (2014) of 24 h MeJA + CD-treated cv. 'Gamay Fréaux' cells (Figure 6).

We first manually curated and improved the current MapMan grapevine ontology with the complete list of genes within the newly created stilbene pathway terms, among which we created 'Trans-resveratrol di-O-methyltransferase activity' and 'Resveratrol Glycosyltransferase activity' (Dataset S3). The combined DAP-Seq results of MYB13/14/15 were included in the context of these pathways, highlighting those genes that are bound

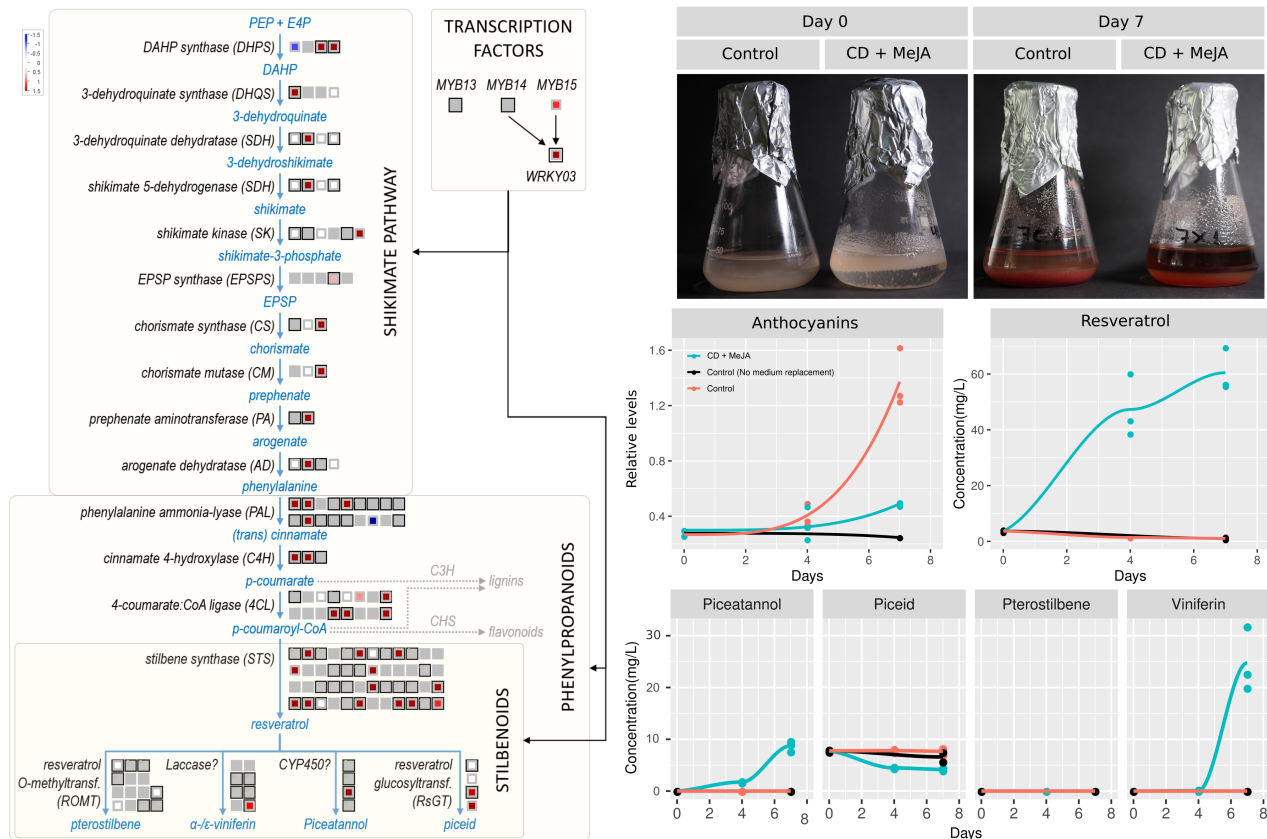


Figure 6. MYB13/14/15 DNA binding of shikimate, early phenylpropanoid and stilbenoid pathway genes shown together with microarray gene expression of methyljasmonate and cyclodextrins (MeJA + CD) elicited cells (Almagro et al., 2014).

Left panel: genes are ordered from left to right and top to bottom. Genes that are surrounded by a black box correspond to MYB-bound genes (within -10 kb of the transcription start site, the gene body itself, or $+2$ kb from the end of the gene). Significant (0.05 P -value threshold) positive \log_2 fold-changes are shown in red and negative changes in blue. Genes with no associated microarray probe are greyed out while white boxes mark genes with no significant differential gene expression. *WRKY03* is included as it has been shown to cooperate with R2R3-MYBs for inducing *STS* gene expression (Vannozzi et al., 2018). Right panel, top: grape cell suspensions at the beginning of the elicitation experiment (Day 0) and 7 days after (Day 7). Bottom: anthocyanin content, measured by spectrophotometry, and stilbenoid quantification by liquid chromatography–mass spectrometry (i.e. resveratrol, piceatannol, pterostilbene, piceid and viniferin in the grape cells at Days 0, 4, and 7. An additional sample, corresponding to non-elicited grape cells subcultured in the old same growth media (i.e. with no sugar replenishment) was taken at Day 7.

by the S2 MYBs. The metabolite profile observed in our experiment correlates with the activation of *MYB15* and almost all its bound genes (no probe available for *MYB14/ MYB13*), including resveratrol-modifying candidate genes. Production of viniferin and piceatannol also matched the upregulation of their putative related enzymes. In particular, *LAC59* and several cytochrome P450 *CYP76C4* genes, identified as MYB14/15 HCTs, could be directly responsible for the enzymatic reactions producing these compounds in our elicited grape cells. The microarray data also support the lack of pterostilbene accumulation, as no upregulation of *ROMT1* or *ROMT*-like genes was observed.

We inspected additional public RNA-Seq datasets to see if MYB15/14 targets with potential resveratrol-modifying activity were co-expressed with their corresponding stilbenoid metabolites. For instance, by reanalysing the transcriptome responses upon *Botrytis* infection stage 2 in cv.

‘Semillon’ (Blanco-Ulate et al., 2015) we see a large group of potential resveratrol glycosyl-transferase (*RsGTs*), *CYP450* and *laccases*, many of which were identified as HCT, being highly induced (Figure S10). In this case, the expression of laccases and *RsGTs* matched the differential accumulation of viniferins and piceid in response to infection.

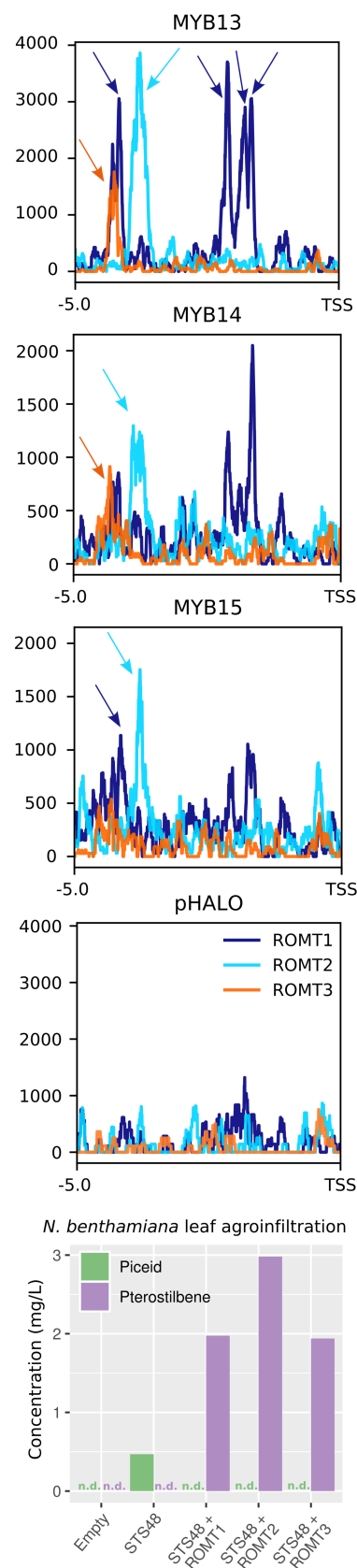
As both *ROMT* transcripts and pterostilbene were undetected in the elicited-grape cells or in the *Botrytis*-infected samples, we further inspected their expression in other transcriptomics datasets to look for their condition-specific induction. The cv. ‘Corvina’ atlas (Fasoli et al., 2012) represents a suitable dataset with >50 samples corresponding to all types of vegetative and reproductive tissues and green-to-mature stages. The stilbenoid pathway and the regulators *WRKY03* and *MYB14/15* show high expression in leaf and berry development, in particular at senescent or

mature stages for both types of organs. For the case of *ROMT1* and *ROMT-like* genes, we see an almost specific upregulation in senescent leaves (Figure S11). Thus, we extracted RNA from these senescent grapevine organs and amplified two *ROMT-like* genes, here named *ROMT2* and *ROMT3*. Together with the previously characterized *ROMT1* gene used as a positive control, we transiently overexpressed them in *Nicotiana benthamiana* leaves in combination with *STS48* that is also an MYB14/MYB15 HCT. All these three *ROMT* genes, which show several MYB-binding events upstream of their TSSs, were able to promote pterostilbene accumulation in tobacco leaves as seen by liquid chromatography (LC)–mass spectrometry (MS) analysis (Figure 7). On the contrary, the sole overexpression of *STS48* only produced piceid accumulation while the agroinfiltrated empty vector did not produce any stilbenes. Overall, the DAP-Seq data presented here allowed us to select candidate pathways genes for their enzymatic characterization *in planta*.

DISCUSSION

DNA binding and gene co-expression as a proxy for gene regulatory networks

GCNs are based on the ‘guilt-by-association’ principle whereby correlation in expression implies biological association (Wolfe et al., 2005). This results in a promising tool for predicting gene regulatory networks, in particular to identify gene targets and their regulators. Network aggregation has been shown to improve performance with the use of microarray grapevine datasets (Wong, 2020), and here it has been applied to all the publicly available leaf and fruit RNA-Seq data for grapevine. The resulting network performance, estimated through the AUROC measurement (using MapMan BIN categories) was in our case approximately 0.75, which is an improvement over previously published RNA-Seq aggregate networks (Wong, 2020). Although network performance was high, a co-expression relationship does not necessarily indicate a biological connection (Gillis & Pavlidis, 2012) and the overlap with other datasets of experimental origin is required. This study represents a successful application of genome-wide DNA-



binding site interrogation in combination with reciprocal GCNs used to draw the gene regulatory networks of TFs controlling specialized metabolism.

The three MYBs studied here belong to S2 of the R2R3-MYB subfamily. Some members of this clade have been functionally characterized in other species. For instance, in *A. thaliana* *AtMYB13* is involved in response to UVB (Qian et al., 2020), *AtMYB14* is related to cold tolerance (Chen et al., 2013) and *AtMYB15* is involved in the lignin biosynthesis and immune responses (Kim et al., 2020). In this study, we show that the grape members of this subgroup are all tightly connected to the shikimate pathway and PPP, particularly the early steps of the latter and most of the stilbenoid branch within this pathway. We show that not only *STS* genes are S2 MYB targets but also resveratrol-modifying enzymes such as glycosyltransferases and *O*-methyltransferases, characterized here. We also suggest laccases and CYP450 hydroxylases as potential viniferin- and piceatannol-producing enzymes based on DAP-Seq, co-expression networks, and the inspection of large transcriptomic datasets where these genes are expressed in correlation with stilbenoid accumulation.

In this work, we show considerable overlap between MYB-related GCNs and their cistrome data, validating the use of co-expression relationships as a predictor of regulation. In addition, many of the HCTs identified by the overlap of these datasets was also confirmed by the transcriptomic analysis of transient overexpressing plants. Most of the statistically significant induced genes upon *MYB15* overexpression are indeed HCTs (Figure 5). This highlights the power of combining both DAP-Seq and reciprocal GCN data as a similar biological meaning is drawn from this approach when compared with the overexpression analysis. However, in the case of *MYB13*, due to its low expression in both leaf and fruit SRA experiments, we could not generate it, as it would not have been very informative. Beyond target identification, GCNs can also point to MYB-interacting partners such as WRKY03 or WRKY43, which interact with MYB14 to increase *STS* expression (Vannozzi et al., 2018). Our GCNs also show a high co-expression of S2 MYBs with *MYB154* from S14, which has been recently shown to activate a few *STS* gene promoters in *V. quinquangularis* (Jiang et al., 2021).

Shared and exclusive regulatory features of S2 MYB TFs

MYB13, MYB14, and MYB15 share a high proportion of bound genes as well as identical DNA-binding motifs suggesting a partial overlap of their regulatory roles. This conserved motif has already been described as a MYB *cis*-regulatory element; in particular, the AC-element motif (CACCT/A)ACC) previously identified for *AtMYB15* in *A. thaliana* (Romero et al., 1998). As described in Kelemen et al. (2015), Arabidopsis S2 R2R3-MYBs, as well as other subgroups such as S1, S3, S13, and S24, show a high

degree of specificity towards the AC-element motif. Within the stilbenoid pathway, we found that MYB-binding patterns, i.e. in terms of their sequence and positions, across individual *STS* genes were remarkably similar for these three R2R3-MYBs, providing strong evidence that the recent tandem duplications observed for the *STS* gene family involved upstream regions to a certain length extension. MYB13/14/15 TFs bind in two distinct *STS* promoter elements; the first at 400 bp from the TSS that is present in most *STS*s, and a second at 2000 bp, which is only present for a subgroup of them. This observation suggests that *STS*s may possess gene-specific differences in their regulation. The presence of two binding sites in some of them also offers the possibility of synergistic or dual regulation. MYB TFs regulating other specialized metabolic pathways have been known to interact with other proteins from different TF families such as bHLHs or WDRs as well as other MYBs. According to Höll et al. (2013) MYB15 and MYB14 lack a bHLH-interacting motif within their R3 repeat, which is present in other bHLH-interacting MYBs (e.g. those from S5 and S6) but it is unclear if they have dimerization domains. Further studies could address if MYB14 and MYB15 homo- or heterodimerize. If this is the case, the idea of simultaneous binding at different regions of the promoter gains strength.

Our GCNs may help to point out *in vivo* regulatory differences between TFs with otherwise identical *in vitro* DNA-binding profiles. This appears to be the case for MYB14/15 as the exclusive presence of 28 *STS* genes in the *MYB14* fruit network points to a very interesting MYB14-specific relationship for a subset of *STS* genes. This observed contrast in MYB14 and MYB15 co-expressed genes could be explained by promoter differences in *MYB14/15* leading to separate regulation by upstream TFs. On the other hand, structural changes between the TFs, leading to distinct interactions with other co-regulators, could also explain this observation. Most *STS*s are also MYB14 and MYB15 fruit HCTs. The two previously reported *STS* targets of MYB14/MYB15 (Höll et al., 2013) were either HCTs (*STS41*) or present in the *MYB14/15* GCNs (*STS29*). *STS29* was not defined as a fruit or leaf HCT as no binding events were automatically called by the peak detection software. However, a potential peak is observed for the three MYBs when examining the mapped sequencing reads in the promoter region of the gene.

The enrichment analyses of leaf and fruit HCTs, resulting from the overlap of GCNs and DAP-Seq data, corroborate the newly found connections of *MYB13/14/15* with shikimate and early phenylpropanoid genes. Several *PAL* genes appear as HCTs in addition to *STS*s, supporting the idea that MYB14 and MYB15 may have the ability to increase and favour the activity of the PPP for the formation of stilbenes, in detriment of lignins or flavonoids. In fact, we may suggest that MYB14/15 might be involved in the

repression, direct or not, of chalcone synthases, which are direct competitors of *STS*s activity. We also suggest that these MYB TFs may rewire the whole metabolic grid of the cell and favour the increase of carbon flux into the PPP by activating the shikimate pathway. Direct DNA-binding observed for at least one isoenzyme of every shikimate and early PPP step highlights the central character of the studied R2R3-MYB regulators in activating stilbenoid metabolism. Despite not being able to perform leaf or fruit GCNs for *MYB13*, we suggest that these three proteins are highly redundant in their regulatory potential. This redundancy is still partial as GCNs also suggest differences between these two TFs.

TFs act in a hierarchical manner and based on our data, *MYB14* and *MYB15* appear to be near the top of this hierarchy, not only because they are able to control a large portion of the shikimate/early phenylpropanoid and stilbene pathways but also because approximately 21% of their targets are TFs, respectively. In fact, at least nine *WRKY* TF genes are present among the common *MYB14/15* HCTs. *WRKY* TFs have been suggested as MYB co-regulators in the control of *STS* expression (Wang et al., 2020; Xi et al., 2014; Xu et al., 2019) and in many defence responses to pathogens. In particular, *WRKY03* is bound by both TFs and present in *MYB14* and *MYB15* leaf and fruit GCNs. This observation adds one more layer of complexity to the hierarchical regulation of *MYB14*, which requires *WRKY03* to increase *STS* gene expression (Vannozzi et al., 2018).

The data presented here suggest that grapevine S2 members act mainly as positive regulators of transcription; however, based on the reduction of gene expression of some bound genes upon *MYB15* induction, S2 MYBs could also be involved in negative regulation possibly requiring other co-repressors. An alternative hypothesis could be that *MYB15* may be competing for promoter space with other stronger transcriptional activators that recognize the same target motifs, as suggested by Tamagnone et al. (1998). An interesting example of this repression is observed among circadian rhythm-related genes. Given the known circadian behaviour of *STS* gene expression (Carbonell-Bejerano et al., 2014), we further explored the potential relationship of *MYB15* with the core molecular components of the circadian clock. *MYB15* leaf overexpression led to a strong downregulation of *LHY* (Dataset S3), one of the two components of the main oscillator within the system. Moreover, different components of the clock such as *LHY* itself, *PPR7a/b*, *TOC-like*, *GI*, and *ELF4* are bound by either of *MYB13/14/15* (Dataset S1), which further suggest a direct connection with the circadian rhythm. RVE8-type MYBs have been previously shown to have roles regarding the control of the circadian rhythm in *A. thaliana* (Shalit-Kaneh et al., 2018); however, the involvement of R2R3-MYBs has not been previously described.

Integration of DAP-Seq and expression data as a tool for the identification of putative secondary metabolic pathway enzymes

By overlapping the different datasets generated in this work, we were able to identify putative stilbenoid genes among MYB targets, such as laccases (i.e. multicopper oxidases potentially involved in resveratrol oligomerization), resveratrol hydroxylases, resveratrol *O*-methyltransferases (ROMTs) and resveratrol glycosyltransferases (RsGTs). Within *O*-methyltransferases, *ROMT1* has been the only gene described to date as involved in the synthesis of pterostilbene (Schmidlin et al., 2008), and here we show that *MYB13* and *MYB15* bind within its 5-kb upstream region. In addition, we show that the three S2 MYBs bind in a similar region with respect to *ROMT2* and *ROMT10* promoters. Despite none of these *ROMTs* appearing in *MYB14* or *MYB15* fruit GCNs, some of them do appear in leaf GCNs. This is in line with the lack of expression of *ROMT* genes in fruit or flower tissues (Figure S11), highlighting the importance of tissue-specificity when dealing with co-expression networks. In fact, the transient overexpression of *MYB15* led to an upregulation of all *ROMT1-6* enzymes recognized by the microarray probes. Clustering of microarray data (Figure 4) also showed that the expression profiles of the *ROMT* enzymes was very similar to *MYB15* and hence the gene set enrichment analysis of cluster module ME5, where *MYB15* is found, showed enrichment of the *O*-methyltransferase activity ontology term.

We also identified cytochrome P450 and laccase (multicopper oxidases) among MYB targets. Both of these enzyme classes are suggested to take part in the hydroxylation and oligomerization of resveratrol. This has not been shown experimentally except for the production of piceatannol by a transiently expressed human cytochrome P450 in *Nicotiana bethamiana* (Martínez-Márquez et al., 2016). We identify particular genes from these two families, which are highly induced in stilbenoid-accumulating conditions such as *Botrytis* infection (e.g. at Stage 2 of the infection; Figure S10) as well as being present in S2 MYB GCNs or HCT lists. Nevertheless, further studies are required for their functional demonstration.

The approach of integrating DAP-Seq data and reciprocal GCNs is particularly relevant in non-model species; however, it needs sufficient public transcriptomic data for a given species. Online tools for the inspection of DAP-Seq data would certainly help in the identification of novel TF targets for their selection and further characterization. Therefore, we developed a public grapevine DAP-Seq visualization application named DAP-Browse, present in the Vitis Visualization Platform (Figure S13; VitViz; available at <https://tomsbiolab.com/vitviz>).

EXPERIMENTAL PROCEDURES

DNA affinity purification sequencing

Genomic DNA (gDNA) was purified from young grapevine leaves of cv. 'Pinot Noir' (clone PN-94) according to Chin et al. (2016). gDNA library construction, DAP-Seq, peak calling, and motif analysis were performed as described in Bartlett et al. (2017) with the following modifications. The gDNA samples (800 ng/library) were sonicated into 200-bp fragments on a focused ultrasonicator (Covaris, Woburn, MA, USA). The gDNA library then underwent end-repair, A-tailing, and adapter ligation for subsequent Illumina-compatible sequencing. Successful library construction was verified by gel electrophoresis of sonicated gDNA and qPCR of adapter ligated fragments. *MYB14*, *MYB15*, and *MYB13* were amplified from cv. 'Pinot Noir' and cloned into the pIX-HALO expression plasmid (TAIR Vector: 6530264275) with the HALO-tag at the N-terminal. Clones were verified by *XhoI* digestion. The MYB-expression vectors were used in a coupled transcription/translation system (Promega, Madison, WI, USA) to produce MYB Halo tagged proteins. Protein production was confirmed by western blot using an anti-HaloTag antibody (Promega). HaloTag-ligand conjugated magnetic beads (Promega) were used to pull down the HaloTag-fused TFs. Pulled-down TFs were exposed to DAP-Seq gDNA libraries for MYB-DNA binding. Four hundred nanograms of gDNA library was used per DAP-Seq reaction. The bound DNA was then eluted and sequencing libraries were generated by PCR amplification. Sequencing was carried out using Illumina NextSeq 500 set up to obtain 30 million 1×75 bp single-end reads. An empty pIX-HALO expression vector was used as a negative control to account for non-specific DNA binding as well as copy number variation at specific genomic loci. Two replicates (individual and independent gDNA-TF interaction samples) were used for all experiments including the control.

The latest published genome assembly for *V. vinifera* cv. 'PN40024' is the 12X.v2, which is associated to V1 and VCost.v3 annotations, containing 29 971 and 42 414 gene models, respectively (Canaguier et al., 2017; Jaillon et al., 2007). Although the VCost.v3 annotation has been manually curated for the *STS* family, it also contains non-protein coding genes (such as lncRNAs and miRNAs) most of which have only been automatically annotated. As the main focus of this work was to study the regulation of protein-coding genes in secondary metabolism (stilbenoid pathway in particular), a new annotation file combining all V1 gene models with the updated *STS* VCost.v3 gene models was created. The merged gff3 annotation file is available at <http://tomsbiolab.com/scriptsandfiles>.

DAP-Seq reads were mapped to the 'PN40024' 12X.v2 reference genome using bowtie2 (Langmead & Salzberg, 2012), version 2.0-beta7, with default parameters and post-processing to remove reads that have MAPQ scores lower than 30. Peak detection was performed using GEM peak caller (Guo et al., 2012) version 3.4 with the 12X.v2 genome assembly using the following parameters: '-q 1 -t 1 -k_min 6 -k_max 20 -k_seqs 600 -k_neg_dinu_shuffle', limited to nuclear chromosomes. The replicates were analysed as multi-replicates with the GEM replicate mode. Peak summits called by GEM were associated with the closest gene model in the custom annotation file using the BioConductor package ChIPpeakAnno (Zhu et al., 2010) with default parameters (i.e. NearestLocation). *De novo* motif discovery was performed by retrieving 200-bp sequences centred at GEM-identified binding events for the 600 most enriched peaks and running the meme-chip tool (Machanic & Bailey, 2011) in MEME suite 4.10.1 with default parameters. Gene set enrichment analysis were conducted for the

three analysed MYB-bound gene lists (Dataset S1). DAP-Seq replicate reproducibility was assessed using DeepTools suite v.3.3.2 (Figure S14). Briefly, for each TF, genomic regions where peaks had been detected were extracted and divided into bins. The DAP-Seq binding 'signal' (read counts normalized by library size) was extracted in all the bins and averaged to create an overall binding signal for each peak. This was done independently for replicate 1 (y-axis) and replicate 2 (x-axis) for all peaks, computing the Spearman correlation coefficient for each TF. This analysis allowed us to demonstrate the high reproducibility between TF replicates, with correlations over 0.8. We computed the same plots for the pIXHALO replicates across the peaks detected for each TF, resulting in low correlation, as expected, because the reads sequenced in these experiments are mostly random noise. We thus corroborate the robustness of the results derived from using two replicates, as in shown in O'Malley et al. (2016) (Figure S1b), improved compared with ChIP-seq, where replicate correlation ranges between 0.6 and 0.7).

Generation of condition-dependent aggregate whole genome co-expression networks and extraction of gene-centred co-expression networks

Transcriptomic RNA-Seq SRA studies (Illumina sequencing) from fruit/flower and leaf tissue samples were downloaded. SRA studies were manually inspected and filtered to keep those that were correctly annotated and contained four or more data sets (i.e. runs) while also excluding SRA studies concerning microRNAs, small RNAs and non-coding RNAs.

In total, 35 and 42 SRA studies were obtained, encompassing 807 and 670 runs from fruit/flower and leaf samples, respectively (Dataset S2). Reads were trimmed with fastp (Chen et al., 2018), version 0.20.0 with the following parameters: '-detect_adapter_for_pe -n_base_limit 5 cut_front_window_size 1 cut_front_mean_quality 30 -cut_front cut_tail_window_size 1 cut_tail_mean_quality 30 -cut_tail -l 20'. After trimming, the runs are aligned with STAR (Dobin et al., 2012), version 2.7.3a, using default parameters. Raw counts were computed using FeatureCounts (Liao et al., 2013), version 2.0.0, and the VCost.v3_27.gff3 gene models (available at <http://www.integrate.eu>). Each SRA study was analysed individually to build a highest reciprocal rank (HRR) matrix (Mutwil et al., 2011). Each raw counts matrix was normalized to FPKMs, and genes with less than 0.5 FPKMs in every run of the SRA study were removed. The Pearson's correlation coefficients of each gene against the remaining genes was then calculated for each SRA study (across all run matrices) and ranked in descending order. Ranked Pearson's correlation coefficient values were used to compute HRRs among the top 420 ranked genes (420 roughly equals 1% of all VCost.v3 gene models), using the following formula: $HRR(A,B) = \max(rank(A,B), rank(B,A))$, generating a HRR matrix for each SRA study. To construct the aggregate whole genome co-expression network, the frequency of co-expression interaction(s) across individual HRR matrices was used as edge weights, and after ranking in descending order, the top 420 frequency values for each gene were chosen to build the final aggregate networks. The top 420 most highly co-expressed genes for any gene of interest, was used to generate individual GCNs.

Network functional connectivity (i.e. performance) across all given annotations and genes was assessed as in Wong (2020) by neighbour-voting, a machine learning algorithm based on the guilt-by-association principle, which states that genes sharing common functions are often co-ordinately regulated across multiple experiments (Verleyen et al., 2014). The evaluation was performed using the EGAD R package (Ballouz et al., 2016) with

default settings. The network was scored by the AUROC across MapMan V4 BIN functional categories associated to the VCost.v3 annotation (Dataset S2) using threefold cross-validation. MapMan BIN ontology annotations were limited to groups containing 20–1000 genes to ensure robustness and stable performance when using the neighbour-voting algorithm.

Prediction of HCTs in grape reproductive and vegetative organs

To identify MYB14 and MYB15 HCTs, TF-bound genes identified with DAP-Seq were overlapped with data extracted from the aggregate whole genome co-expression network derived from fruit/flower and leaf transcriptomic data. Individual GCNs were extracted from the whole genome network for each MYB14/15 bound gene and for both MYBs to check if a particular MYB-target gene pair had at least one co-expression relationship. This relationship was considered positive if either the DNA bound gene was present in the MYB GCN or the respective MYB gene was present in the GCN of the candidate gene. Thus, bound genes for each MYB TF with a positive co-expression relationship with their respective MYB TF were considered as fruit/flower or leaf HCTs. GO and KEGG enrichment analyses for HCTs were performed using the gprofiler2 R package (Kolberg et al., 2020) with default settings. A significance threshold of 0.05 was chosen for *P*-values adjusted with the Benjamini–Hochberg correction procedure (Benjamini & Hochberg, 1995).

Transient MYB15 overexpression, microarray time-course, and *trans*-piceid quantification in leaves

Vitis vinifera cv. ‘Shiraz’ leaves from 10-week old plants were transformed via *Agrobacterium*-mediated infiltration as in Merz et al. (2015) with MYB15 (35S::VviMYB15) or an empty vector control. A green fluorescent protein vector (35S,GFP-GUS) was used in each agroinfiltration to estimate transformation efficiency. Samples were subsequently collected at five different time-points (8, 24, 48, 72, and 96 h) in separate replicate pools (L1 and L2) consisting of three leaves from independent plants per time-point ($n = 2$ biological replicates). Microarray analysis, using the Affymetrix *V. vinifera* Grape Array (A-AFFY-78), was carried out on extracted leaf RNA from three of the time-points described for both control and MYB15 infiltrated samples. Microarray data are available at Dataset S3 (Microarray dataset sheet). Affymetrix probe-to-gene associations were reassigned as here explained. Affymetrix probe sequences (originally associated to V1) were realigned using bowtie with default parameters against the ‘PN40024’ transcriptome derived from the VCost.v3 annotation to improve V1 probe-to-gene assignments (Dataset S3). As each Affymetrix probe has a different number of sequences associated to it (ranging from 8 to 20), a probe-to-gene association was accepted when more than 40% of the probe’s sequences matched the same gene with no mismatches. The percentage was increased to 60% for hits with one mismatch, 70% with two mismatches and 80% with three mismatches. From, in total, 16 602 probes, 59.8% were assigned to a VCost.v3 gene model.

The LIMMA R package (Ritchie et al., 2015) was used for RMA normalization of fluorescence values after which they were \log_2 -transformed. Normalized probes were analysed with the NOISeq R package (Tarazona et al., 2011). Briefly, batch effect was corrected using the ARSvNseq function with default parameters, and then differentially expressed genes were obtained comparing the two MYB15 overexpression lines against the two control lines for each time-point using the NOISeqBIO function. Genes were considered as differentially expressed when posterior probability

>0.95 and $|\log_2FC|>0.53$ in at least one time-point, obtaining 2297 probes corresponding to differentially expressed genes. RMA values of the filtered probes were clustered applying the weighted GCN analysis (WGCNA) R package (Langfelder & Horvath, 2008) generating 18 modules for the signed network. Briefly, WGCNA R package blockwiseModules function was used with a soft-threshold power value of 18 (as no free-scale topology was reached), a deepSplit of 4 and a mergeCutHeight of 0.1, obtaining, in total, 18 modules for the signed network (i.e. considering the sign of correlation).

Microarray analysis results are available at Dataset S3 (MA analysis sheet). The pheatmap R package was used to represent the modules using the calculated Z-scores across samples. The probe-assigned genes in each module were used for GO and KEGG enrichment analyses as described for HCTs. Quantitative expression analysis of MYB15 or STS genes and *trans*-piceid quantification were conducted using L2 leaf pools across the different samples. *Trans*-piceid was quantified on a reverse phase high-performance LC (HPLC) instrument (Krionton, Reno, NV, USA). Quantitative PCR was run in triplicate per time-point and treatment using primers against the endogenous (aligning to the UTR) and transgenic MYB15, STS25/27/29, and STS41/45. Transcript levels were corrected using the housekeeping gene *Ubiquitin1*. All used primers are shown in Dataset S3. The fold of induction reported for both microRNAs and metabolites were calculated relative to background levels in non-infiltrated leaves.

Grapevine cell culture elicitation time-course

MeJA-CD elicitation of cell suspensions was carried out as described in Bru et al. (2006) with the following modifications. Liquid cell suspensions were initiated by transferring solid calli into 250-ml flasks containing liquid culture media (with Gamborg basal salts and Morel vitamins) described in Bru et al. (2006) and grown in an orbital shaker at 120 rpm. Liquid cell cultures were maintained in dark conditions for 2 months, with subcultures being performed every 2 weeks by moving half of the extracellular volume to a new flask and restoring the original volume with fresh media (half medium replenishment). For the elicitation time-course, four flasks of cell suspensions were mixed together and three-quarters of the suspension was filtered using a sterile glass filter (90–150 μm pore size) and a vacuum filtration system to separate the cells from the media. In total, 15 100 ml flasks were prepared with three biological replicates for each time-point and treatment. Day 0 cells were immediately collected after filtering and frozen in liquid nitrogen. For day 4 and day 7 samples, 8 g of filtered cells and 32 g of new liquid media were added in each flask (complete medium replenishment). Treated (elicited) cells were grown in the liquid media with 50 mM methyl-beta-cyclodextrin (CAS no. 128446-36-6) and 84 μl of 47.5 mM MeJA dissolved in 100% MeOH while untreated (control) cell suspensions were grown in liquid media with 84 μl of 100% MeOH. The rest of the original suspension (one-quarter) was subcultured as usual and kept growing at dark for being used as an additional control for half medium replenishment at day 7. All the biological replicates were grown by orbital shaking at 120 rpm in the conditions described above. Cell cultures were sampled at their corresponding time-point, filtered using a vacuum-aided system, and then washed with cold sterile water before freezing liquid nitrogen. Cells were lyophilized at -52°C and 0.63 mbar for 48 h and then transferred to -80°C .

STS and ROMT overexpression in tobacco leaves

ROMT2-3 were amplified using TOPO-D (Invitrogen, Waltham, MA, USA) compatible primers from cDNA derived from grapevine

senescent leaves. RNA extraction and cDNA synthesis were performed using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) and NZY First-Strand cDNA Synthesis Kit, respectively. The final *ROMT2/3*-containing pENTR/D-TOPO plasmids were verified by Sanger sequencing. *ROMT2* and *ROMT3* were transferred into the binary destination vector pB2GW7 including Cauliflower mosaic virus (CaMV) 35S promoter through Gateway LR clonase recombination (Invitrogen). Together with *ROMT1* and *STS48* constructs (Santos-Rosa et al., 2008; Schmidlin et al., 2008) all these were chemically transformed in *Agrobacterium tumefaciens* strain C58 for subsequent agroinfiltration of *N. benthamiana* leaves. The bacterial suspension was kept at an OD₆₀₀ of 0.2 for each construct before infiltration of leaves on their abaxial side (three leaves of one plant for each experimental combination) using a 1 ml syringe without a needle. Agroinfiltration was conducted with the individual empty vector-, *STS*-, or combined *STS*- and *ROMT*-containing bacterial cultures. After 72 h, leaf-tissue samples were collected by dissection to avoid leaf nerves. Collected samples were frozen at -80°C .

Stilbenoid and anthocyanin quantification of grape cell cultures and *Nicotiana* agroinfiltrated leaves

Stilbenoid and anthocyanin quantification of grape cell cultures and *Nicotiana* agroinfiltrated leaves was performed as in Martínez-Márquez et al. (2016) and Nakata and Ohme-Takagi (2014), respectively, with modifications. Stilbenoid metabolites were extracted using 10 mg of cells in 1 ml of 80% methanol with 220 rpm overnight shaking at 4°C . The extracted solution was centrifuged at 14 000 *g* for 10 min, and collected in another tube for HPLC–MS detection and quantification, as described in Hurtado-Gaitán et al. (2017). For the case of tobacco samples, leaves were weighed and grounded in liquid nitrogen. Extraction using 100% MeOH was carried out keeping a ratio of 8 ml of MeOH per 1 g of fresh tissue with 220 rpm overnight shaking at 4°C as described in Martínez-Márquez et al. (2016). After centrifuging at 14 000 *g* for 10 min the supernatant was used for HPLC–MS determination and quantification, as described in Hurtado-Gaitán et al. (2017). Anthocyanin quantification was performed as described in Nakata and Ohme-Takagi (2014). Briefly, lyophilized cells (0.2 ml) were treated with 1 ml of extraction buffer (45% MeOH and 5% HCL). After a 10 min vortex and centrifugation at 12 000 *g* for 5 min, the supernatant was collected into a new tube and centrifuged again under the same conditions. Supernatant absorbance was measured at 530 nm and 637 nm and the anthocyanin relative levels were calculated using the formula: $(\text{Abs}_{530} - (0.25 \times \text{Abs}_{637})) \times 5/(\text{mg of cells})$.

Re-annotation of the grapevine MapMan ontology and representation of previously published microarray and RNA-Seq data

MapMan v4 (Schwacke et al., 2019) annotation was generated based on EGAD, resulting in high-quality annotations for 13 318 of 41 413 VCost.v3 genes (32%) (Dataset S3). A pathway image for shikimate, phenylpropanoid, and stilbenoid pathway was created and annotated using the MapMan tool (Thimm et al., 2004) for both Vcost.v3 and V1 of the grapevine genome annotations. This image includes the DAP-Seq results whether a gene bound by MYB14, MYB15, and MYB13 is shown in black squares. All files are available for download from the GoMapMan exports page (Ramšak et al., 2014) (gomapman.nib.si/export). Significant \log_2 fold-change values from a microarray study involving jasmonate elicitation of cv. 'Gamay Fréaux' cell cultures (Almagro et al., 2014) and from an RNA-Seq study of Botrytis-infected berries of cv. 'Semillon'

(Blanco-Ulate et al., 2015) were represented within the newly generated MapMan pathway image.

ACKNOWLEDGEMENTS

This work was supported by Grant PGC2018-099449-A-I00 and by the Ramón y Cajal program (grant RYC-2017-23 645), both awarded to JTM, and to the FPI scholarship (PRE2019-088044) granted to LO from the Ministerio de Ciencia, Innovación y Universidades (MCIU, Spain), Agencia Estatal de Investigación (AEI, Spain), and Fondo Europeo de Desarrollo Regional (FEDER, European Union). CZ is supported by China Scholarship Council (CSC; no. 201906300087). KG and ZR were supported by the Slovenian Research Agency (grants P4-0165 and Z7-1888). SCH is partially supported by the National Science Foundation (grant PGRP IOS-1916804). This article is based upon work from COST Action CA 17111 INTEGRAPPE, supported by COST (European Cooperation in Science and Technology). Data have been treated and uploaded in public repositories according to the FAIR principles, in accordance to the guidelines found at INTEGRAPPE. The genomic data presented here will be re-analysed and associated to the PN40024.v4 assembly and its structural annotation upon its release. Grapevine cell cultures in solid media were kindly provided by Roque Bru (Universidad de Alicante). Special thanks to Pere Mestre and Philippe Huguency (INRAE Colmar) for providing the 35S:*STS48* construct and the 35S:*ROMT1* positive control used for agroinfiltration, to Anne-Francoise Adam-Blondon and Nicolas Francillon (URGI INRAE Versailles) for guidance in the adaptation of JBrowse during an INTEGRAPPE Short-Term Scientific Mission of LO, and to Susana Selles-Marchart, for her help with stilbenoid quantification.

AUTHOR CONTRIBUTIONS

LO, DN-P, and JTM designed the research; ML, JH, PM, GM, PR, and CZ collected plant material and performed the experiments; LO, AS, DN-P, S-sCH, ZR, and DW conducted all bioinformatic analyses; LO, DN-P, and JTM wrote the paper. JH, PM, JB, AV, and DC contributed with reagents. All authors contributed to interpretations and revisions of the manuscripts.

CONFLICT OF INTERESTS

The authors declare that they have no competing interests.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Gene set enrichment analysis of high confidence targets (HCTs) and STS gene synteny.

Figure S2. MYB14 and MYB15 high confidence targets (HCTs) among shikimate, phenylpropanoid, and stilbenoid pathway genes.

Figure S3. Endogenous MYB15 and STS25/2729 expression in grapevine leaves overexpressing MYB15.

Figure S4. GO and KEGG gene set enrichment analysis for selected WGCNA cluster modules in grapevine leaves overexpressing MYB15.

Figure S5. Phylogenetic and protein motif analyses of Subgroup 2 R2R3-MYB proteins and their closest relatives.

Figure S6. MYB13 DAP-Seq binding peak distribution and overlap of MYB13/14/15 bound genes.

Figure S7. Genome Browser screenshot of MYB13/14/15 DAP-Seq detected peaks for STS10.

Figure S8. MYB13, MYB14, and MYB15 DNA binding among the shikimate, stilbenoid, lignin, and flavonoid pathways.

Figure S9. Lack of anthocyanin pigmentation in jasmonate (MeJA) + cyclodextrins (CD)-elicited grapevine cell cultures, compared with control cells at day 7.

Figure S10. MYB13, MYB14, and MYB15 DNA binding with respect to differential gene expression of shikimate, early phenylpropanoid, and stilbenoid pathway genes during stage 2 noble rot infection of berries, reanalysed from Blanco-Ulate et al. (2015).

Figure S11. Microarray expression of shikimate, early phenylpropanoid, and stilbenoid pathway genes across the *Vitis vinifera* cv. 'Corvina' expression atlas (Fasoli et al., 2012).

Figure S12. Piceid (Pc) and pterostilbene (Pt) accumulation in tobacco leaves upon overexpression of STS48 and *ROMT1-3* genes.

Figure S13. JBrowse adaptation for the visualization of DAP-Seq data in grapevine (DAP-Browse App, available at <https://tomsbiolab.com/vitviz>).

Figure S14. Reproducibility and similarity between DAP-Seq replicates.

Dataset S1. VviMYB13/14/15 DAP-Seq peaks, Gene Set Enrichment Analysis (GSEA) tables, STS genes and housekeeping genes.

Dataset S2. RNA-Seq SRA runs for network construction, *MYB14* and *MYB15* GCNs, HCTs, GSEA tables, and VCost.v3 to MapMan Bin associations.

Dataset S3 Primers, *MYB15* overexpression cluster modules, affymetrix microarray dataset, microarray analysis (affymetrix probe-to-gene assignments and log₂ fold-changes) and updated MapMan ontology.

OPEN RESEARCH BADGES



This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE180450>.

DATA AVAILABILITY STATEMENT

DAP-Seq and microarray data can be found in the Gene Expression Omnibus (GEO) database of the NCBI under the accessions GSE180450 and GSE195646, respectively. All relevant data can be found within the manuscript and its supporting materials.

REFERENCES

- Almagro, L., Carbonell-Bejerano, P., Belchi-Navarro, S., Bru, R., Martínez-Zapater, J.M., Lijavetzky, D. et al. (2014) Dissecting the transcriptional response to elicitors in *Vitis vinifera* cells. *PLoS ONE*, **9**, e109777. <https://doi.org/10.1371/journal.pone.0109777>.
- Bai, R., Luo, Y., Wang, L., Li, J., Wu, K., Zhao, G. et al. (2019) A specific allele of MYB14 in grapevine correlates with high stilbene inducibility triggered by Al³⁺ and UV-C radiation. *Plant Cell Reports*, **38**, 37–49. <https://doi.org/10.1007/s00299-018-2347-9>.
- Ballouz, S., Weber, M., Pavlidis, P. & Gillis, J. (2016) EGAD: ultra-fast functional analysis of gene networks. *Bioinformatics*, **33**, 612–614. <https://doi.org/10.1093/bioinformatics/btw695>.
- Bartlett, A., O'Malley, R., Huang, S.-S., Galli, M., Nery, J., Gallavotti, A. et al. (2017) Mapping genome-wide transcription-factor binding sites

using dap-seq. *Nature Protocols*, **12**, 1659–1672. <https://doi.org/10.1038/nprot.2017.055>.

- Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological*, **57**, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x> <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x>.
- Blanco-Ulate, B., Amrine, K.C., Collins, T.S., Rivero, R.M., Vicente, A.R., Morales-Cruz, A. et al. (2015) Developmental and metabolic plasticity of white-skinned grape berries in response to botrytis cinerea during noble rot. *Plant Physiology*, **169**, 2422–2443. <https://doi.org/10.1104/pp.15.00852>.
- Bru, R., Sellés, S., Casado-Vela, J., Belchi-Navarro, S. & Pedreño, M.A. (2006) Modified cyclodextrins are chemically defined glucan inducers of defense responses in grapevine cell cultures. *Journal of Agricultural and Food Chemistry*, **54**, 66–71. <https://doi.org/10.1021/jf051485j>.
- Canaguier, A., Grimplet, J., Di Gaspero, G., Scalabrin, S., Duchêne, E., Choise, N. et al. (2017) A new version of the grapevine reference genome assembly (12x.v2) and of its annotation (vcost.v3). *Genomics Data*, **14**, 56–62. <https://doi.org/10.1016/j.gdata.2017.09.002> <http://www.sciencedirect.com/science/article/pii/S2213596017301459>.
- Carbonell-Bejerano, P., Rodríguez, V., Royo, C., Hernáiz, S., Moro-González, L.C., Torres-Viñals, M. et al. (2014) Circadian oscillatory transcriptional programs in grapevine ripening fruits. *BMC Plant Biology*, **14**, 1–15. <https://doi.org/10.1186/1471-2229-14-78>.
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Chen, Y., Zhangliang, C., Kang, J., Gu, H. & Qin, G. (2013) Atmyb14 regulates cold tolerance in arabidopsis. *Plant Molecular Biology Reporter*, **31**, 87–97. <https://doi.org/10.1007/s11105-012-0481-z>.
- Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A. et al. (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, **13**, 1050–1054. <https://doi.org/10.1038/nmeth.4035>.
- Davies, K.M., Jibrán, R., Zhou, Y., Albert, N.W., Brumell, D.A., Jordan, B.R. et al. (2020) The evolution of flavonoid biosynthesis: a bryophyte perspective. *Frontiers in Plant Science*, **11**, 7. <https://doi.org/10.3389/fpls.2020.00007>.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S. et al. (2012) STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics*, **29**, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Dubrovina, A.S. & Kiselev, K.V. (2017) Regulation of stilbene biosynthesis in plants. *Planta*, **246**, 597–623. <https://doi.org/10.1007/s00425-017-2730-8>.
- Fasoli, M., Dal Santo, S., Zenoni, S., Tornielli, G.B., Farina, L., Zamboni, A. et al. (2012) The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *Plant Cell*, **24**, 3489–3505. <https://doi.org/10.1105/tpc.112.100230>.
- Gillis, J. & Pavlidis, P. (2012) "Guilt by association" is the exception rather than the rule in gene networks. *PLoS Computational Biology*, **8**, 1–13. <https://doi.org/10.1371/journal.pcbi.1002444>.
- Guo, Y., Mahony, S. & Gifford, D.K. (2012) High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Computational Biology*, **8**, 1–14. <https://doi.org/10.1371/journal.pcbi.1002638>.
- Hall, D. & De Luca, V. (2007) Mesocarp localization of a bi-functional resveratrol/hydroxycinnamic acid glucosyltransferase of Concord grape (*Vitis labrusca*). *Plant Journal*, **49**, 579–591. <https://doi.org/10.1111/j.1365-3113.2006.02987.x>.
- Höll, J., Vannozzi, A., Czemmel, S., D'Onofrio, C., Walker, A., Rausch, T. et al. (2013) The r2r3-myb transcription factors myb14 and myb15 regulate stilbene biosynthesis in *vitis vinifera*. *The Plant Cell*, **25**, 4135–4149. <https://doi.org/10.1105/tpc.113.117127>.
- Huang, L., Yin, X., Sun, X., Yang, J., Rahman, M., Chen, Z. et al. (2018) Expression of a grape vqsts36-increased resistance to powdery mildew and osmotic stress in arabidopsis but enhanced susceptibility to botrytis cinerea in arabidopsis and tomato. *International Journal of Molecular Sciences*, **19**, 2985. <https://doi.org/10.3390/ijms19102985>.
- Hurtado-Gaitán, E., Sellés-Marchart, S., Martínez-Márquez, A., Samper-Herrero, A. & Bru-Martínez, R. (2017) A focused multiple reaction

- monitoring (MRM) quantitative method for bioactive grapevine stilbenes by ultra-high-performance liquid chromatography coupled to triple-quadrupole mass spectrometry (UHPLC-QqQ). *Molecules*, **22**, 418. <https://doi.org/10.3390/molecules22030418>.
- Jailon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A. et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467. <https://doi.org/10.1038/nature06148>.
- Jiang, C., Wang, D., Zhang, J., Xu, Y., Zhang, C., Zhang, J. et al. (2021) VqMYB154 promotes polygene expression and enhances resistance to pathogens in Chinese wild grapevine. *Horticulture Research*, **8**, 1–17. <https://doi.org/10.1038/s41438-021-00585-0>.
- Kelemen, Z., Sebastian, A., Xu, W., Grain, D., Salsac, F., Avon, A. 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**, 1–22. <https://doi.org/10.1371/journal.pone.0141044>.
- Kenrick, P. & Crane, P. (1997) The origin and early evolution of plants on land. *Nature*, **389**, 33–39. <https://doi.org/10.1038/37918>.
- Kim, S.H., Lam, P.Y., Lee, M.-H., Jeon, H.S., Tobimatsu, Y. & Park, O.K. (2020) The arabidopsis r2r3 myb transcription factor myb15 is a key regulator of lignin biosynthesis in effector-triggered immunity. *Frontiers in Plant Science*, **11**, 1456. <https://doi.org/10.3389/fpls.2020.583153> <https://www.frontiersin.org/article/10.3389/fpls.2020.583153>.
- Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. (2020) gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research*, **9**, 709. <https://doi.org/10.12688/f1000research.24956.2>.
- Langfelder, P. & Horvath, S. (2008) Wgcna: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 1–13.
- Langmead, B. & Salzberg, S. (2012) Langmead b, salzberg sl.. fast gapped-read alignment with bowtie 2. *Nature Methods*, **9**, 357–359. <https://doi.org/10.1038/nmeth.1923>.
- Lecourieux, F., Kappel, C., Pieri, P., Charon, J., Pillet, J., Hilbert, G. et al. (2017) Dissecting the biochemical and transcriptomic effects of a locally applied heat treatment on developing Cabernet Sauvignon grape berries. *Frontiers in Plant Science*, **8**, 53. <https://doi.org/10.3389/fpls.2017.00053>.
- Liao, Y., Smyth, G.K. & Shi, W. (2013) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
- Machanic, P. & Bailey, T. (2011) Meme-chip: motif analysis of large dna datasets. *Bioinformatics (Oxford, England)*, **27**, 1696–1697. <https://doi.org/10.1093/bioinformatics/btr189>.
- Martínez-Márquez, A., Morante-Carriel, J.A., Ramírez-Estrada, K., Cusidó, R.M., Palazon, J. & Bru-Martínez, R. (2016) Production of highly bioactive resveratrol analogues pterostilbene and piceatannol in metabolically engineered grapevine cell cultures. *Plant Biotechnology Journal*, **14**, 1813–1825. <https://doi.org/10.1111/pbi.12539>.
- Merz, P.R., Moser, T., Höll, J., Kortekamp, A., Buchholz, G., Zyprian, E. et al. (2015) The transcription factor VvWRKY33 is involved in the regulation of grapevine (*Vitis vinifera*) defense against the oomycete pathogen *Plasmopara viticola*. *Physiologia Plantarum*, **153**, 365–380. <https://doi.org/10.1111/ppl.12251>.
- Mutwil, M., Klie, S., Tohge, T., Giorgi, F.M., Wilkins, O., Campbell, M.M. et al. (2011) PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. *The Plant Cell*, **23**, 895–910. <https://doi.org/10.1105/tpc.111.083667>.
- Nakata, M. & Ohme-Takagi, M. (2014) Quantification of anthocyanin content. *Bio-Protocol*, **4**, e1098. <https://doi.org/10.21769/BioProtoc.1098>.
- Navarro-Payá D., Santiago A., Orduña L., Zhang C., Amato A., Fattorini C. et al. (2022) The grape gene reference catalogue as a standard resource for gene selection and genetic improvement. *Frontiers in Plant Science*, **12**, 803977. <https://doi.org/10.3389/fpls.2021.803977>.
- O'Malley, R., Huang, S.S.C., Song, L., Lewsey, M., Bartlett, A., Nery, J. et al. (2016) Cistrome and epistrome features shape the regulatory dna landscape. *Cell*, **165**, 1280–1292. <https://doi.org/10.1016/j.cell.2016.04.038> <http://www.sciencedirect.com/science/article/pii/S0092867416304810>.
- Parage, C., Tavares, R., Réty, S., Baltenweck-Guyot, R., Poutaraud, A., Renault, L. et al. (2012) Structural, functional, and evolutionary analysis of the unusually large stilbene synthase gene family in grapevine. *Plant Physiology*, **160**, 1407–1419. <https://doi.org/10.1104/pp.112.202705> <http://www.plantphysiol.org/content/160/3/1407>.
- Pasquereau, S., Nehme, Z., Haidar Ahmad, S., Daouad, F., Van Assche, J., Wallet, C. et al. (2021) Resveratrol inhibits HCoV-229E and SARS-CoV-2 coronavirus replication in vitro. *Viruses*, **13**, 1–11. <https://doi.org/10.3390/v13020354> <http://www.ncbi.nlm.nih.gov/pubmed/33672333>.
- Qian, C., Chen, Z., Liu, Q., Mao, W., Chen, Y., Tian, W. et al. (2020) Coordinated transcriptional regulation by the uv-b photoreceptor and multiple transcription factors for plant uv-b responses. *Molecular Plant*, **13**, 777–792. <https://doi.org/10.1016/j.molp.2020.02.015> <https://www.sciencedirect.com/science/article/pii/S1674205220300629>.
- Ramsak, Z., Baebler, S., Rotter, A., Korbar, M., Mozetič, I., Usadel, B. et al. (2014) GoMapMan: integration, consolidation and visualization of plant gene annotations within the MapMan ontology. *Nucleic Acids Research*, **42**, 1167–1175. <https://doi.org/10.1093/nar/gkt1056>.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. et al. (2015) limma powers differential expression analyses for RNA-Sequencing and microarray studies. *Nucleic Acids Research*, **43**, e47. <https://doi.org/10.1093/nar/gkv007>.
- Romero, I., Fuertes, A., Benito, M.J., Malpica, J.M., Leyva, A. & Paz-Ares, J. (1998) More than 80R2R3- MYB regulatory genes in the genome of Arabidopsis thaliana. *Plant Journal*, **14**, 273–284. <https://doi.org/10.1046/j.1365-313X.1998.00113.x>.
- Santos-Rosa, M., Poutaraud, A., Merdinoglu, D. & Mestre, P. (2008) Development of a transient expression system in grapevine via agro-infiltration. *Plant Cell Reports*, **27**, 1053–1063. <https://doi.org/10.1007/s00299-008-0531-z>.
- Schmidlin, L., Poutaraud, A., Claudel, P., Mestre, P., Prado, E., Santos-Rosa, M. et al. (2008) A stress-inducible resveratrol O- methyltransferase involved in the biosynthesis of pterostilbene in grapevine. *Plant Physiology*, **148**, 1630–1639. <https://doi.org/10.1104/pp.108.126003>.
- Schwacke, R., Ponce-Soto, G.Y., Krause, K., Bolger, A.M., Arsova, B., Hallab, A. et al. (2019) MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. *Molecular Plant*, **12**, 879–892. <https://doi.org/10.1016/j.molp.2019.01.003>.
- Shalit-Kaneh, A., Kumimoto, R.W., Filkov, V. & Harmer, S.L. (2018) Multiple feedback loops of the Arabidopsis circadian clock provide rhythmic robustness across environmental conditions. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, 7147–7152. <https://doi.org/10.1073/pnas.1805524115>.
- Tamagnone, L., Merida, A., Parr, A., Mackay, S., Cullanez-Macia, F.A., Roberts, K. et al. (1998) The AmMYB308 and AmMYB330 transcription factors from antirrhinum regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *The Plant Cell*, **10**, 135–154. <https://doi.org/10.1105/tpc.10.2.135>.
- Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. (2011) Differential expression in RNA-seq: a matter of depth. *Nucleic Acids Research*, **43**, e140. <https://doi.org/10.1101/gr.124321.111>.
- Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P. et al. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal*, **37**, 914–939. <https://doi.org/10.1111/j.1365-313X.2004.02016.x>.
- Vannozzi, A., Dry, I., Fasoli, M., Zenoni, S. & 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 Biology*, **12**, 8. <https://doi.org/10.1186/1471-2229-12-130>.
- Vannozzi, A., Wong, D.C.J., Höll, J., Himmam, I., Matus, J.T., Bogs, J. et al. (2018) Combinatorial Regulation of Stilbene Synthase Genes by WRKY and MYB Transcription Factors in Grapevine (*Vitis vinifera* L.). *Plant and Cell Physiology*, **59**, 1043–1059. <https://doi.org/10.1093/pcp/pcy045>.
- Verleyen, W., Ballouz, S. & Gillis, J. (2014) Measuring the wisdom of the crowds in network-based gene function inference. *Bioinformatics (Oxford, England)*, **31**, 10. <https://doi.org/10.1093/bioinformatics/btu715>.
- Wang, D., Jiang, C., Li, R. & Wang, Y. (2019) VqZIP1 isolated from Chinese wild *Vitis quinqueangularis* is involved in the ABA signaling pathway and regulates stilbene synthesis. *Plant Science*, **287**, 110202. <https://doi.org/10.1016/j.plantsci.2019.110202>.
- Wang, D., Jiang, C., Liu, W., Wang, Y. & Hancock, R. (2020) The WRKY53 transcription factor enhances stilbene synthesis and disease resistance by interacting with MYB14 and MYB15 in Chinese wild grape. *Journal of Experimental Botany*, **71**, 3211–3226. <https://doi.org/10.1093/jxb/era097>.
- Wang, L. & Wang, Y. (2019) Transcription factor VqERF114 regulates stilbene synthesis in Chinese wild *Vitis quinqueangularis* by interacting with

- VqMYB35. *Plant Cell Reports*, **38**, 1347–1360. <https://doi.org/10.1007/s00299-019-02456-4>.
- Waters, E.R.** (2003) Molecular adaptation and the origin of land plants. *Molecular Phylogenetics and Evolution*, **29**, 456–463. <https://doi.org/10.1016/j.ympev.2003.07.018> <http://www.sciencedirect.com/science/article/pii/S1055790303003130nn>.
- Wolfe, C.J., Kohane, I.S. & Butte, A.J.** (2005) Systematic survey reveals general applicability of 'guilt-by-association' within gene coexpression networks. *BMC Bioinformatics*, **6**, 1–10. <https://doi.org/10.1186/1471-2105-6-227>.
- Wong, D.** (2020) Network aggregation improves gene function prediction of grapevine gene co-expression networks. *Plant Molecular Biology*, **103**, 425–441. <https://doi.org/10.1007/s11103-020-01001-2>.
- Wong, D.C.J. & Matus, J.T.** (2017) Constructing integrated networks for identifying new secondary metabolic pathway regulators in grapevine: recent applications and future opportunities. *Frontiers in Plant Science*, **8**, 1–8. <https://doi.org/10.3389/fpls.2017.00505>.
- Wong, D.C.J., Schlechter, R., Vannozzi, A., Höll, J., Hmam, I., Bogs, J. et al.** (2016) A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation. *DNA Research*, **23**, 451–466. <https://doi.org/10.1093/dnares/dsw028>.
- Xi, H., Ma, L., Liu, G., Wang, N., Wang, J., Wang, L. et al.** (2014) Transcriptomic analysis of grape (*Vitis vinifera* L.) leaves after exposure to ultraviolet C irradiation. *PLoS ONE*, **9**, 1–24. <https://doi.org/10.1371/journal.pone.0113772>.
- Xu, W., Fuli, M., Li, R., Zhou, Q., Yao, W., Jiao, Y. et al.** (2019) Vpsts29/sts2 enhances fungal tolerance in grapevine through a positive feedback loop. *Plant, Cell & Environment*, **42**, 2979–2998. <https://doi.org/10.1111/pce.13600>.
- Yin, X., Singer, S.D., Qiao, H., Liu, Y., Jiao, C., Wang, H. et al.** (2016) Insights into the mechanisms underlying ultraviolet-c induced resveratrol metabolism in grapevine (*V. amurensis* Rupr.) cv. "Tonghua-3". *Frontiers in Plant Science*, **7**, 1–16. <https://doi.org/10.3389/fpls.2016.00503>.
- Zhu, L., Gazin, C., Lawson, N., Pagès, H., Lin, S., Lapointe, D. et al.** (2010) Chippeakanno: a bioconductor package to annotate chip-seq and chip-chip data. *BMC Bioinformatics*, **11**, 237. <https://doi.org/10.1186/1471-2105-11-237>.