bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Orchestration of the stilbene synthase gene family and their regulators by subgroup 2 MYB genes

Luis Orduña[1], Miaomiao Li[2], David Navarro-Payá[1], Chen Zhang[1], Živa Ramšak[3], Kristina Gruden[3], Janine Höll[4], Patrik Merz[4], Alessandro Vannozzi[5], Dario Cantu[6], Jochen Bogs[4], Darren C. J. Wong[7], Shao-shan Carol Huang[2], José Tomás Matus[1].

[1]Institute for Integrative Systems Biology, I2SysBio (University of Valencia-CSIC), Paterna, Valencia, Spain.

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

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

[4]Dienstleistungszentrum Ländlicher Raum Rheinpfalz, Viticulture and Enology Group, Neustadt/W, Germany.

[5]Department of Agronomy, Food, Natural resources, Animals, and Environment (DAFNAE), University of Padova, Legnaro 35020, Italy.

[6]Department of Viticulture and Enology, University of California Davis, Davis, California, USA.

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

1 Abstract

The control of plant specialised metabolism is exerted by transcription factors and co-regulators acting on *cis*-regulatory DNA sequences of pathway-structural genes, determining when, where, and how metabolites are accumulated. A particularly interesting case for studying the transcriptional control of metabolism is represented by stilbenoids, produced within the phenylpropanoid pathway, as their ability to inhibit infection by coronaviruses MERS-CoV and SARS-CoV has been recently demonstrated *in vitro*. Integrative omic studies in grapevine (*Vitis vinifera* L.), including gene co-expression networks, have previously highlighted several transcription factors (TFs) from different gene families as potential modulators of stilbenoid accumulation, offering an ideal framework for gene function characterisation using genome-wide approaches. In the context of non-model plant species, DNA affinity purification sequencing (DAP-Seq) results a novel and potentially powerful tool for the analysis of novel uncharacterised regulators, however, it has not yet been applied in fruit crops. Accordingly, we tested as a proof-of-concept the binding of two previously characterised R2R3-MYB TFs to their known targets of the stilbene pathway, MYB14 and MYB15, obtaining 5,222 and 4,502 binding events assigned to 4,038 and 3,645 genes for each TF, respectively. Bound genes (putative targets) were overlapped with aggregated gene centred co-expression networks resulting in shared and exclusive High Confidence Targets (HCTs) suggesting a high, but not complete, redundancy. Our results show that in addition to the previously known but few *STS* targets, these regulators bind to almost half of the complete *STS* family in addition to other phenylpropanoid- and stilbenoid-related genes. We also suggest they are potentially involved in other processes such as the circadian rhythm or the synthesis of biotin. We searched the activated transcriptomes of

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transiently *MYB15*-overexpressing grapevine plants and observed a large activation of its high confidence targets, validating our methodological approach. Our results also show that MYB15 seems to play a role in regulating other stilbenoid-related TFs such as WRKY03.

2 Introduction

The evolved complexity of plant specialised metabolism can be traced back to the colonisation of dry land by green algal ancestors (Waters, 2003). Since their origin, land plants have been accompanied by the continuous presence of a wide range of abiotic and biotic stresses such as UV radiation, desiccation, or unfavourable microbial communities that have led to the selective emergence of novel protective metabolites derived from primary metabolism (Kenrick and Crane, 1997). A clear example of this consequence is represented in the phenylpropanoid pathway responsible for synthesising a plethora of vital compounds such as lignans, flavonoids and stilbenes, all of which start their biosynthesis from the catabolism of the amino acids phenylalanine and tyrosine. This pathway is a major source of aromatic secondary metabolites in plants, with many of their roles concerning tolerance and adaptation to the above-mentioned stresses (e.g. anthocyanins protecting from excessive radiation). Whilst some branches of the pathway are ubiquitous in the plant kingdom (such as those producing the lignin polymer or flavonoids), others such as the stilbene pathway are restricted to a number of species across at least 50 unrelated families such as Vitaceae (e.g. *Vitis vinifera* L.) and Moraceae (e.g. *Morus alba*) (Dubrovina and Kiselev, 2017). Grapevine is not only an important crop species but an interesting model for studying the complexity of the stilbene pathway given the remarkable expansion of the stilbene synthase (*STS*) family in its genome, attributable to tandem gene and chromosome segmental duplications, reaching up to a total of 47 genes (Parage et al., 2012; Vannozzi et al., 2012).

Stilbenes are phytoalexins; small, lipophilic compounds with key roles in plant defence, which 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 specificity when dealing with different biotic 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 but increased susceptibility to *Botrytis cinerea* (Huang et al., 2018).

STS enzymes are direct competitors of chalcone synthases (CHS) for pathway precursors. Both STS and CHS are very closely related type-III polyketide synthases, generating a tetraketide intermediate from the condensation of p-coumaroyl-CoA with 3 molecules of malonyl-CoA that depending on the STS/CHS activity will generate resveratrol or naringenin chalcone, respectively, thus defining the entry point of the stilbene and flavonoid branches. Different stimuli have been shown to favour one branch over the other, e.g. UV-C irradiation and downy mildew infection promote *STS* gene expression whilst downregulating *CHS* expression (Vannozzi et al., 2012). Other enzymes from the stilbene pathway in grapevine result in a broad range of stilbenes such as pterostilbene, viniferins, piceid, and piceatannol which involve methoxylation, oligomerisation, glycosylation and hydroxylation, respectively. The enzymes catalysing these reactions are mostly uncharacterised in grapevine except for a resveratrol O-methyltransferase (ROMT) responsible for the production of pterostilbene (Schmidlin et al., 2008) and a resveratrol glycosyl transferase (Rs-GT), in *Vitis labrusca*, for the production of piceid (Hall and 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 thought to be mainly regulated at the transcriptional level through R2R3-MYB-type transcription factors (TFs). In particular, the R2R3-type MYB14 and MYB15 (subgroup 2 members) have been shown to specifically activate a few *STS* promoters (*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

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

stilbene accumulation (Wong et al., 2016) although it has not been yet validated *in planta*. Gene regulatory networks of TFs can be initially explored and interrogated by the use of GCNs. For instance, members of the AP2/ERF, bZIP and WRKY gene families have been suggested as regulators of *STS* expression through these approaches (Wong and Matus, 2017), however, experimental evidence of binding is necessary to prove regulatory causality. Nevertheless, the 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 characterisation of additional stilbene pathway regulators such as bZIP1, ERF114, MYB35A and WRKY53 in recent years (Wang et al., 2019; Wang and Wang, 2019; Vannozzi et al., 2018), proving their efficacy in hypothesis driven research for TF discovery. Very 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).

Amongst the identified *STS*) regulators, MYB14 and MYB15 have been considered to be upstream TFs (i.e. major modulators) in the regulatory cascade governing stilbenoid accumulation. The detailed characterisation of these two potential orchestrators is hence of great importance. In addition, no other processes controlled by these TFs have been identified. The following study combines genome-wide TF binding-site interrogation using DNA Affinity Purification (DAP)-seq, and aggregated weighted gene co-expression networks (aggWGCN) to lay out MYB14 and MYB15 cistrome landscapes and identify their complete repertoire of target genes. Our results suggest that both MYB14 and MYB15 simultaneously bind to regulatory elements in most members of the *STS*) gene family, other stilbenoid genes and WRKY regulators of stilbene synthesis.

3 Methods

3.1 DNA Affinity Purification followed by sequencing (DAP-Seq) experiments

3.1.1 Experimental setup

Genomic DNA (gDNA) was purified from young grapevine leaves of cv. 'Pinot Noir' (clone PN-94) according to (Chin et al., 2016). Genomic DNA library and DAP-Seq were performed following the protocol described in (Bartlett et al., 2017), with slight modifications. Briefly, the gDNA sample (800ng/library) was sonicated into 200 bp fragments on a Covaris Focus-ultrasonicator instrument, which underwent end-repair, A-tailing and adapter ligation to attach Illumina-compatible sequencing adapters to the DAP-Seq library. We verified successful adapter ligation by qPCR of the DAP-Seq library with primer sequences that anneal to the adapter, as well as gel electrophoresis of the sonicated genomic DNA and the DAP-Seq library. *MYB14* and *MYB15* were amplified from cv. 'Pinot Noir' and cloned into the pIX-HALO expression plasmid (TAIR Vector:6530264275) to create an expression vector that contained a HALO-tag at the N-terminal. Clones were verified by digestion with restriction enzyme Xho. The *MYB14/15*-expression vectors were used in a coupled transcription/translation system (Promega) to produce *MYB14/15* protein with the HaloTag-fusion. Production of MYB14/15 proteins was confirmed by western blot using an anti-HaloTag antibody (Promega). The protein expression reactions were mixed with HaloTag-ligand conjugated magnetic beads (Promega) to pull down the HaloTag-fused TF. The pulled down TFs were then exposed to the DAP-Seq library for MYB-DNA binding. 400 ng library was used in each DAP-Seq reaction. The bound DNA was then eluted and PCR amplified to generate sequencing libraries that were sequenced on an Illumina NextSeq 500 (sequencing of libraries was set at 30 million and 1x75bp single-end reads). As a negative control, we performed DAP-Seq with the pIX-HALO expression vector without any ORF inserted, accounting for possible non-specific DNA binding, as well as copy number variations at specific genomic loci. All experiments included two biological replicates.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.1.2 Setup of assembly and annotation files and bioinformatic pipeline for identifying TF-bound genes

The most updated genome assembly available to date in *Vitis vinifera* is the 12X.2, associated to the V1 and V.Cost annotation versions, containing 29,971 and 42,414 gene models, respectively [(Jaillon et al., 2007), (Canaguier et al., 2017)]. Although the V.Cost 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 the structural genes coding for enzymes of any secondary metabolic pathway (including the stilbenoid branch) by the MYB14 and MYB15 TFs, we merged the V1 and V.COST annotation versions to conduct the DAP-Seq bioinformatic pipeline. A new annotation file was created, containing all the V1 gene models allocated in the 12X.2 assembly except for the *STS* gene models, belonging to the V.Cost annotation version (designated as V1_COST_STS annotation, and available at http://tomsbiolab.com/scriptsandfiles).

DAP-Seq reads were mapped to the 12X.2 reference genome sequence (Canaguier et al., 2017) using bowtie2 (Langmead and Salzberg, 2012) version 2.0-beta7 with default parameters and post-processing of reads with multiple alignments, conducted with a home-made Python script. Peak detection was performed using GEM peak caller (Guo et al., 2012) version 3.4 with the 12X.2 genome assembly using the following parameters: "–relax –f SAM –k_min 6 –kmax 20 –k_seqs 600 –k_neg_dinu_shuffle" limited to nuclear chromosomes only. The biological replicates were analysed as multi-replicates with the GEM replicate mode. Peak summits called by GEM were associated with the closest gene model using the BioConductor package ChIPpeakAnno (Zhu et al., 2010) with default parameters (e.g. NearestLocation). For this association, the V1_COST_STS annotation version was given to ChIPpeakAnno. *De novo* motif discovery was performed using 200 bp sequences centred at GEM-identified binding events for the 600 most enriched peaks, as in (Bailey et al., 2009).

3.2 Aggregated Whole Genome Co-expression Network (aggWGCN) and generation of individual Gene-Centred Networks (aggGCCN)

52 different RNA-Seq experiments from fruit or flower transcriptome samples (sequenced with Illumina technology) were downloaded from the Sequence Read Archive (SRA) database. Experiments were manually inspected and filtered to keep those with more than 6 data sets (i.e.runs), also excluding those that were wrongly annotated or those that only included microRNAs, sRNAs and non-coding RNA sequencing data, keeping a total of 40 experiments, encompassing 868 runs (655 single-end and 213 paired-end runs, **Supplementary Table 1**).

Single and paired-end runs were trimmed separately using fastp (Chen et al., 2018) version 0.20.0, aligned with STAR (Dobin et al., 2012) version 2.7.3a. Raw counts were computed using FeatureCounts (Liao et al., 2013) version 2.0.0 and the 12X.2 V.Cost gene models. Each experiment was analysed individually in order to build a Highest Reciprocal Rank (HRR) matrix. Briefly, all the raw counts for each run of the experiment were summarised in a unique raw counts matrix, obtaining 40 raw count matrices. Raw counts were normalised to FPKMs, and genes that had less than 0.5 FPKMs in all runs of the experiment were removed. The Pearson's correlation coefficient (PCC) of each gene against the remaining 42,413 genes was then calculated and ranked according to descending PCC values. The ranked PCC values were used to compute HRRs between the TOP 1% remaining genes, using the following formula: $HRR(A, B) = max(rank(A,B), rank(B,A))$. To construct the aggWGCN, the frequency of co-expression interaction(s) present across individual HRR matrices was used as edge weights, and ranked in descending order, taking the TOP 1% frequency values for each gene in order to build the final matrix (network) with 32,857 genes.

Network functional connectivity (performance) across all given annotations and genes was assessed as in (Wong, 2020) by guilty by association (GBA) neighbour voting, a machine learning algorithm based on the GBA principle, which states that genes sharing common functions are often coordi-

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nately regulated across multiple experiments (Verleyen et al., 2014). The test was performed using *EGAD* R package (Ballouz et al., 2016) with default settings. The network was scored by the area under the receiver operator characteristic curve (AUROC) across MapMan BIN functional categories 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. The output from a homemade Python script, which selected the TOP 1% most highly co-expressed genes (the TOP 1% equals 328 genes) for any gene of interest, was used to generate the aggGCCNs.

3.3 Identification of High Confidence Targets (HCTs)

In order to identify potential targets of *MYB14* and *MYB15*, the lists of TF-bound genes obtained by DAP-Seq were overlapped with data extracted from the aggWGCN. Briefly, we first extracted all TF-bound genes for each MYB and checked if they had a positive relationship in the aggGCCNs. This network relationship was considered positive if either the DAP-Seq bound candidate gene was present in the *MYB* aggGCCN or the respective *MYB* gene was present in the aggGCCN of the candidate gene. Bound DAP-Seq genes for each MYB TF with a positive relationship in the bidirectional aggGCCNs were thus considered HCTs for that particular MYB TF. HCTs of both *MYB14* and *MYB15* were overlapped to obtain common HCTs. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses for individual and common HCTs were performed using the gprofiler2 package (Kolberg et al., 2020) to produce Manhattan plots for the enriched terms. P-values for each term were adjusted using the false discovery rate parameter and terms with adjusted p-values below a 0.05 threshold were considered significant.

3.4 Transcriptome analysis of *MYB15* transient overexpression in grapevine plantlets

Vitis vinifera L. plantlets were transfected with 35S::MYB15 constructs according to (Foria et al., 2020). 35S::GPF plants were used as negative control. RNA was extracted from two biological replicates and three times points after agroinfiltration: 24, 48 and 96 hours. RNA was used for microarray analysis using the Affymetrix *Vitis vinifera* Grape Array.

Affymetrix probe sequences were re-aligned using bowtie against the transcriptome of *Vitis vinifera* L. cv. Pinot Noir derived from the V3 annotation on the 12X.V2 genome assembly in order to improve original V1 probe-*to*-gene assignments. Since 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 1 mismatch, 70% with 2 mismatches and 80% with 3 mismatches. From a total of 16602 probes, 59.8% were positively assigned to a V3 gene model.

The fluorescence values were normalised by the RMA method using LIMMA (Ritchie et al., 2015) R package and transformed to log scale. Normalised values were then clustered applying the Weighted Gene Co-Expression Network Analysis (WGCNA) (Langfelder and Horvath, 2008) R package using the blockwiseModules with a soft-threshold power value of 6 (chosen from the scale independence graph shown in **Supplementary Figure 1A**), a deepSplit of 1 and a mergeCutHeight of 0.2, obtaining a total of 35 modules (**Supplementary Figure 1B**). To analyse the behaviour of the modules and compare it between them, they were visualised as a heatmap using the z-scores calculated individually for each biological replicate (n=2) and time point. The heatmap was constructed using kmeans clustering with the pheatmap (https://CRAN.R-project.org/package=pheatmap) R package. GO and KEGG enrichment analyses were carried out, as previously described for HCTs (**Section 3.3**), for the probe-assigned genes in the different modules.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4 Results

4.1 *MYB14* and *MYB15* bind upstream of *STS* gene transcription start sites

Our DAP-Seq analysis reported 5,222 and 4,502 TF binding events (i.e. peaks), that were assigned to 4,038 and 3,645 different genes for *MYB14* and *MYB15*, respectively (**Supplementary Table 2**). An initial inspection of all binding events showed that 30.93% and 32.07% of the peaks are present between -3 kb upstream and 1 kb downstream from the Transcription Start Sites (TSSs) of annotated genes for *MYB14* and *MYB15*, respectively, as shown in **Figure 1A**. *MYB14* and *MYB15* shared a total of 2,717 genes. In addition, the *de novo* motif discovery reported almost identical motifs for *MYB14* and *MYB15*, as shown in **Figure 1B**. This DNA binding motif is very similar to the AC-element motif (CACC[T/A]ACC) for *AtMYB15* described in *Arabidopsis thaliana* (Romero et al., 1998). The similar number of total peaks, a high proportion of shared genes, and a close to identical DNA binding motif suggests that both transcription factors bind similar DNA sequences and therefore could share many of their roles in the control of gene expression.

A closer look at the *STS* gene family revealed that both MYB14 and MYB15 TFs produced a clear DNA binding signal, predominantly located in the TSS proximal region, and not observed in the pHALO non-specific DNA binding control (**Figure 2**). The inclusion of housekeeping genes (**Supplementary Table 2**) showed no specific DNA binding of MYB14 or MYB15 TFs. Furthermore, the binding patterns across individual *STS* genes are remarkably similar for both TFs providing strong evidence that both do not only bind the promoters of *STS* genes but they do so in a similar manner. Both TFs show two distinct DNA binding enriched regions upstream of *STS* genes, a first region located at 400 bp upstream, present in most *STSs* and a second close to 2000 bp upstream that is exhibited only by a subgroup of *STSs*, suggesting gene-specific differences in their regulation.

4.2 Aggregated gene-centred co-expression networks (aggGCCNs) evidence shared connections of *MYB14* and *MYB15* with several stilbene pathway genes

Gene co-expression networks are based on the 'guilt-by-association' principle whereby correlation in expression implies biological association (Wolfe et al., 2005). These networks can be constructed by taking all the available public expression data and computing them as a single data set. However, it has been previously shown that these networks can improve their biological predictability if the persistence of a gene-*to*-gene expression correlation across independent transcriptomic experiments is taken into account rather than the correlation strength in the computed single data set. Deriving a network from different experimental data sets in this manner is known as the aggregation of whole genome co-expression networks (aggWGCNs). As described in (Wong, 2020), the aggregation of individual WGCNs across multiple experiments improves network performance in grapevine studies. We generated a condition-dependent (i.e. flower- and fruit-derived) aggWGCN and evaluated the effect of the aggregation. Network performance was assessed through the Area Under the Receiver Operator characteristic Curve (AUROC) measurements across iterative aggregation steps. Briefly, we generated ten randomly chosen pairs of experiments, and for each pair, an aggWGCN was computed and assessed, obtaining ten different AUROC values. The same process was repeated, but with groups formed by 4, 6, 8... randomly chosen experiments until evaluating the final aggWGCN with all the 40 experiments. The AUROC value improved consistently as more experiments were included in the iteration subset (**Supplementary Figure 2**), approaching a plateau of the AUROC value (AUROC = 0.76) near the total of 40 experiments selected in this study, suggesting that the number of experiments chosen was enough for establishing 'guilt-by-association' relationships between all annotated grapevine genes.

The analysis of the *MYB14* and *MYB15* aggregated gene-centred co-expression networks (i.e. the

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

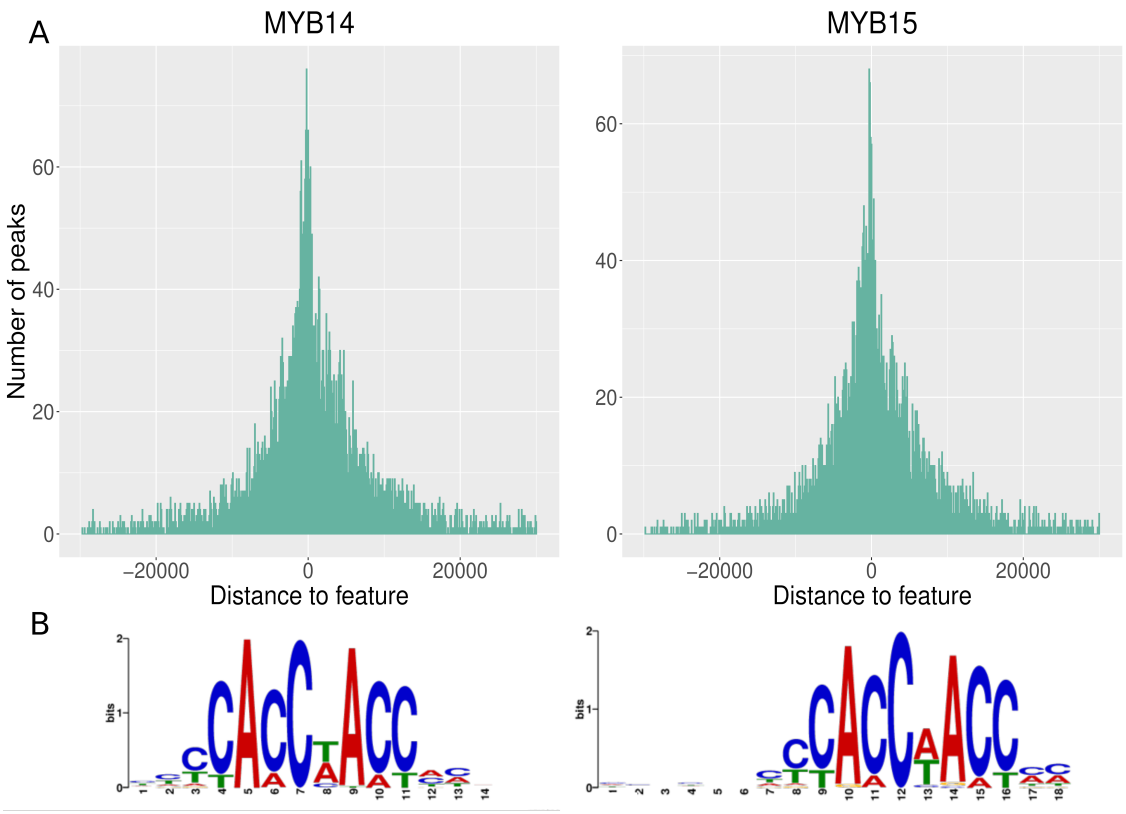


Figure 1: Transcription factor cistrome landscapes in cv. 'Pinot Noir' and binding motifs for MYB14 and MYB15, obtained by DAP-Seq. (A) Histogram of distances for all binding events (peaks) with respect to the transcriptional start site (TSS) of the gene they have been assigned to. (B) *De novo* binding motifs obtained from the TOP600 most highly-scored peaks of MYB14 and MYB15 as seen in the positive strand.

TOP 1% genes being co-expressed with a particular gene of interest; aggGCCNs) (**Figure 3A**, **Supplementary Table 1**) revealed that both networks shared 103 common genes out of the 328 genes present in each aggGCCN. Importantly, 20 *STS* genes were present in both aggGCCNs. In addition, 19 *STS* genes were exclusively present in the aggGCCN of *MYB14* meaning that almost all known *STS* genes in the grapevine genome, 39 out of 47, were present in either of the two constructed aggGCCNs. The exclusive presence of 19 *STS* genes in the MYB14 network points to a very interesting TF-specific relationship for a subset of *STS* genes which is not suggested by the DNA binding evidence presented above. This could point towards MYB14 specific co-regulators potentially conferring differences in the regulation of *STS* genes on behalf of MYB14 and MYB15 which are very similar TFs otherwise. Nonetheless, MYB14 could also have a more direct role in the regulation of *STS* genes than MYB15.

4.3 High Confidence Targets integrated from aggGCCNs and DAP-Seq

We further integrated both the co-expression and TF-binding results by overlapping the DAP-Seq genes with the MYB14/MYB15-centred aggGCCNs. The overlap between both data sets showed between 20% and 22% of co-expressed genes being supported by DAP-Seq data, representing potential target genes (**Figure 3A**). Despite this good level of support, we additionally inspected the aggGCCNs of DAP-Seq bound genes to look for MYB14 and MYB15 within these networks. Given the expected key regulatory role of both MYB-TFs in a high number of biologically relevant processes it is logical to expect that not all regulated genes appear within the TOP1% co-expression relationships which are used to build the aggGCCNs. We thus defined as High Confidence Targets (HCTs) those

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

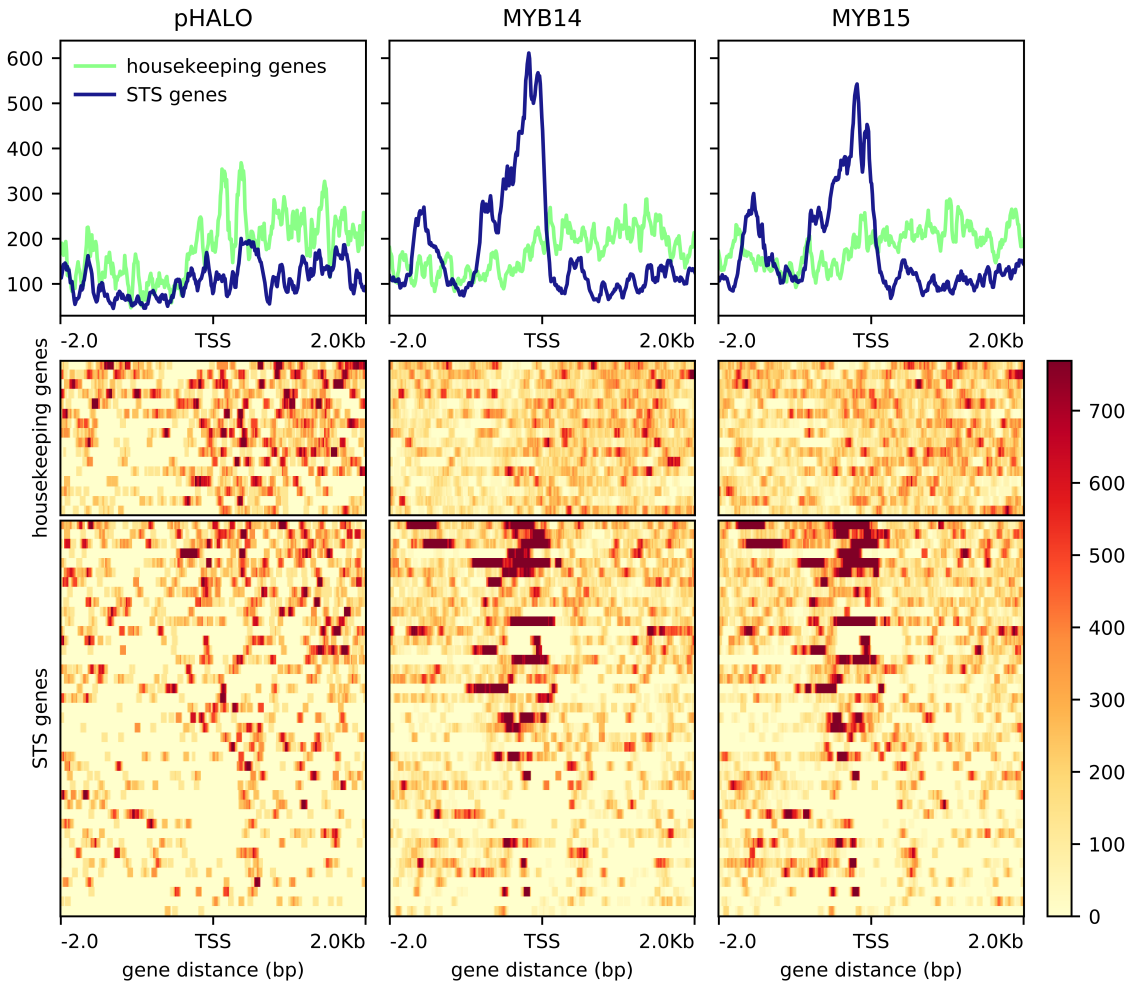


Figure 2: Preferential localisation of binding events in promoter regions of MYB14/MYB15 stilbene synthase *STS* putative targets. DAP-Seq binding signal for MYB14 and MYB15 at (-2kb,+2kb) from the TSS (x axis) in 22 *STS* genes, compared to background housekeeping genes.

bound genes with a co-expression relationship present in at least one of the two aggGCCNs (i.e. presenting a single or double network relationship). The results showed 170 and 181 HCT genes for MYB14 and MYB15, respectively, sharing 60 common HCTs. **Figure 3B** shows the complete *STS* gene family, including pseudogenes, following a colour scheme to illustrate the different levels of evidence of MYB14/MYB15 regulation provided by this study, for each individual gene. The individual and common HCT lists are provided in **Supplementary Table 3**. The common HCTs contained 9 different *STS* genes (*STS6*, *STS16*, *STS36*, *STS41*, *STS23*, *STS17*, *STS8*, *STS48*, *STS45*), confirming the importance and partial redundancy of both MYB14 and MYB15 in the regulation of stilbene synthesis. The two previously reported *STS* targets of MYB14/MYB15 (Höll et al., 2013) were present as HCTs; *STS41* being a common HCT, while *STS29* being only present as a MYB14 HCT. The manual inspection of MYB15-peaks in the PN40024 genome browser showed potential binding events in *STS29* that were not detected by our computational method (**Supplementary Data File 1**). This manual observation of peaks also reported potential binding sites for *STS3, STS9, STS27* and *STS46* as indicated by an asterisk in **Figure 3B**. These represent potential additional targets that cannot be demonstrated at least in this study.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

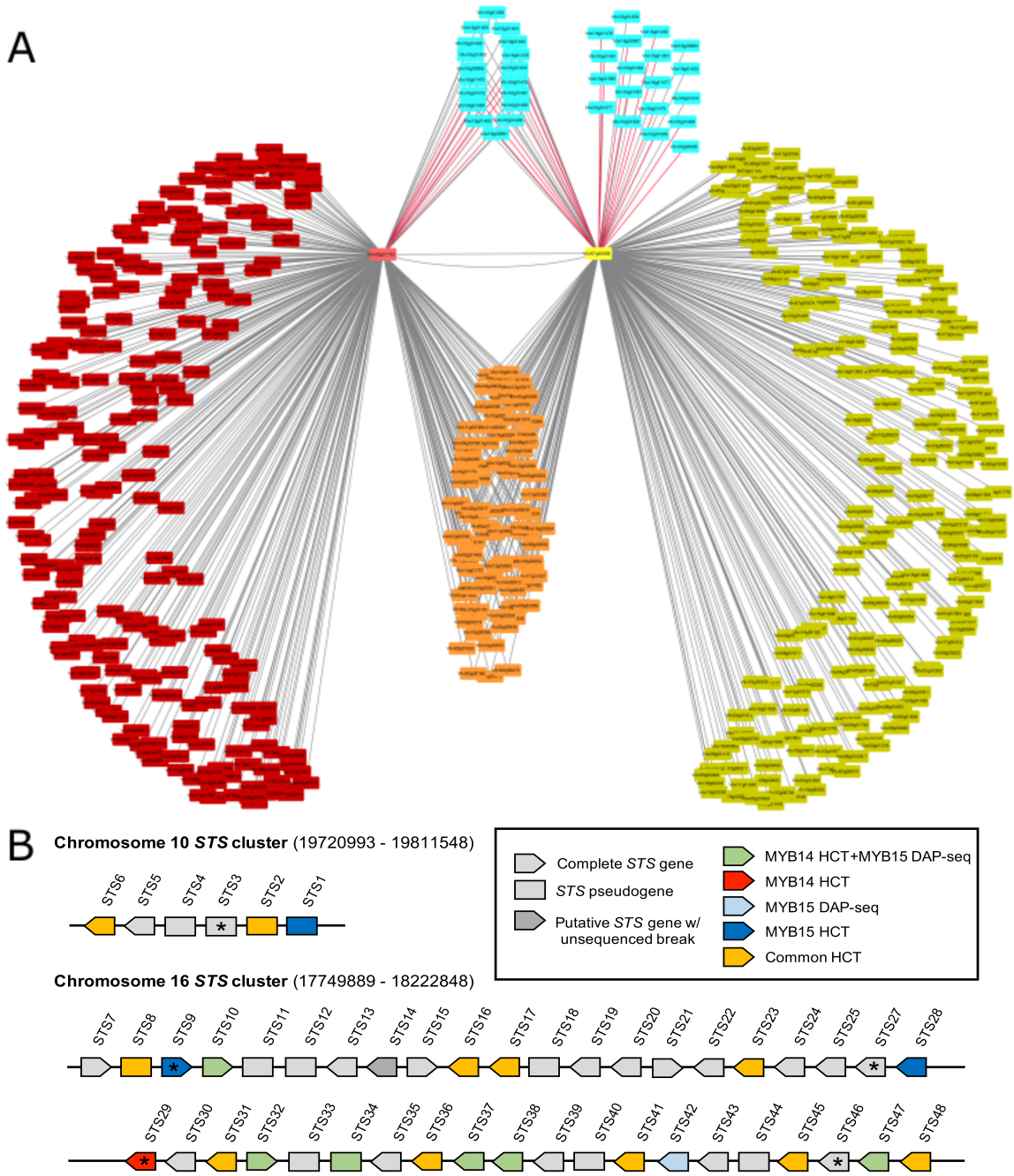


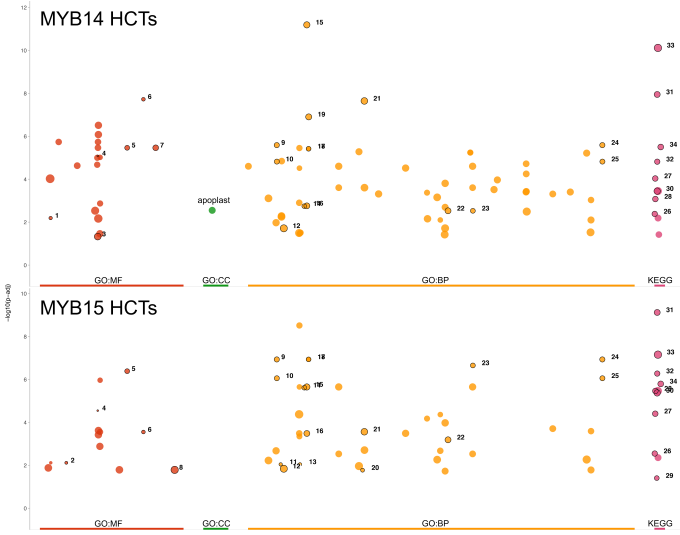
Figure 3: MYB14 and MYB15 aggGCCNs share many co-expressed genes overlapping with DAP-Seq experiments. (A) MYB14 (yellow) and MYB15 (red) aggGCCNs. Orange genes are the non-*STS* shared genes. The *STS* genes are coloured blue. The red edges are the *STS* genes that are also present in the DAP-Seq experiments. (B) Synteny of the *STS* genes along chromosome 10 and 16 *STS* clusters. The shape of each *STS* gene informs about its functionality, while its colour depicts the type of evidence supporting each regulation by MYB14 and MYB15. Diagram adapted from (Parage et al., 2012).

Several *PAL* genes appear as HCTs in addition to *STSs*, supporting the idea that MYB14 and MYB15 may have the ability to increase the carbon flux into the phenylpropanoid pathway.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

This may have indirect consequences in other phenylpropanoid branches such as those producing flavonols or anthocyanins. Three *WRKY* transcription factor genes are present among the common MYB14/MYB15 HCTs. WRKY TFs have been suggested as co-regulators of MYB action in the control of *STS* expression [(Wang et al., 2020), (Xi et al., 2014), (Xu et al., 2019)]. In fact, *WRKY03* is bound by both TFs and supported in all three MYB14-, MYB15- and WRKY03-aggGCCNs. This observation adds one more layer of complexity to the hierarchical regulation of MYB14, which requires *WRKY03* to increase *STS29* expression (Vannozzi et al., 2018).

Two independent GO and KEGG enrichment analyses were carried out for each TF-HCT list. The results of both enrichment analyses (**Figure 4 and Supplementary Table 3**) show strong term enrichment for a number of pathways and molecular functions of interest amongst which the shikimate, phenylpropanoid and stilbenoid pathway terms stand out as highly relevant.



id	source	term_id	term_name	term_size	pvalue MYB14	pvalue MYB15
1	GO:MF	GO:0003849	3–deoxy–7–phosphoheptulonate synthase activity	5	6.4e–03	NA
2	GO:MF	GO:0004764	shikimate 3–dehydrogenase (NADP+) activity	4	NA	7.6e–03
3	GO:MF	GO:0016709	oxidoreductase activity, acting on paired donors	294	4.7e–02	NA
4	GO:MF	GO:0016710	trans–cinnamate 4–monooxygenase activity	3	8.1e–06	2.8e–05
5	GO:MF	GO:0045548	phenylalanine ammonia–lyase activity	18	3.4e–06	4.1e–07
6	GO:MF	GO:0050350	trihydroxystilbene synthase activity	6	1.9e–08	2.8e–04
7	GO:MF	GO:0052716	hydroquinone:oxygen oxidoreductase activity	57	3.4e–06	NA
8	GO:MF	GO:0140110	transcription regulator activity	961	NA	1.6e–02
9	GO:BP	GO:0006558	L–phenylalanine metabolic process	27	2.6e–06	1.2e–07
10	GO:BP	GO:0006559	L–phenylalanine catabolic process	23	1.5e–05	8.7e–07
11	GO:BP	GO:0006768	biotin metabolic process	5	NA	9.0e–03
12	GO:BP	GO:0006952	defense response	570	1.9e–02	1.4e–02
13	GO:BP	GO:0009102	biotin biosynthetic process	5	NA	9.0e–03
14	GO:BP	GO:0009423	chorismate biosynthetic process	15	1.8e–03	2.4e–06
15	GO:BP	GO:0009698	phenylpropanoid metabolic process	130	6.4e–12	2.2e–06
16	GO:BP	GO:0009699	phenylpropanoid biosynthetic process	71	1.7e–03	3.3e–04
17	GO:BP	GO:0009800	cinnamic acid biosynthetic process	16	3.8e–06	1.2e–07
18	GO:BP	GO:0009803	cinnamic acid metabolic process	16	3.8e–06	1.2e–07
19	GO:BP	GO:0009808	lignin metabolic process	92	1.2e–07	NA
20	GO:BP	GO:0019632	shikimate metabolic process	7	NA	1.7e–02
21	GO:BP	GO:0019748	secondary metabolic process	238	2.2e–08	2.7e–04
22	GO:BP	GO:0044550	secondary metabolite biosynthetic process	81	2.2e–03	6.4e–04
23	GO:BP	GO:0046417	chorismate metabolic process	18	2.9e–03	2.2e–07
24	GO:BP	GO:1902221	erythrose 4–phosphate/phosphoenolpyruvate family amino acid metabolic	27	2.6e–06	1.2e–07
25	GO:BP	GO:1902222	erythrose 4–phosphate/phosphoenolpyruvate family amino acid catabolic	23	1.5e–05	8.7e–07
26	KEGG	KEGG.00130	Ubiquinone and other terpenoid–quinone biosynthesis	38	4.1e–03	2.8e–03
27	KEGG	KEGG.00360	Phenylalanine metabolism	46	9.2e–05	3.9e–05
28	KEGG	KEGG.00400	Phenylalanine, tyrosine and tryptophan biosynthesis	46	8.3e–04	3.4e–06
29	KEGG	KEGG.00780	Biotin metabolism	16	NA	3.9e–02
30	KEGG	KEGG.00940	Phenylpropanoid biosynthesis	159	3.5e–04	4.2e–06
31	KEGG	KEGG.00941	Flavonoid biosynthesis	78	1.1e–08	7.5e–10
32	KEGG	KEGG.00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	33	1.5e–05	5.3e–07
33	KEGG	KEGG.01110	Biosynthesis of secondary metabolites	1130	7.6e–11	6.9e–08
34	KEGG	KEGG.04712	Circadian rhythm – plant	59	3.1e–06	1.6e–06

Figure 4: MYB14 and MYB15 HCTs are enriched in stilbene pathway ontologies/bins. GO and KEGG enrichment analysis results for both MYB14 and MYB15 HCTs.

Interestingly, a number of unexpected terms appeared in the analysis, such as circadian rythm for both MYBs or biotin metabolism specifically for MYB15, pointing towards new aspects to consider in the study of MYB14 and MYB15 transcriptional regulation. On the other hand, there is a notable absence of O-methyltransferase-like terms given that the resveratrol O-methyltransferase enzyme ROMT1, part of the stilbenoid pathway, was found amongst the HCTs of MYB15 and has been previously shown to be involved in the synthesis of pterostilbene (Schmidlin et al., 2008). This absence from the enrichment analysis for MYB15 could be due to the fact that only one gene pertaining to this class of enzymes may not have been enough to show statistical significance despite the biological relevance. Nonetheless, the enrichment analysis shows many terms regarding shikimate or phenylpropanoid pathway enzymes such as phenylalanine ammonia-lyase (PAL) activity which refers to the first step of the phenylpropanoid pathway (i.e. the conversion of phenylalanine to cinnamate), which is enriched in both TFs. The involvement of MYB14 and MYB15 in the early steps of the phenylpropanoid pathway has not been previously reported and is hence of high interest given that these steps precede stilbenoid as well as lignin and flavonoid biosynthesis. These terms also appear in the enrichment analysis although not always for both TFs pointing towards interesting role specificity as well as redundancy for these two regulators. Lignin biosynthesis appears to be linked to *MYB14* exclusively whilst shikimate 3-dehydrogenase activity is specifically associated to *MYB15*. In addition, *MYB15* is specifically linked to the terms biotin metabolic process and biotin metabolism which could be interesting as biotin is an essential cofactor for a small number of enzymes involved in carboxylation and decarboxylation reactions (Alban et al., 2000).

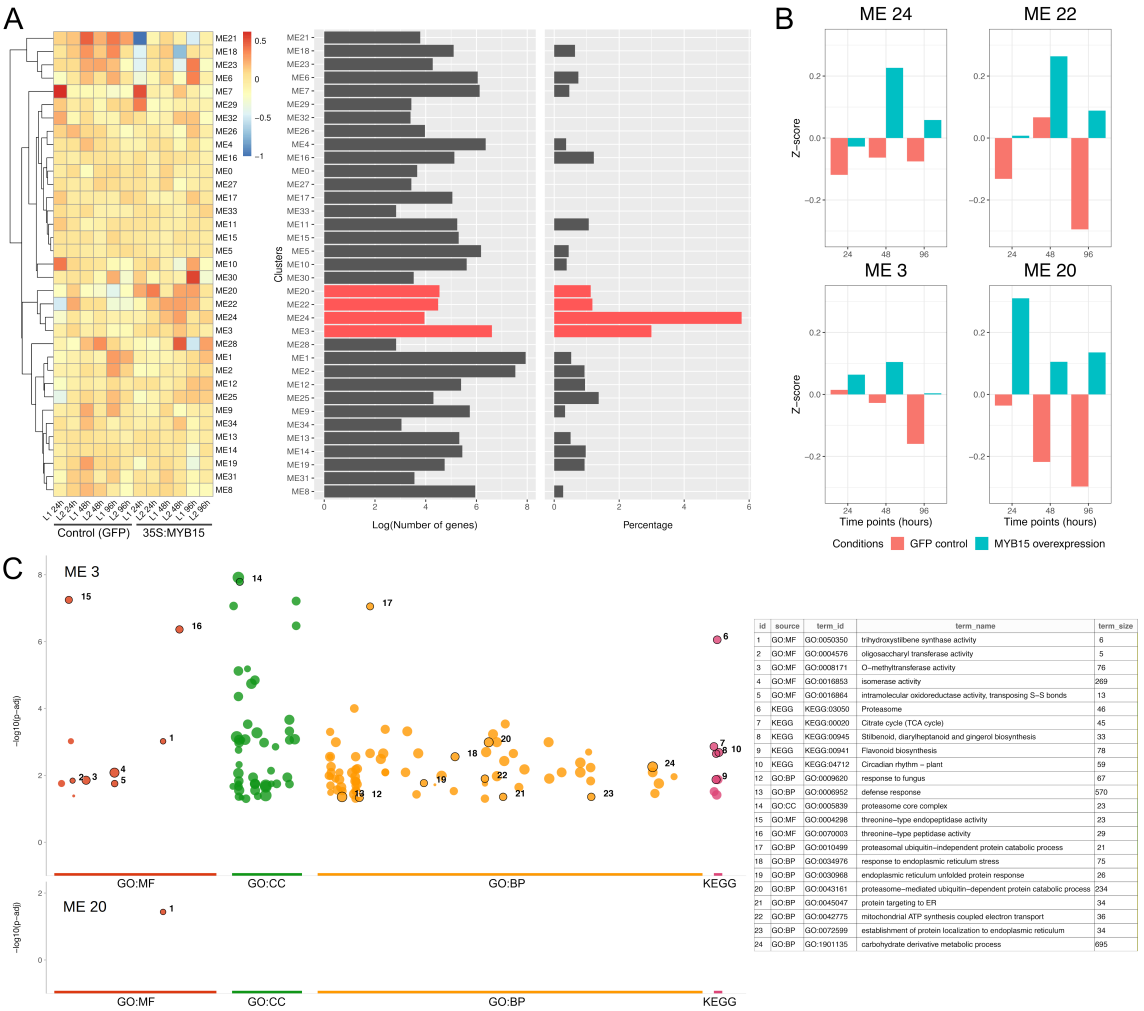
bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Other unexpected and yet interesting terms also appear such as the term for circadian rhythm for which no related roles have been previously described regarding *MYB14* or *MYB15*. However, LHY- and RVE8-type MYBs, different to the R2R3-type MYBs, have been previously shown to have roles regarding the control of circadian rhythm in *Arabidopsis thaliana* (Shalit-Kaneh et al., 2018). Altogether, this analysis suggests that the MYB14 and MYB15 TFs have active roles in the regulation of the stilbenoids, as expected, whilst also opening up the possibility of involvement in upstream pathways such as the early phenylpropanoid or shikimate routes and/or other biological processes unknown to date.

4.4 *MYB15* HCTs are supported by its transient over-expression in grapevine

We tested the prediction of MYB15 high confidence targets by inspecting them in the transiently-activated transcriptomes of MYB15 over-expressing grapevine plants. Microarray analysis (MA) was conducted at 24, 48 and 96 hours after *in planta* agroinfiltration with a 35S:MYB15 construct, and compared to a 35S:GFP control. Fluorescence microarray values were normalised and clustered using the WGCNA R package. The resulting 35 modules were inspected and visualised on a heatmap by independently computing their mean z-scores for each line and time point. MYB15 probe was assigned to module ME20, showing higher expression in 35S:MYB15 lines compared to controls. ME20 was clustered with three additional modules (modules ME22, ME24 and ME3 (**Figure 5A, left panel**). By inspecting both the total number of genes and presence of HCTs on each module, we suggest these four modules as potentially holding MYB15 real targets (**Figure 5A, middle and right panels, highlighted in red**). The expression behaviour of these four modules was plotted for each condition (**Figure 5B**), confirming that across all the considered time points they all presented higher gene expression in *MYB15* over-expressed lines compared to GFP control lines. GO and KEGG enrichment analysis was carried out for each module, showing that only ME3 and ME20 possessed significantly enriched terms (modules ME22 and ME24 have fewer genes thus probably affecting GO enrichment analysis). Modules ME3 and ME20 show both the term GO:0050350 (trihidroxystilbene synthase activity) as enriched (**Figure 5C**), supporting the role of *MYB15* in the regulation of *STS* genes. Whilst module ME20, with a total of 84 genes, only shows significant enrichment for this term; ME3, with a total of 650 genes, shows a wide range of term associations. The KEGG:00945 term for stilbenoid biosynthesis which again supports the stilbenoid regulatory roles of *MYB15*. Interestingly, some of the terms seem to be related to the stilbenoid pathway such as O-methyltransferase activity which could refer to enzymes catalysing the conversion of resveratrol to pterostilbene (Schmidlin et al., 2008). Similarly the term oligosaccharyl transferase activity could be referring to the glycosylation of resveratrol into piceid (Härtl et al., 2017).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



id	source	term_id	term_name	term_size
1	GO:MF	GO:0050350	trihydroxystilbene synthase activity	6
2	GO:MF	GO:0004576	oligosaccharyl transferase activity	5
3	GO:MF	GO:0008171	O−methyltransferase activity	76
4	GO:MF	GO:0016853	isomerase activity	269
5	GO:MF	GO:0016864	intramolecular oxidoreductase activity, transposing S−S bonds	13
6	KEGG	KEGG:03050	Proteasome	46
7	KEGG	KEGG:00020	Citrate cycle (TCA cycle)	45
8	KEGG	KEGG:00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	33
9	KEGG	KEGG:00941	Flavonoid biosynthesis	78
10	KEGG	KEGG:04712	Circadian rhythm − plant	59
12	GO:BP	GO:0009620	response to fungus	67
13	GO:BP	GO:0006952	defense response	570
14	GO:CC	GO:0005839	proteasome core complex	23
15	GO:MF	GO:0004298	threonine−type endopeptidase activity	23
16	GO:MF	GO:0070003	threonine−type peptidase activity	29
17	GO:BP	GO:0010499	proteasomal ubiquitin−independent protein catabolic process	21
18	GO:BP	GO:0034976	response to endoplasmic reticulum stress	75
19	GO:BP	GO:0030968	endoplasmic reticulum unfolded protein response	26
20	GO:BP	GO:0043161	proteasome−mediated ubiquitin−dependent protein catabolic process	234
21	GO:BP	GO:0045047	protein targeting to ER	34
22	GO:BP	GO:0042775	mitochondrial ATP synthesis coupled electron transport	36
23	GO:BP	GO:0072599	establishment of protein localization to endoplasmic reticulum	34
24	GO:BP	GO:1901135	carbohydrate derivative metabolic process	695

Figure 5: Stilbene pathway genes are enriched in MYB15 and in MYB15-related modules. (A, left panel) Heatmap of the mean of z-scores for each time point and line. (A, central panel) Barplot of the numer of genes per module. (A, right panel) Percentage of MYB15 HCTs present in each module. (B) Barplot of mean z-scores across the different plant lines in the different time points for cluster 24, 22, 3 and 20. (C) GO and KEGG enrichment analysis results for modules 20 and 3.

5 Funding and Acknowledgements

This work was supported by Grant PGC2018-099449-A-I00 and by the Ramón y Cajal program grant RYC-2017-23645, both awarded to J.T.M. and to the FPI scholarship PRE2019-088044 granted to L.O. 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). C.Z. is supported by China Scholarship Council (CSC) no. 201906300087. This article is based upon work from COST Action CA 17111 INTEGRAPE, supported by COST (European Cooperation in Science and Technology). Data has been treated and uploaded in public repositories according to the FAIR principles.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6 Supplementary Material

Supplementary Table 1. Experimental runs used for the aggWGCN and MYB14/15 gene-centred aggGCCNs.

Supplementary Table 2. DAP-Seq binding peak calls for MYB14 and MYB15 and list of housekeeping genes.

Supplementary Table 3. MYB14 and MYB15 individual and common HCTs, with their corresponding GO enrichment results.

Supplementary Table 4. MYB15 overexpression modules with each affymetrix probe-*to*-gene assignments and microarray log-normalized expression data.

Supplementary Figures. SupFig1. WGCNA R package output. SupFig2. AUROC validation of the aggWGCN. SupFig3. GO and KEGG enrichment analysis for MYB14 and MYB15 common HCTs.

Supplementary Data File 1. Screenshots of genome browser (JBrowse) with DAP-Seq peaks for MYB14 and MYB15 bound genes related to stilbenoid, including putative and common high confidence targets (HCTs). Replicates are shown separately for MYB-TF and pHALO-Control samples.

Supplementary Data File 2. Manhattan plots using gprofiler2 derived from the GO and KEGG enrichment analysis for each MYB15-responsive WGCNA co-expression module.

References

C. Alban, D. Job, and R. Douce. Biotin metabolism in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 51:17–47, 2000.

T. Bailey, M. Bodén, F. Buske, M. Frith, C. Grant, L. Clementi, J. Ren, W. Li, and W. Noble. Meme suite: tools for motif discovery and searching. *Nucleic acids research*, 37:W202–8, 06 2009. doi: 10.1093/nar/gkp335.

S. Ballouz, M. Weber, P. Pavlidis, and J. Gillis. EGAD: ultra-fast functional analysis of gene networks. *Bioinformatics*, 33(4):612–614, 11 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btw695. URL `https://doi.org/10.1093/bioinformatics/btw695`.

A. Bartlett, R. O'Malley, S.-S. Huang, M. Galli, J. Nery, A. Gallavotti, and J. Ecker. Mapping genome-wide transcription-factor binding sites using dap-seq. *Nature Protocols*, 12:1659–1672, 08 2017. doi: 10.1038/nprot.2017.055.

A. Canaguier, J. Grimplet, G. Di Gaspero, S. Scalabrin, E. Duchêne, N. Choisne, N. Mohellibi, C. Guichard, S. Rombauts, I. Le Clainche, A. Bérard, A. Chauveau, R. Bounon, C. Rustenholz, M. Morgante, M.-C. Le Paslier, D. Brunel, and A.-F. Adam-Blondon. A new version of the grapevine reference genome assembly (12x.v2) and of its annotation (vcost.v3). *Genomics Data*, 14:56 – 62, 2017. ISSN 2213-5960. doi: https://doi.org/10.1016/j.gdata.2017.09.002. URL `http://www.sciencedirect.com/science/article/pii/S2213596017301459`.

S. Chen, Y. Zhou, Y. Chen, and J. Gu. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17):i884–i890, 09 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty560. URL `https://doi.org/10.1093/bioinformatics/bty560`.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

C. S. Chin, P. Peluso, F. J. Sedlazeck, M. Nattestad, G. T. Concepcion, A. Clum, C. Dunn, R. O'Malley, R. Figueroa-Balderas, A. Morales-Cruz, G. R. Cramer, M. Delledonne, C. Luo, J. R. Ecker, D. Cantu, D. R. Rank, and M. C. Schatz. Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, 13(12):1050–1054, 2016. ISSN 15487105. doi: 10.1038/nmeth.4035. URL `http://dx.doi.org/10.1038/nmeth.4035`.

A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, 10 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts635. URL `https://doi.org/10.1093/bioinformatics/bts635`.

A. S. Dubrovina and K. V. Kiselev. Regulation of stilbene biosynthesis in plants. *Planta*, 246(4): 597–623, 2017. ISSN 14322048. doi: 10.1007/s00425-017-2730-8.

S. Foria, D. Copetti, B. Eisenmann, G. Magris, M. Vidotto, S. Scalabrin, R. Testolin, G. Cipriani, S. Wiedemann-Merdinoglu, J. Bogs, G. Di Gaspero, and M. Morgante. Gene duplication and transposition of mobile elements drive evolution of the rpv3 resistance locus in grapevine. *The Plant Journal*, 101(3):529–542, 2020. doi: https://doi.org/10.1111/tpj.14551. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.14551`.

Y. Guo, S. Mahony, and D. K. Gifford. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLOS Computational Biology*, 8(8):1–14, 08 2012. doi: 10.1371/journal.pcbi.1002638. URL `https://doi.org/10.1371/journal.pcbi.1002638`.

D. Hall and V. De Luca. Mesocarp localization of a bi-functional resveratrol/hydroxycinnamic acid glucosyltransferase of Concord grape (Vitis labrusca). *Plant Journal*, 49(4):579–591, 2007. ISSN 09607412. doi: 10.1111/j.1365-313X.2006.02987.x.

K. Härtl, F. C. Huang, A. P. Giri, K. Franz-Oberdorf, J. Frotscher, Y. Shao, T. Hoffmann, and W. Schwab. Glucosylation of Smoke-Derived Volatiles in Grapevine (Vitis vinifera) is Catalyzed by a Promiscuous Resveratrol/Guaiacol Glucosyltransferase. *Journal of Agricultural and Food Chemistry*, 65(28):5681–5689, 2017. ISSN 15205118. doi: 10.1021/acs.jafc.7b01886.

L. Huang, X. Yin, X. Sun, J. Yang, M. Rahman, Z. Chen, and X. Wang. 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, 09 2018. doi: 10.3390/ijms19102985.

J. Höll, A. Vannozzi, S. Czemmel, C. D'Onofrio, A. Walker, T. Rausch, M. Lucchin, P. Boss, I. Dry, and J. Bogs. The r2r3-myb transcription factors myb14 and myb15 regulate stilbene biosynthesis in vitis vinifera. *The Plant cell*, 25, 10 2013. doi: 10.1105/tpc.113.117127.

O. Jaillon, J.-M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Hugueney, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyère, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C. Del Fabbro, M. Alaux, G. Di Gaspero, V. Dumas, N. Felice, S. Paillard, I. Juman, M. Moroldo, S. Scalabrin, A. Canaguier, I. Le Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M. E. Pè, G. Valle, M. Morgante, M. Caboche, A.-F. Adam-Blondon, J. Weissenbach, F. Quétier, P. Wincker, and French-Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449(7161):463—467, September 2007. ISSN 0028-0836. doi: 10.1038/nature06148. URL `https://doi.org/10.1038/nature06148`.

P. Kenrick and P. Crane. The origin and early evolution of plants on land. *Nature*, 389:33–39, 09 1997. doi: 10.1038/37918.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

L. Kolberg, U. Raudvere, I. Kuzmin, J. Vilo, and H. Peterson. gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research*, 9:709, 2020. doi: 10.12688/f1000research.24956.2.

P. Langfelder and S. Horvath. Wgcna: an r package for weighted correlation network analysis. *BMC Bioinformatics*, 9:1–13, 01 2008.

B. Langmead and S. Salzberg. Langmead b, salzberg sl.. fast gapped-read alignment with bowtie 2. nat methods 9: 357-359. *Nature methods*, 9:357–9, 03 2012. doi: 10.1038/nmeth.1923.

Y. Liao, G. K. Smyth, and W. Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, 11 2013. ISSN 1367-4803. doi: 10.1093/bioinformatics/btt656. URL `https://doi.org/10.1093/bioinformatics/btt656`.

C. Parage, R. Tavares, S. Réty, R. Baltenweck-Guyot, A. Poutaraud, L. Renault, D. Heintz, R. Lugan, G. A. Marais, S. Aubourg, and P. Hugueney. Structural, functional, and evolutionary analysis of the unusually large stilbene synthase gene family in grapevine. *Plant Physiology*, 160(3):1407–1419, 2012. ISSN 0032-0889. doi: 10.1104/pp.112.202705. URL `http://www.plantphysiol.org/content/160/3/1407`.

M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47–e47, 01 2015. ISSN 0305-1048. doi: 10.1093/nar/gkv007. URL `https://doi.org/10.1093/nar/gkv007`.

I. Romero, A. Fuertes, M. J. Benito, J. M. Malpica, A. Leyva, and J. Paz-Ares. More than 80R2R3-MYB regulatory genes in the genome of Arabidopsis thaliana. *Plant Journal*, 14(3):273–284, 1998. ISSN 09607412. doi: 10.1046/j.1365-313X.1998.00113.x.

L. Schmidlin, A. Poutaraud, P. Claudel, P. Mestre, E. Prado, M. Santos-Rosa, S. Wiedemann-Merdinoglu, F. Karst, D. Merdinoglu, and P. Hugueney. A stress-inducible resveratrol O-methyltransferase involved in the biosynthesis of pterostilbene in grapevine. *Plant Physiology*, 148(3):1630–1639, 2008. ISSN 00320889. doi: 10.1104/pp.108.126003.

A. Shalit-Kaneh, R. W. Kumimoto, V. Filkov, and S. L. Harmer. 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(27):7147–7152, 2018. ISSN 10916490. doi: 10.1073/pnas.1805524115.

A. Vannozzi, I. Dry, M. Fasoli, S. Zenoni, and M. Lucchin. Genome-wide analysis of the grapevine stilbene synthase multigenic family: Genomic organization and expression profiles upon biotic and abiotic stresses. *BMC plant biology*, 12:130, 08 2012. doi: 10.1186/1471-2229-12-130.

A. Vannozzi, D. C. J. Wong, J. Höll, I. Hmmam, J. T. Matus, J. Bogs, T. Ziegler, I. Dry, G. Barcaccia, and M. Lucchin. Combinatorial Regulation of Stilbene Synthase Genes by WRKY and MYB Transcription Factors in Grapevine (Vitis vinifera L.). *Plant and Cell Physiology*, 59(5):1043–1059, 2018. ISSN 14719053. doi: 10.1093/pcp/pcy045.

W. Verleyen, S. Ballouz, and J. Gillis. Measuring the wisdom of the crowds in network-based gene function inference. *Bioinformatics (Oxford, England)*, 31, 10 2014. doi: 10.1093/bioinformatics/btu715.

D. Wang, C. Jiang, R. Li, and Y. Wang. VqbZIP1 isolated from Chinese wild Vitis quinquangularis is involved in the ABA signaling pathway and regulates stilbene synthesis. *Plant Science*, 287(July):110202, 2019. ISSN 18732259. doi: 10.1016/j.plantsci.2019.110202. URL `https://doi.org/10.1016/j.plantsci.2019.110202`.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.31.424746; this version posted January 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D. Wang, C. Jiang, W. Liu, Y. Wang, and R. Hancock. 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(10):3211–3226, 2020. ISSN 14602431. doi: 10.1093/jxb/eraa097.

L. Wang and Y. Wang. Transcription factor VqERF114 regulates stilbene synthesis in Chinese wild Vitis quinquangularis by interacting with VqMYB35. *Plant Cell Reports*, 38(10):1347–1360, 2019. ISSN 1432203X. doi: 10.1007/s00299-019-02456-4. URL `https://doi.org/10.1007/s00299-019-02456-4`.

E. R. Waters. Molecular adaptation and the origin of land plants. *Molecular Phylogenetics and Evolution*, 29(3):456 – 463, 2003. ISSN 1055-7903. doi: https://doi.org/10.1016/j.ympev.2003.07.018. URL `http://www.sciencedirect.com/science/article/pii/S1055790303003130`. Plant Molecular Evolution.

C. J. Wolfe, I. S. Kohane, and A. J. Butte. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics*, 6:1–10, 2005. ISSN 14712105. doi: 10.1186/1471-2105-6-227.

D. Wong. Network aggregation improves gene function prediction of grapevine gene co-expression networks. *Plant Molecular Biology*, 103(4-5):425–441, 2020. ISSN 15735028. doi: 10.1007/s11103-020-01001-2. URL `https://doi.org/10.1007/s11103-020-01001-2`.

D. C. J. Wong and J. T. Matus. Constructing integrated networks for identifying new secondary metabolic pathway regulators in grapevine: Recent applications and future opportunities. *Frontiers in Plant Science*, 8(April):1–8, 2017. ISSN 1664462X. doi: 10.3389/fpls.2017.00505.

D. C. J. Wong, R. Schlechter, A. Vannozzi, J. Höll, I. Hmmam, J. Bogs, G. B. Tornielli, S. D. Castellarin, and J. T. Matus. A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation. *DNA Research*, 23(5):451–466, 2016. ISSN 17561663. doi: 10.1093/dnares/dsw028.

H. Xi, L. Ma, G. Liu, N. Wang, J. Wang, L. Wang, Z. Dai, S. Li, and L. Wang. Transcriptomic analysis of grape (Vitis vinifera L.) leaves after exposure to ultraviolet C irradiation. *PLoS ONE*, 9(12), 2014. ISSN 19326203. doi: 10.1371/journal.pone.0113772.

W. Xu, M. Fuli, R. Li, Q. Zhou, W. Yao, Y. Jiao, C. Zhang, J. Zhang, X. Wang, Y. Xu, and Y. Wang. Vpsts29/sts2 enhances fungal tolerance in grapevine through a positive feedback loop. *Plant, Cell & Environment*, 42, 07 2019. doi: 10.1111/pce.13600.

L. Zhu, C. Gazin, N. Lawson, H. Pagès, S. Lin, D. Lapointe, and M. Green. Chippeakanno: A bioconductor package to annotate chip-seq and chip-chip data. *BMC bioinformatics*, 11:237, 05 2010. doi: 10.1186/1471-2105-11-237.