ARTICLE

Open Access

Genomic analysis uncovers functional variation in the C-terminus of anthocyanin-activating MYB transcription factors

Jessica A. Rodrigues¹, Richard V. Espley¹ and Andrew C. Allan^{1,2}

Abstract

MYB transcription factors regulate diverse aspects of plant development and secondary metabolism, often by partnering in transcriptional regulatory complexes. Here, we harness genomic resources to identify novel MYBs, thereby producing an updated eudicot MYB phylogeny with revised relationships among subgroups as well as new information on sequence variation in the disordered C-terminus of anthocyanin-activating MYBs. BLAST® and hidden Markov model scans of gene annotations identified a total of 714 MYB transcription factors across the genomes of four crops that span the eudicots: apple, grape, kiwifruit and tomato. Codon model-based phylogenetic inference identified novel members of previously defined subgroups, and the function of specific anthocyanin-activating subgroup 6 members was assayed transiently in tobacco leaves. Sequence conservation within subgroup 6 highlighted one previously described and two novel short linear motifs in the disordered C-terminal region. The novel motifs have a mix of hydrophobic and acidic residues and are predicted to be relatively ordered compared with flanking protein sequences. Comparison of motifs with the Eukaryotic Linear Motif database suggests roles in protein–protein interaction. Engineering of motifs and their flanking regions from strong anthocyanin activators into weak activators, and vice versa, affected function. We conclude that, although the MYB C-terminal sequence diverges greatly even within MYB clades, variation within the C-terminus at and near relatively ordered regions offers opportunities for exploring MYB function and developing superior alleles for plant breeding.

Introduction

MYB transcription factors (TFs) regulate diverse aspects of metabolism and development across eukaryotes¹. In plants, the MYB family has undergone relatively large expansions during the transition to life on land and the development of the seed habit². Seed-bearing plants possess hundreds of MYB TFs³, with different MYBs regulating diverse pathways such as stress responses, hormone signalling, organ development, reproduction, and secondary metabolite production^{4,5}. Several fruit and vegetable crop traits are MYB-regulated, including colour,

flavour, texture, nutritional content, and storage life⁴. Given the many adaptive and reproductive plant traits which MYBs regulate, MYB family expansion was likely an underlying factor in the complex assortment of cell fates plants evolved in order to thrive as sessile land organisms^{3,5}.

The MYB TF family is defined by one or more characteristic conserved MYB repeats in the DNA-binding domain, each of which consists of a helix–turn–helix motif⁶. Most plant MYB proteins contain two tandem N-terminal MYB repeats that are similar to the second and third domains of vertebrate MYBs. Thus, MYBs of this class are referred to as 2R or R2R3 MYBs. Other plant MYB family members contain either just a single R3 MYB repeat (R3 MYBs), three MYB repeats (R1R2R3 or 3R MYBs, the most common vertebrate class), or four

Correspondence: Andrew C. Allan (Andrew.Allan@plantandfood.co.nz)

¹The New Zealand Institute for Plant and Food Research Limited, 120 Mount Albert Road, Sandringham, Auckland 1025, New Zealand

²School of Biological Sciences, University of Auckland, 3A Symonds St, Auckland 1010, New Zealand

© The Author(s) 2021

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

MYB repeats (R1R2R2R1/2 or 4R MYBs)⁷. In contrast to the well-structured and conserved DNA-binding domains of the N-terminus, the larger C-terminal regions of R2R3 MYBs appear to be principally unstructured, varying greatly in length and sequence⁸. Such extended regions of intrinsic disorder are enriched among eukaryotic TFs compared with other proteins, associated with the increased presence of molecular recognition features (MoRFs) that participate in dynamic protein–protein interactions⁹. MoRFs are often identified as short linear motifs (SLiMs) of 3 to 11 amino acids that are evolutionarily variable, yet contain a few relatively conserved specificity-determining residues favouring structural order within a larger disordered context^{10–13}. Interactions at SLiMs tend to be specific yet low affinity, thus ideal for regulation, signalling and post-translational modification^{10,14}. It is widely thought that the loose constraints on protein structure and sequence at MoRFs have facilitated the evolution of TF regulation by varied cellular signalling pathways and interacting co-factors, thereby contributing to eukaryotic diversity and complexity^{9,10,12–14}. Hence, although the C-termini of plant MYBs are poorly conserved, they probably confer functional specialisation and adaptive regulation in a family where members vary greatly in their roles¹⁵.

Previous studies have categorised MYBs into over 25 subgroups based on phylogeny and function^{3,7}, with several subgroups displaying SLiM-like C-terminal sequence motifs that are conserved among their members¹⁶. Some of these conserved C-terminal sequence motifs are verified interaction sites, such as that with co-activating basic helix–loop–helix (bHLH) TFs at the subgroup 12 motif⁸ and that with repressive chromatin modifiers at one of the subgroup 4 motifs¹⁷. It is noteworthy that, although MYBs of subgroups 2, 5, 6, 15, 16, and 19 also require a bHLH partner to activate target promoters, the bHLH-interacting site for subgroups 5, 6, 15, and 16 occurs within the N-terminal R3 MYB sequence^{5,8,18–20}. Thus, bHLH interaction of subgroup 12 MYBs appears to have evolved convergently with that of subgroups 5, 6, 15, and 16, illustrating the power of SLiMs in *ex nihilo* evolution of protein regulation^{8,14}. Indeed, most SLiM classes include examples of convergent evolution¹¹. In addition to co-activator requirements, MYB function is regulated at transcriptional, post-transcriptional and post-translational levels^{4,7,21–24}. The C-terminal motif previously described for subgroup 6¹⁶, which includes anthocyanin-activating MYBs, coincides with the target sequence of trans-acting small interfering RNA TAS4-siRNA81(-)²⁵. It is likely that some conserved C-terminal motifs also contribute to post-translational regulation, serving either as docking or target sites for modification machinery¹².

Here, we explore MYB diversity and its functional implications, to enhance resources for crop improvement. We examine diversity and molecular evolution with the latest evolutionary models and resources for five mature reference genomes that span eudicots: *Arabidopsis thaliana*, *Malus × domestica* (apple), *Vitis vinifera* (grape), *Actinidia chinensis* (kiwifruit), and *Solanum lycopersicum* (tomato). We identify several new MYB family members in apple and kiwifruit while providing firmer support for some previously ambiguous phylogenetic relationships of MYB subgroups. Finally, by characterising novel members of the horticulturally important clade of anthocyanin-activating MYBs, we demonstrate how novel putative MoRFs in the disordered C-terminus enhance anthocyanin activation in functional assay.

Results

Enhanced annotation of MYB genes

Since *Arabidopsis thaliana* and *Solanum lycopersicum* (tomato) genomes have had extensive collaborative input in assembly and gene annotation^{26,27}, we performed hidden Markov model and BLAST® searches on the latest gene annotations, i.e. TAIR10 for *Arabidopsis* and ITAG4.0 for tomato. Our approach did not identify any *Arabidopsis* MYB genes beyond those already described⁷ but found four new genes for tomato²⁸. For the newer apple, kiwifruit and grape genomes, we first enhanced existing gene annotations by harnessing in-house and publicly available RNA sequencing (RNA-seq) data that spanned tissues and cultivars. Subsequent hidden Markov model and BLAST search identified MYBs spanning R3, R2R3 and 3R classes (Fig. 1a, b; Dataset S3). Our analyses excluded a minor subset of genes (~10%) that produce non-functional proteins from the reference genome sequence based on frameshifts or deletions of the N-terminal DNA binding domain. Thus, we finally identified 215 functional R2R3 MYBs in the kiwifruit genome compared with 155²⁹ and 93³⁰ from previous studies that used either the same reference genome or an earlier version from a different cultivar, respectively. We identified 199 functional apple R2R3 proteins, of which 23 were absent from the lists of 186³¹ and 222³² R2R3 MYBs generated using an earlier version of the apple genome³³. Unresolved heterozygosity resulted in the unintentional inclusion of both haplotypes in the sequence of this earlier assembly, and thus alleles were erroneously identified as distinct genes³⁴. We did not identify functional grape MYBs beyond those previously described³⁵.

RNA-seq data corrected the dominant splice isoform annotation for 64 MYB genes across apple, kiwifruit and grape genomes and identified a further 33 MYB genes that were not previously annotated, with three apple genes and one kiwifruit gene failing to map to the reference genome (Fig. 1b). The overall improvement to existing annotations

Fig. 1 High quality RNA-sequencing data enhanced genome-wide detection of MYB genes and improved the annotation of dominant splice isoforms. a Bioinformatic pipeline to obtain improved and novel gene annotations of published reference genomes using RNA-sequencing data. Publicly available tools used are indicated in brackets; dashed boxes indicate steps applied to some but not all datasets. **b** Numbers of R3, R2R3 and 3R MYBs identified in apple, kiwifruit and grape reference genomes, including novel genes not in published genome annotations and ones for which the dominant splice sites were corrected using RNA-sequencing data. Genes that contain frameshift mutations or premature termination codons in the MYB domain are indicated as non-functional

based on our approach was assessed by comparing de novo transcripts over 1 kb long, the revised gene annotations derived from them and the original published gene annotations with which they overlap (Fig. S1). On average, de novo transcripts and the revised annotations derived from them produced similar results in terms of transcript length, protein length and mapping of RNA-seq reads, to outperform the original published annotations for apple and kiwifruit and a significant improvement for grape. This could be attributed to enhanced definition of splice sites and untranslated regions.

Our approach of validating gene annotations through mapping of de novo assembled transcripts rather than

mapping of unassembled reads reduced mis-mapping at recently duplicated genes. For example, *Vitvi02g01019* (*MYBA1*) is one of 12 tandemly duplicated grape *MYBA* genes clustered on chromosome 2. Although unassembled reads map to *MYBA1* in the Sauvignon blanc dataset, de novo assembled transcripts from the dataset map only to *MYBA3* (*Vitvi02g01024*) and *MYBA2* (*Vitvi02g01015*). A search for *VvMYBA1*-specific single nucleotide polymorphisms (SNPs) across datasets from different grape varieties found that only one (Nebbiolo) de novo transcript possessed signature SNPs within the coding sequence; however, the rest of the transcript displayed characteristic *VvMYBA2* and *VvMYBA3* SNPs.

Fig. 2 Phylogenetic summary of functional R2R3 and 3R MYBs identified in *Arabidopsis*, apple, grape, kiwifruit and tomato genomes. Numbers on the maximum likelihood phylogenetic tree indicate support based on 10,000 ultrafast bootstrap replicates, while numbers in the table indicate the number of subgroup members in each species. R2R3 subgroup nomenclature, function, and presence in early diverging plants is based on previous publications^{3,4,7,50}. MYBs across the phylogeny that do not fall into the subgroups originally described⁷ are placed in the “Others” category. Reference genomes and/or transcriptomes used are: At *Arabidopsis thaliana* TAIR10, Sl *Solanum lycopersicum* (tomato) ITAG4.0, Md *Malus x domestica* (apple) GDDH13 v1.1, Vv *Vitis vinifera* 12x v3 with the Vcost.V3 annotation, Ac *Actinidia chinensis* (kiwifruit) Red5

Thus, there was no significant support for the expression of *VvMYBA1* in any of the datasets we analysed, but there was support for the expression of *VvMYBA3* in Sauvignon blanc, Pinot noir and Nebbiolo and *VvMYBA2* in Sauvignon blanc, Pinot noir, Tempranillo and Nebbiolo datasets.

Improved support for clading within and among subgroups

Our process of applying the same MYB annotation pipeline to multiple species allowed direct comparison of subgroup membership across species (Fig. 2). Variation in total numbers of MYBs among species reflects relatively recent genome duplications in apple and kiwifruit compared with grape, tomato and *Arabidopsis*. Both the maximum likelihood phylogenetic tree built with iqtree using a codon alignment of functional R2R3 and 3R MYBs (*Arabidopsis*, apple, grape, kiwifruit and tomato; Figs. 2, S2), as well as that built with RAXML using an amino acid

alignment of functional and non-functional R3, R2R3 and 3R MYBs (*Arabidopsis*, apple, grape and kiwifruit; Fig. S3), indicate with moderate to high support (bootstrap of 92 in Fig. S2 and 79 in Fig. S3) that all land plant-specific eudicot R2R3 subgroups except subgroups 22 and 23 belong to a single land plant-specific clade. Based on the functions of the early diverging subgroups within this large clade, it is likely that its ancestral algal R2R3 MYB played a role in growth and development before the transition to life on land. Both phylogenetic trees also converge in placing the clade of subgroups 21, 22 and 23 as sister to the large land plant-specific clade (bootstrap of 98 in Fig. S2 and 75 in Fig. S3). Current roles for these subgroups also indicate an ancestral role in growth and development. Thus, all land plant-specific eudicot R2R3 MYB subgroups owe their origin to only two of at least five ancient algal R2R3 MYB clades. There appears to have been stricter selection against expansion and diversification in the ancient 3R MYB subgroup

and R2R3 subgroups 29 (CDC5-like), 28 (AtMYB88, AtMYB124) and 25. This is likely due to lower tolerance of perturbations in the pathways they regulate: cell cycle, innate immunity, cell fate and gametogenesis, respectively^{4,7}.

The two trees also concur in defining membership within R2R3 MYB subgroups, confirming the monophyly of all subgroups except 18, 20, and 24. However, the two trees do differ in support and topology of relationships among different subgroups. Codon evolutionary models generally outperform amino acid models^{36,37}; nonfunctional MYB genes are more likely to possess homoplasies, and sampling of fewer species is more likely to result in failure to detect homoplasies, resulting in an inferior phylogeny³⁸. In line with this, inter-subgroup relationships are more strongly supported in the tree built using only functional R2R3 and 3R MYBs from *Arabidopsis*, apple, grape, kiwifruit and tomato (Fig. S2) compared with the other constructed tree (Fig. S3).

Within the clade of land plant-specific subgroups, the early divergence of development-related subgroups 18, 26 and 27 results in a large highly supported clade (bootstrap of 100) of subgroups 1 to 17, 19, 20 and 24, where regulation of secondary metabolism is the predominant function. MYB family expansion during the evolution of the seed habit is due to the expansion of this clade of secondary metabolism regulators. The highly supported clade (bootstrap of 94) of anthocyanin- and proanthocyanin-regulating subgroups AtMYB5, 5 and 6 evolved during that large expansion, probably from a flavonoid regulator based on the flavonol- and flavone-regulating functions of its sister, subgroup 7. These form a moderately supported (bootstrap of 76) clade together with subgroups 19 and 20, which mostly play roles in development and response to biotic and abiotic stress. The placement of angiosperm-specific organ patterning-related subgroup 15 as the closest sister to subgroup 6 is well-supported (bootstrap of 91).

Early diverging subgroup 6 members do not activate anthocyanin

Subsequent analysis focused on MYB subgroup 6, whose membership was identical in both phylogenetic trees and included previously undescribed members. For all species except grape, the closest homologs of described anthocyanin-promoting members were other members from the same species rather than putative orthologs from other species. Analysis could not differentiate whether this resulted from gene conversion among subgroup 6 members of the same species, restrictions placed by co-evolution with binding partners or independent replication from a single ancestral eudicot anthocyanin regulator in all five species. Grape anthocyanin activators fall into a clade of 12 tandem replicates located on chromosome 2

and a second clade of three tandem replicates on chromosome 14. The chromosome 2 clade includes genes that mostly contribute to fruit colour, such as the functionally redundant *VvMYBA1* and *VvMYBA2*³⁹, while the chromosome 14 clade appears to colour vegetative tissues⁴⁰.

A third clade of grape MYBs on chromosome 1, consisting of *Vitvi01g00094* (*VvMYBAL1*) and *Vitvi01g00095* (*VvMYBAL2*), is the earliest diverging of all the subgroup 6 members analysed. Although *VvMYBAL1* and *VvMYBAL2* have been previously assigned to subgroup 6, their function remains unclear³⁵. Another early diverging MYB is the newly identified apple MYB MD04G1235800. Like *VvMYBAL1* and *VvMYBAL2*, MD04G1235800 clades separately form MYBs with demonstrated anthocyanin-promoting function such as *MdMYB10* (MD09G1278600)⁴¹ and *MdMYB110a* (MD17G1261000)⁴². We included these genes in transient expression assays in tobacco leaves that compare function of novel and previously described subgroup 6 MYBs that represent the diversity across grape, apple and kiwifruit (Figs. 3a, b, S4, and S5a-h). MYB function was assessed by two measures: quantitatively as stimulation of anthocyanin production by activation of endogenous pathway genes of *Nicotiana tabacum* (Figs. 3a, Fig. 3, S4), and qualitatively by dual-luciferase assay in *Nicotiana benthamiana* to test activation of promoters of apple *dihydroflavonol-4-reductase 1* (*MdDFR1*) and apple *UDP-glucose:flavonoid 3-O-glycosyltransferase 1* (*MdUGFT1*), key genes in the late anthocyanin biosynthesis pathway (Figs. 3a, S5a-h). Since some subgroup 6 MYBs from species distantly related to tobacco, such as *MdMYB10*⁴¹, appear to be unable to partner with endogenous tobacco bHLHs in order to activate gene expression, we included *MdbHLH3* in all infiltrations so tested MYBs are likely to find suitable co-factors for activity.

None of the early diverging subgroup 6 MYBs significantly activated either the *N. tabacum* biosynthetic pathway or the *MdDFR1* and *MdUGFT1* promoters. Apple *MdMYB10*, *MdMYB110a* and kiwifruit *AcMYB10* and *AcMYB110* had high activity in both assays, as previously demonstrated^{41–43}. *VvMYBA1r* also returned positive results, as expected³⁹. *VvMYBA10* is expressed along with *VvMYBA1r* and/or other MYBs from the chromosome 2 locus during anthocyanin activation³⁵; however, it remains unclear whether it promotes anthocyanin production. Our results show that *VvMYBA10* does promote anthocyanin, but not as strongly as *VvMYBA1r*. Similarly, novel kiwifruit MYB *AcMYB310* promotes anthocyanin less strongly than *AcMYB10* and *AcMYB110*. Phylogeny suggests that *AcMYB310* is the earliest diverging kiwifruit subgroup 6 MYB (Figs. S2, S3). Additionally, *AcMYB110* and *AcMYB210* are tandem duplicates on linkage group 9, while *AcMYB10* is their homeologue on linkage group 1 from a recent whole-genome duplication specific to the kiwifruit lineage⁴⁴.

AcMYB310 on linkage group 27 appears to have a different origin. The two anthocyanin-activating grape loci on chromosomes 2 and 14 also do not appear to be homeologues from the ancient hexaploidization event in a common ancestor of dicots⁴⁵. *MdMYB10* on apple chromosome 9 is a homeologue of chromosome 17 *MdMYB110a* from a recent whole-genome duplication³⁴. Although novel subgroup 6 apple genes *MD01G1017100* and *MD05G1276500* form a well-

supported clade with *MdMYB10*, *MdMYB110a* along with their tandem duplicates (bootstrap of 97), *MD01G1017100* and *MD05G1276500* are not homeologues of each other from the recent apple whole genome duplication. *MD01G1017100* did not activate either assay for anthocyanin-promoting ability, while *MD05G1276500* activated *MdDFR1* and *MdUGT1* promoters in *N. benthamiana* but only weakly activated anthocyanin production in *N. tabacum*.

Conserved C-terminal motifs of anthocyanin-activating subgroup 6 members display molecular recognition feature (MoRF)-like characteristics

The C-terminus of subgroup 6 members is more divergent than the R2 and R3 DNA binding domains (Fig. S6), and the extent to which this variation correlates with variation in activation strength is unclear. To identify functional C-terminal regions, we first searched sequence conservation among MYBs that moderately or strongly activated anthocyanin (Fig. 4). We defined moderate to strong activation as a colour development score > 1 in our study or similar colour development in the literature. Three regions were identified of 15 to 17 codons that were more conserved than average, and these were named S6A, S6B and S6C. As the S6A region encompassed the previously described 7-codon motif that characterises subgroup 6, we equated the S6A SLiM with the previously termed “subgroup 6 motif”¹⁶. IUPred2A⁴⁶ and DISOPRED3⁴⁷ were then used to predict protein disorder and protein-binding potential across anthocyanin activating and non-activating MYBs (Figs. 5, S5i). Results indicate that the highly conserved seven amino acids of the S6A motif as well as the core five to eight amino acids of S6B and S6C motifs possessed a greater tendency toward structural order than flanking intrinsically disordered

regions (Figs. 5, S5i), a characteristic of MoRFs¹⁰. Additionally, both S6B and S6C motifs were enriched in hydrophobic and acidic residues and localised near the C-terminus in at least one MYB.

It is interesting that the S6A motif, which has a confirmed role in post transcriptional repression²⁵, also demonstrated MoRF-like features which are functional in the protein. The Eukaryotic Linear Motif (ELM) database⁴⁸ lists the similar SLiM [RK]-x(2,4)-[LIVMP]-x-[LIV]-x-[LIVMF] (ELM identifier: DOC_MAPK_MEF2A_6; described here in PROSITE notation) to be a kinase docking motif that mediates interaction with the ERK1/2 and p38 subfamilies of MAP kinases. *Arabidopsis* MAP KINASE4, which belongs to the ERK subfamily, stabilises subgroup 6 MYB *AtMYB75* and is essential for light-induced anthocyanin accumulation⁴⁹. In line with the hypothesis that S6A variation fine-tunes an interaction with MAP KINASE4, *AcMYB110* is the only MYB to fully match the ELM database DOC_MAPK_MEF2A_6 motif with its sequence R-P-Q-P-R-I-F-M (Fig. S6), and was also the strongest activator in our assay.

Like the S6A motif, S6B did not fully match any known ELM database entries, but its core SLiM is similar to the SLiM x-x-x-x-W-F-x-x-L (ELM identifier: LIG_PALB2_WD40_1) required by the BRCA2 DNA repair

Fig. 5 Protein regions corresponding to conserved C-terminal motifs of anthocyanin-activating subgroup 6 R2R3 MYBs are predicted to be less disordered than flanking regions, with non-activators (MD04G1235800, Vitvi01g00094) diverging at S6A and S6B motifs compared with strong anthocyanin activators (AcMYB110, VvMYBA1r) and moderate activators (AcMYB310, MdMYB10). **a** Disorder and protein binding scores along select subgroup 6 R2R3 MYB protein sequences, as predicted by IUPred2A⁴⁶ and DISOPRED3⁴⁷ programs for protein structure analysis. Amino acid position is plotted along the x-axis with conserved MYB DNA binding domains and C-terminal motifs denoted by filled grey and blue boxes, respectively. Clear boxes outlined in blue highlight regions that align with the sequence motifs of other MYBs but are considerably diverged from the consensus. Predicted disordered regions are defined by a disorder score of 0.5 to 1, while protein binding regions are defined by a binding score of 0.5 to 1. Note that, unlike ANCHOR2, DISOPRED3.1 specialises in predicting protein binding probability for disordered regions and does not make predictions for ordered regions. **b** A representative sample of sequence variation at C-terminal motifs within subgroup 6. Amino acids are shown in red when they differ from the activator-based consensus motif and bold when part of a perfect match to a Eukaryotic Linear Motif database entry (DOC_MAPK_MEF2A_6 for S6A and LIG_WD40_WDR5_VDV_2 for S6C)

protein for binding to WD40 protein PALB2. The core SLiM of S6C was the only one to fully match an ELM database entry: [EDSTY]-x(0,4)-[VIPLA]-[TSDEKR]-[ILVA] (ELM identifier: LIG_WD40_WDR5_VDV_2). This is a fungi-specific motif that binds to WD40 protein WDR5, which mediates histone modification complex assembly. S6C also shares similarities with [ED]-x(0,3)-[VIL]-D-[VI]

(ELM identifier: LIG_WD40_WDR5_VDV_1; the metazoan version of LIG_WD40_WDR5_VDV_2), and F-[EDQS]-[MILV]-[ED]-[MILV]-((x(0,1)-[ED]) | (*)) (ELM identifier: DOC_WD40_RPTOR_TOS_1; occurs either at the N- or C-terminal end and binds to the WD40 protein Raptor, which is a part of the TOR protein kinase complex). In general, MYBs with no or poor anthocyanin-promoting

ability possessed distinct sequence variation at S6A, S6B and S6C regions compared with stronger activators (Figs. 5, S6), but this was most pronounced at S6A.

Engineering of motifs from weak anthocyanin activators into strong anthocyanin activators and vice versa reveals functional effects of sequence variation

As the functional importance of S6A is already underscored by its role as a small RNA target site, we focused on testing whether S6B and S6C motifs contributed to MYB function. Unlike for subgroup 12, subgroup 6 C-terminal motifs are not expected to play a role in bHLH interaction, since the C-terminus of subgroup 6 AtMYB75 cannot recruit bHLH AtGL3, while that of subgroup 12 AtMYB29 recruits bHLH AtMYC4⁸. Eleven modified gene sequences were created by swapping motif sequences between closely related subgroup 6 MYBs that varied in colour activation and sequence (Fig. 6a, Supplementary Materials and Methods). For most motif swaps, flanking sequences were included on either side of the motif, as they also play a role in defining specificity and affinity of MoRF interactions^{10,12}.

The modified MD05G1276500-1 allele, where the natural S6B sequence is replaced with that of stronger anthocyanin activator MdMYB10, showed greater anthocyanin production than the un-modified allele in *N. tabacum* leaves, with a score of 2 versus 1 (Figs. 6a, S4). The insertion of MdMYB10 S6B sequence into a modified MD01G1017100 allele did not increase colour development from the initial score of 0, but this was probably due to the effect of mutations in the R2R3 DNA binding domain of MD01G1017100 that cannot be ameliorated by an improved S6B region. MD01G1017100 contains several non-synonymous mutations in the MYB DNA binding domains that are not shared by any anthocyanin-activating MYBs: R17K, R24Q, Q25H, C26F, K29Q, H36R, T69G, I77V, L83V, L89M and N107S (Fig. S6). Replacement of the MdMYB10 S6B motif with S6B motifs from either MD01G1017100 or MD05G1276500 decreased the colour activation score from 3 to 2 (Figs. 6a, b, S4), supporting the hypothesis that this region fine-tunes the activity of the MYB.

For the S6C motif, replacement of the strong activator AcMYB110 motif sequence into the weaker AcMYB310 improved anthocyanin activation in AcMYB310-1, while engineering of the weak AcMYB310 S6C motif into AcMYB110 decreased anthocyanin activation in AcMYB110-1. Although MdMYB110a is a naturally truncated allele, where a frameshift mutation soon after S6B produces a premature termination codon with loss of S6C, it is a stronger anthocyanin activator than MdMYB10, which contains both S6B and S6C motifs. Addition of the MdMYB10 S6C motif to MdMYB110a did not further increase colour activation in MdMYB110a-1. This suggests

that the MdMYB10 S6C motif is ineffective in the assay or functionally redundant to at least one other motif of MdMYB110a, probably the S6B motif given sequence similarity. Vitvi02g01308 is also mutated near the C-terminus with apparent loss of S6C. However, although the sequence replacing S6C in Vitvi02g01308 does not match the WDR5-binding motif and differs from the TOS motif by substituting acidic residues [ED] with G, it is similar to S6C in being enriched for acidic and hydrophobic residues and is predicted to be even more ordered than the VvMYBA1r S6C motif (Fig. S6, S5i). For VvMYBA1r and Vitvi02g01308, we tested the effect of swapping either a smaller fragment that included the core S6C SLiM but excluded flanking sequence or a larger fragment that included 13 to 21 codons of flanking sequence. Neither size of the VvMYBA1r sequence swapped into Vitvi02g01308 improved anthocyanin production, just as when the larger Vitvi02g01308 sequence was swapped into VvMYBA1r. However, when the smaller Vitvi02g01308 S6C sequence was swapped into VvMYBA1r, colour activation score decreased from 7 for VvMYBA1r to 6 for VvMYBA1r-2. This highlights how flanking disordered sequences are important for interaction at SLiMs^{10,12,13}.

Discussion

Codon models and improved genomic resources enhance phylogenetic analyses

This work builds upon previous studies which explored MYB diversity by harnessing recent improvements to reference genomes and increased availability of gene expression data. Codon models of evolution outperform nucleotide and amino acid models in phylogenetic analysis^{36,37}, and our codon model-based phylogeny better resolves relationships between MYB subgroups. The topological improvements made by sampling genes from an additional species, tomato, and excluding non-functional proteins, are evident by comparison of the two different trees in the analysis. Some well-supported conclusions of our phylogenetic analysis have been indicated with weaker support in other phylogenies, including those built with multiple plant species^{3,50} and species-focused phylogenies for grape³⁵ and kiwifruit³⁰. These comprise the early divergence of R2R3 subgroups 29 (CDC5-like), 28 (AtMYB88, AtMYB124) and 25 relative to subgroups that expanded during the transition to life on land, and the early divergence of subgroups 18, 22, 23, 26 and 27 within the larger clade of land plant-specific subgroups. The indication that most land plant-specific MYB regulators of secondary metabolism evolved from an ancient algal regulator of growth and development may be unsurprising given that development-related subgroup 15 is the closest sister to anthocyanin-regulating subgroup 6.

The placement of subgroups 4, 19 and 20 as the closest relatives of the flavonol- and anthocyanin-regulating clade

is mirrored in multi-species phylogenies of *Arabidopsis*, peach, strawberry and apple³¹, as well as single-species phylogenies of *Arabidopsis*⁷ and grape³⁵. However, multi-species phylogenies that span land plants^{3,50} place development- and stress-related subgroups 19 and 20 closer to cell wall metabolism-related subgroups 13 and 16, although with lower support than our analysis. At this stage it is unclear if this latter result is due to increased taxon sampling, reducing the effect of homoplasy³⁸, or whether codon models more effectively resolve deep divergence. In support of our placement, the subgroup 20 MYB AtMYB112 shares similarities with subgroup 6 members⁵¹ in promoting anthocyanin formation under abiotic stress and specifically activating subgroup 4 repressor MYBs AtMYB6, AtMYB7 and AtMYB32⁵². If

true, this placement would be another indication of a short evolutionary distance between secondary metabolite regulation and developmental regulation. Most previous publications support our clading of repressor MYB subgroup 4 with flavonol and anthocyanin regulators^{7,31,35,50}, suggesting that repression evolved alongside activation from a common ancestral subgroup.

C-terminal S6A, S6B and S6C motifs of subgroup 6 MYBs are putative MoRFs, suggesting roles in protein–protein interactions

Each relatively conserved region within the C-terminus (i.e. S6A, S6B and S6C) was also predicted to have greater protein structural order than the rest of the disordered C-terminus. This suggests they are MoRFs, which are often

disordered when isolated but form secondary structures upon interaction¹¹. Subgroup 6 MYBs, like many other environmentally responsive TFs, can participate in large interactomes, and this is facilitated by having intrinsically disordered protein regions that allow flexible interaction^{12,13}. Disorder predictions varied between DISOPRED3 and IUPred2A at times, possibly owing to differences in algorithm and datasets used to train the predictors via machine learning.

The implication that motif S6A plays a role in regulation by TAS4 as well as in protein–protein interaction, is not only supported by our analysis of predicted structure but also by a study that found protein sequence in tobacco and *Arabidopsis* was conserved better than DNA sequence despite eventual mismatches to TAS4-siRNA81(-)²⁵. Evolution may favour the overlap of small RNA target sites with regions of conserved protein function, as seen in NB-LRR genes⁵³. Additionally, many subgroup 6 MYBs are the targets of the microRNAs miRNA-858 and miRNA-828 at sites in the R2R3 MYB DNA binding domain important for protein function^{24,54}. The similarity between S6A and ELM database⁴⁸ entry DOC_MAPK_MEF2A_6, along with the previous finding that *Arabidopsis* MAP KINASE4 stabilises subgroup 6 MYB AtMYB75 and is essential for light-induced anthocyanin accumulation⁴⁹, implies a role for S6A in interaction with MAP KINASE4. Our transient assays in tobacco leaves were strongly influenced by the presence of light, possibly accounting for S6A variation being a greater predictor of activation strength than that at S6B and S6C. The discrepancy between the S6A consensus and DOC_MAPK_MEF2A_6 may be due to a lack of plant instances for this motif in the ELM database. MoRFs conserved across biological kingdoms have more greatly diverged instances among phyla and subphyla than do ordered protein domains, so divergence between plant and mammalian instances is expected^{11,14}.

Core SLiMs of S6B and S6C motifs differed from those of S6A, being enriched in hydrophobic and acidic residues and matching a generic SLiM Φ -x-x- Φ - Φ (where Φ indicates a hydrophobic residue) seen in the activation domain of many TFs¹³. S6B is conserved in *Antirrhinum* subgroup 6 MYBs⁵⁵, while an S6C-like motif is required for activity of maize subgroup 5 anthocyanin activator ZmC1⁵⁶. Subgroup 6 MYBs show no or severely decreased activity after loss of S6B and S6C in *Petunia*⁵⁷, or even S6C alone in *Petunia* and *Trifolium*⁵⁸. Sequence variation at and near the ZmC1 S6C-like motif alters activation strength⁵⁹, as does that at a similar essential SLiM of plant AP2/ERF TFs⁶⁰. We show the functional importance of S6B in motif swaps between apple MYBs. The S6B SLiM is similar to ELM database entry LIG_PALB2_WD40_1, the motif of DNA-binding protein BRCA2 that binds WD40 protein PALB2. Although functional differences between BRCA2 and subgroup 6

MYBs make it unlikely that S6B recruits a plant PALB2 homolog, it is likely that another WD40 protein plays a role. Many anthocyanin-activating MYBs require the presence of a WD40 scaffold protein along with a bHLH cofactor to upregulate target anthocyanin biosynthetic enzymes like DFR, UFGT, and leucoanthocyanidin dioxygenase (LDOX; also known as anthocyanidin synthase or ANS)^{22,61}, with the WD40 partner undiscovered in many cases⁵. The S6C motif also matched ELM database motifs for interaction with WD40 proteins, although it is possible that putative WD40 interactors include WDR5 and key proteins of the TOR protein kinase complex, since some of their functions appear to be conserved between metazoans, fungi and plants^{62–65} and appear to affect anthocyanin accumulation⁶⁶. Whilst several SLiM-binding domain families are experimentally verified to bind non-canonical motifs, WD40 family members are especially known to target physicochemically diverse motifs, possibly explaining why the S6C replacement sequence in Vitvi02g01308 sustains function¹². As S6C sequence variation has a strong effect in the motif swap between AcMYB110 and AcMYB310, it is likely that S6C loss in MdMYB110a and S6C swaps in grape were ineffectual because S6C function is redundant to that of S6B, or our assay was less suitable for apple and grape MYBs than for kiwifruit MYBs. SLiM redundancy has been previously described for mammalian proteins¹³ and species-specific bias of the tobacco leaf assay is highlighted by increased interaction of kiwifruit MYBs with endogenous tobacco bHLHs⁴³ compared with MYBs from apple, which is a more distantly related species^{41,67}.

Notably, S6B and S6C motifs are similar in relative location and sequence to previously described motifs for closely related subgroups 15 (W-V-x-x-D-x-F-E-L-S-x-L) and 5 (D-E-x-W-R-L-x-x-T), respectively¹⁶. The subgroup 15 motif of *Arabidopsis* WEREWOLF is sufficient for transcriptional activation in yeast one-hybrid assay⁶⁸, but its loss in an allele of *Arabidopsis* GLABROUS 1 (AtGL1) appears to be partially rescued by fellow subgroup 15 member AtMYB23 as well as AtGL1 interaction with the bHLH GLABROUS 3 (AtGL3) that recruits WD40 protein TRANSPARENT TESTA GLABRA 1 (AtTTG1)⁶⁹. Like subgroup 6, subgroups 15 and 5 function in complex with bHLH and WD40 cofactors⁵, with bHLH interaction occurring through the MYB R3 sequence²⁰. Subgroup 15 and 5 motifs probably recruit regulators similar to those recruited by S6B and S6C, thereby affecting transcriptional activation strength. C-terminal motifs enriched in hydrophobic and acidic residues also occur at similar positions in widely diverged subgroups 1 to 6, 15, 16 and 21¹⁶. Motifs for subgroups 9, 11 and 19 are also enriched in hydrophobic and acidic residues but positioned more similarly to S6A¹⁶. Other subgroups may possess homologous or analogous regions that are difficult to detect

owing to the rapid evolution of SLiMs, as seen by comparing the S6C motif of VvMYBA1r with the equivalent region in Vitvi02g01308.

S6A, S6B and S6C motifs provide opportunities for dissecting and harnessing MYB function

Although sequence variation within MoRFs is well tolerated, evidence suggests that the sequence within and also flanking interaction sites fine-tunes interaction strength^{12,13}. Future work aimed at characterising interactors of S6A, S6B and S6C motifs will elucidate regulatory mechanisms that are conserved across eudicot MYB subgroup 6 members and possibly even across eudicot MYB subgroups. It is possible that S6A, S6B and S6C contain further overlapping SLiMs undetected in our analysis. Experimental validation of SLiMs will help build valuable resources for the detection of plant SLiMs. This provides opportunities for both engineering MYB TFs with optimal function as well as sequence-based germplasm screening to identify ideal lines for breeding elite crops.

Materials and methods

Genomic analysis

Gene annotation employed the *Actinidia chinensis* Red5 kiwifruit genome⁴⁴, doubled haploid *Malus x domestica* ‘Golden Delicious’ apple genome³⁴ and updated 12xV3 version of the *Vitis vinifera* “Helfensteiner x Pinot noir” grape genome^{45,70}. The method summarised in Fig. 1 is detailed in the Supplementary Materials and Methods. *Arabidopsis* and tomato transcripts were obtained from the TAIR10²⁷ and ITAG4.0²⁶ annotations, respectively. Gene products were scanned for the MYB Hidden Markov Model (PF00249) using hmmer version 3.1b2 (<http://hmmer.org>). R2R3 and 3R MYBs were identified by the presence of tandem MYB domains. Two iterative BLAST⁷¹ searches detected R2R3 and 3R MYBs that failed screening with PF00249, as well as single domain R3 MYBs. The resultant list was manually curated to exclude MYB-related genes and correct instances where MYB gene annotations were fused with that of another gene.

Phylogenetic analysis

To infer the R2R3 MYB phylogeny, we only used tandem R2R3 domain sequences free of frameshift mutations and premature termination codons. For our codon-model analysis, codons of 846 functional R2R3 and 3R MYBs from apple, *Arabidopsis*, grape, kiwifruit and tomato were aligned with MACSE v2⁷² and manually refined in Geneious 10.0.9 (<https://www.geneious.com>) to generate a 336 nucleotide-long alignment after masking non-informative positions. A maximum likelihood phylogeny with 10,000 bootstrap replicates under the MGK + F3X4 + R9 codon model of evolution was generated with iqtree version 1.6.11 using the command “iqtree -s alignment.phy -st

CODON -nt 1 -seed 5385 -bb 10000 -bnni”^{73–76}. A second maximum likelihood phylogeny was generated with RAXML 8.2.11⁷⁷ using a 221 amino acid-long alignment of 791 R3, R2R3 and 3R MYBs of apple, *Arabidopsis*, grape and kiwifruit and the JTTDCMUT substitution model. The command used was “raxmlHPC -f a -x 621986 -p 621986 -N autoMRE -m PROTGAMMAAUTO”. The amino acid alignment was generated by manual refining an alignment with default settings of the MAFFT⁷⁸ plug-in in Geneious 10.0.9 (<https://www.geneious.com>). Informative positions are provided in the alignments in Supporting information Datasets S1 and S2. MYB subgroup membership was inferred phylogenetically, with previously published information on MYB diversity of algae and early diverging land plants^{3,50} serving to identify ancient subgroups and orient the unrooted trees.

Characterisation of C-terminal sequence motifs

Full-length nucleotide sequences of novel and previously described MYBs that represent the sequence diversity of subgroup 6 were aligned using MACSEv2⁷² with manual refinement in Geneious 10.0.9. Conserved regions were detected based on identity across the alignment and motif sequence logos were generated with Geneious. Protein disorder was predicted with IUPred2A⁴⁶ and DISOPRED3⁴⁷ under default settings.

Generation of MYB constructs

Constructs for AcMYB10, AcMYB110, MdMYB10 and MdMYB110a were as produced previously^{43,67}. Other subgroup 6 MYBs were synthesised into a pUC57-Kan plasmid with flanking attL sites (Genewiz® LLC, genewiz.com) based either on the sequences of de novo assembled transcripts or that of the reference genome if not significantly different from de novo assemblies: constructs referred to primarily by annotation number derived from the reference genome, while those referred to by MYB number originate in a different cultivar. Naturally occurring BsaI sites were mutated to facilitate Golden Gate cloning of motifs of weak activators into strong activators and vice versa. Cloning methods are detailed in the Supplementary Materials and Methods, with the Golden Gate cloning approach following recommendations of the laboratory of Jeffrey Barrick at The University of Texas at Austin (<https://barricklab.org/twiki/bin/view/Lab/GoldenGateAssemblyProtocolsMainPage>).

Dual-luciferase assay

The ability of MYB constructs to activate the promoters of apple anthocyanin pathway genes *MdDFR1* and *MdUFGT1* was tested in the presence of MdbHLH3 as a co-factor by a previously described dual-luciferase assay involving transient *Agrobacterium*-mediated transformation of *Nicotiana benthamiana* leaves⁷⁹. Details are in the Supplementary Materials and Methods.

Transient colour development assay in *Nicotiana tabacum*

The ability of MYB constructs to activate endogenous anthocyanin biosynthesis in *Nicotiana tabacum* was tested by transient *Agrobacterium*-mediated transformation of MYB constructs into growing leaves along with the co-factor MdbHLH3, as previously described⁴¹. *Agrobacterium tumefaciens* GV3101 transformants for *MdbHLH3* and subgroup 6 MYB genes were grown, prepared and infiltrated into leaves as described for the dual-luciferase assay, with the exclusion of dual-luciferase constructs from the final inoculation mixture. Leaves from *N. tabacum* were used, as *N. benthamiana* does not produce anthocyanins naturally. Colour development was assayed between 7 and 12 days after infiltration (DAI). Constructs were compared with positive and negative controls on the same leaf, and colour activation scores ranging from 0 to 8 were assigned on the basis of at least 10 biological replicates.

Acknowledgements

We thank Dan Jones, Carole Smerilli, Ross Crowhurst, Tejas Sevak, Eric Burgueno, Tim Millar and Anali Thrimawithana for training in bioinformatic analysis, Kevin Davies for helpful discussions and advice regarding MYB phylogenetics, and Nick Albert and Catrin Guenther for helpful discussion of the manuscript. This work was funded by The New Zealand Institute for Plant and Food Research Limited, as well as the New Zealand Government through a Rutherford Postdoctoral fellowship from the Royal Society Te Apārangi-Rutherford Foundation.

Author contributions

J.A.R. and A.C.A. conceived the project and acquired funding. J.A.R., R.V.E. and A.C.A. designed experiments. J.A.R. performed analyses and experiments. J.A.R., R.V.E. and A.C.A. wrote the paper.

Data and code availability

In addition to inclusion of sequences used for phylogenetic analysis in Datasets S1 and S2, genomic locations of all apple, grape and kiwifruit MYBs, including novel MYBs identified by genomic analysis in this study, are released in Dataset S3. The nextflow pipeline for automated genome annotation enhancement is provided in the Supplementary Information as Code S1.

Conflict of interest

The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41438-021-00514-1>.

Received: 6 December 2020 Revised: 16 February 2021 Accepted: 1 March 2021

Published online: 01 April 2021

References

- Lipsick, J. S. One billion years of Myb. *Oncogene* **13**, 223–235 (1996).
- One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
- Du, H. et al. The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: new insights into subfamily classification and expansion. *Sci. Rep.* **5**, 11037 (2015).
- Allan, A. C. & Espley, R. V. MYBs drive novel consumer traits in fruits and vegetables. *Trends Plant Sci.* **23**, 693–705 (2018).
- Feller, A., Machemer, K., Braun, E. L. & Grotewold, E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* **66**, 94–116 (2011).
- Ogata, K. et al. The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat. Struct. Biol.* **3**, 178–187 (1996).
- Dubos, C. et al. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **15**, 573–581 (2010).
- Millard, P. S., Weber, K., Kragelund, B. B. & Burrow, M. Specificity of MYB interactions relies on motifs in ordered and disordered contexts. *Nucleic Acids Res.* **47**, 9592–9608 (2019).
- Liu, J. et al. Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873–6888 (2006).
- Fuxreiter, M., Tompa, P. & Simon, I. Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* **23**, 950–956 (2007).
- Davey, N. E. et al. Attributes of short linear motifs. *Mol. Biosyst.* **8**, 268–281 (2012).
- Van Roey, K. et al. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* **114**, 6733–6778 (2014).
- Staby, L. et al. Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem. J.* **474**, 2509–2532 (2017).
- Neduva, V. & Russell, R. B. Linear motifs: evolutionary interaction switches. *FEBS Lett.* **579**, 3342–3345 (2005).
- Millard, P. S., Kragelund, B. B. & Burrow, M. R2R3 MYB transcription factors - functions outside the DNA-binding domain. *Trends Plant Sci.* **24**, 934–946 (2019).
- Stracke, R., Werber, M. & Weisshaar, B. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **4**, 447–456 (2001).
- Kagale, S. & Rozwadowski, K. EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression. *Epigenetics* **6**, 141–146 (2011).
- Jang, I. C., Yang, S. W., Yang, J. Y. & Chua, N. H. Independent and interdependent functions of LAF1 and HFR1 in phytochrome A signaling. *Genes Dev.* **21**, 2100–2111 (2007).
- Zhang, F., Gonzalez, A., Zhao, M., Payne, C. T. & Lloyd, A. A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. *Development* **130**, 4859–4869 (2003).
- Zimmermann, I. M., Heim, M. A., Weisshaar, B. & Uhrig, J. F. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J.* **40**, 22–34 (2004).
- Colanero, S., Tagliani, A., Perata, P. & Gonzali, S. Alternative splicing in the anthocyanin fruit gene encoding an R2R3 MYB transcription factor affects anthocyanin biosynthesis in tomato fruits. *Plant Commun.* **1**, 100006 (2020).
- Dooner, H. K., Robbins, T. P. & Jorgensen, R. A. Genetic and developmental control of anthocyanin biosynthesis. *Annu. Rev. Genet.* **25**, 173–199 (1991).
- Pireyre, M. & Burrow, M. Regulation of MYB and bHLH transcription factors: a glance at the protein level. *Mol. Plant* **8**, 378–388 (2015).
- Xia, R., Zhu, H., An, Y. Q., Beers, E. P. & Liu, Z. Apple miRNAs and tasiRNAs with novel regulatory networks. *Genome Biol.* **13**, R47 (2012).
- Velten, J., Cakir, C. & Cazzonelli, C. I. A spontaneous dominant-negative mutation within a 35S::AtMYB90 transgene inhibits flower pigment production in tobacco. *PLoS One* **5**, e9917 (2010).
- Hosmani, P. S. et al. An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. Preprint at *bioRxiv* <https://doi.org/10.1101/767764> (2019).
- Lamesch, P. et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012).
- Li, Z. et al. Genome-wide identification and analysis of the MYB transcription factor superfamily in *Solanum lycopersicum*. *Plant Cell Physiol.* **57**, 1657–1677 (2016).
- Yu, M., Man, Y. & Wang, Y. Light- and temperature-induced expression of an R2R3-MYB gene regulates anthocyanin biosynthesis in red-fleshed kiwifruit. *Int. J. Mol. Sci.* **20**, 5228 (2019).
- Li, W. et al. Kiwifruit R2R3-MYB transcription factors and contribution of the novel AcMYB75 to red kiwifruit anthocyanin biosynthesis. *Sci. Rep.* **7**, 16861 (2017).
- González, M., Carrasco, B. & Salazar, E. Genome-wide identification and characterization of R2R3MYB family in *Rosaceae*. *Genom. Data* **9**, 50–57 (2016).

32. Cao, Z.-H., Zhang, S.-Z., Wang, R.-K., Zhang, R.-F. & Hao, Y.-J. Genome wide analysis of the apple MYB transcription factor family allows the identification of MdoMYB121 gene conferring abiotic stress tolerance in plants. *PLoS One* **8**, e69955 (2013).
33. Velasco, R. et al. The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat. Genet.* **42**, 833–839 (2010).
34. Daccord, N. et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* **49**, 1099–1106 (2017).
35. Wong, D. C. J. et al. A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation. *DNA Res.* **23**, 451–466 (2016).
36. Ren, F., Tanaka, H. & Yang, Z. An empirical examination of the utility of codon-substitution models in phylogeny reconstruction. *Syst. Biol.* **54**, 808–818 (2005).
37. Seo, T.-K. & Kishino, H. Statistical comparison of nucleotide, amino acid, and codon substitution models for evolutionary analysis of protein-coding sequences. *Syst. Biol.* **58**, 199–210 (2009).
38. Philippe, H. et al. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLOS Biol.* **9**, e1000602 (2011).
39. Walker, A. R. et al. White grapes arose through the mutation of two similar and adjacent regulatory genes. *Plant J.* **49**, 772–785 (2007).
40. Matus, J. T. et al. A group of grapevine MYBA transcription factors located in chromosome 14 control anthocyanin synthesis in vegetative organs with different specificities compared with the berry color locus. *Plant J.* **91**, 220–236 (2017).
41. Espley, R. V. et al. Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* **49**, 414–427 (2007).
42. Chagne, D. et al. Mapping a candidate gene (MdMYB10) for red flesh and foliage colour in apple. *BMC Genom.* **8**, 212 (2007).
43. Peng, Y. et al. Differential regulation of the anthocyanin profile in purple kiwifruit (*Actinidia* species). *Hortic. Res.* **6**, 3 (2019).
44. Pilkington, S. M. et al. A manually annotated *Actinidia chinensis* var. *chinensis* (kiwifruit) genome highlights the challenges associated with draft genomes and gene prediction in plants. *BMC Genom.* **19**, 257 (2018).
45. Jaillon, O. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007).
46. Meszaros, B., Erdos, G. & Dosztanyi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).
47. Jones, D. T. & Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857–863 (2015).
48. Gouw, M. et al. The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Res.* **46**, D428–d434 (2018).
49. Li, S. et al. MYB75 phosphorylation by MPK4 is required for light-induced anthocyanin accumulation in *Arabidopsis*. *Plant Cell* **28**, 2866–2883 (2016).
50. Bowman, J. L. et al. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**, 287–304 e215 (2017).
51. Albert, N. W. et al. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell* **26**, 962 (2014).
52. Lotkowska, M. E. et al. The *Arabidopsis* transcription factor MYB112 promotes anthocyanin formation during salinity and under high light stress. *Plant Physiol.* **169**, 1862–1880 (2015).
53. Zhang, Y., Xia, R., Kuang, H. & Meyers, B. C. The diversification of plant NBS-LRR defense genes directs the evolution of microRNAs that target them. *Mol. Biol. Evol.* **33**, 2692–2705 (2016).
54. Tirumalai, V., Swetha, C., Nair, A., Pandit, A. & Shivaprasad, P. V. miR828 and miR858 regulate VvMYB114 to promote anthocyanin and flavonol accumulation in grapes. *J. Exp. Bot.* **70**, 4775–4792 (2019).
55. Schwinn, K. et al. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell* **18**, 831–851 (2006).
56. Goff, S. A., Cone, K. C. & Fromm, M. E. Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor proteins. *Genes Dev.* **5**, 298–309 (1991).
57. Quattrocchio, F. et al. Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color. *Plant Cell* **11**, 1433–1444 (1999).
58. Albert, N. W., Griffiths, A. G., Cousins, G. R., Verry, I. M. & Williams, W. M. Anthocyanin leaf markings are regulated by a family of R2R3-MYB genes in the genus *Trifolium*. *N. Phytol.* **205**, 882–893 (2015).
59. Sainz, M. B., Goff, S. A. & Chandler, V. L. Extensive mutagenesis of a transcriptional activation domain identifies single hydrophobic and acidic amino acids important for activation in vivo. *Mol. Cell. Biol.* **17**, 115–122 (1997).
60. Tiwari, S. B. et al. The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J.* **70**, 855–865 (2012).
61. Gonzalez, A., Zhao, M., Leavitt, J. M. & Lloyd, A. M. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* **53**, 814–827 (2008).
62. Jiang, D., Kong, N. C., Gu, X., Li, Z. & He, Y. *Arabidopsis* COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. *PLoS Genet.* **7**, e1001330 (2011).
63. Odho, Z., Southall, S. M. & Wilson, J. R. Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. *J. Biol. Chem.* **285**, 32967–32976 (2010).
64. Ren, M. et al. Target of rapamycin regulates development and ribosomal RNA expression through kinase domain in *Arabidopsis*. *Plant Physiol.* **155**, 1367–1382 (2011).
65. Son, O., Kim, S., Hur, Y. S. & Cheon, C. I. Molecular details of the Raptor-binding motif on *Arabidopsis* S6 kinase. *Biochem. Biophys. Res. Commun.* **486**, 137–142 (2017).
66. Wang, L. et al. The inhibition of protein translation mediated by AtGCN1 is essential for cold tolerance in *Arabidopsis thaliana*. *Plant Cell Environ.* **40**, 56–68 (2017).
67. Chagne, D. et al. An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. *Plant Physiol.* **161**, 225–239 (2013).
68. Lee, M. M. & Schiefelbein, J. Developmentally distinct MYB genes encode functionally equivalent proteins in *Arabidopsis*. *Development* **128**, 1539 (2001).
69. Yoshida, Y., Sano, R., Wada, T., Takabayashi, J. & Okada, K. Jasmonic acid control of GLABRA3 links inducible defense and trichome patterning in *Arabidopsis*. *Development* **136**, 1039 (2009).
70. Canaguier, A. et al. A new version of the grapevine reference genome assembly (12Xv2) and of its annotation (VCostv3). *Genom. Data* **14**, 56–62 (2017).
71. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
72. Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N. & Delsuc, F. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol. Biol. Evol.* **35**, 2582–2584 (2018).
73. Chernomor, O., von Haeseler, A. & Minh, B. Q. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* **65**, 997–1008 (2016).
74. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
75. Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
76. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
77. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
78. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
79. Hellens, R. P. et al. Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* **1**, 13 (2005).
80. Zhi, J. et al. CRISPR/Cas9-mediated SIAN2 mutants reveal various regulatory models of anthocyanin biosynthesis in tomato plant. *Plant Cell Rep.* **39**, 799–809 (2020).
81. Li, W. et al. Gene expression profiling of development and anthocyanin accumulation in kiwifruit (*Actinidia chinensis*) based on transcriptome sequencing. *PLoS One* **10**, e0136439 (2015).
82. Fraser, L. G. et al. An R2R3 MYB transcription factor determines red petal colour in an *Actinidia* (kiwifruit) hybrid population. *BMC Genom.* **14**, 28 (2013).