Genome-wide identification and characterization of R2R3MYB family in Rosaceae

Máximo González [a,*], Basilio Carrasco [a], Erika Salazar [b]

[a] Facultad de Agronomía e Ingeniería Forestal, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul, Santiago, Chile
[b] Unidad de Recursos Genéticos, CRI La Platina, Instituto de Investigaciones Agropecuarias, Av. Santa Rosa 11610, La Pintana, Santiago, Chile

ARTICLE INFO

ABSTRACT

Transcription factors R2R3MYB family have been associated with the control of secondary metabolites, development of structures, cold tolerance and response to biotic and abiotic stress, among others. In recent years, genomes of Rosaceae botanical family are available. Although this information has been used to study the karyotype evolution of these species from an ancestral genome, there are no studies that treat the evolution and diversity of gene families present in these species or in the botanical family. Here we present the first comparative study of the R2R3MYB subfamily of transcription factors in three species of Rosaceae family (*Malus domestica*, *Prunus persica* and *Fragaria vesca*). We described 186, 98 and 86 non-redundant gene models for apple, peach and strawberry, respectively. In this research, we analyzed the intron–exon structure and genomic distribution of R2R3MYB families mentioned above. The phylogenetic comparisons revealed putative functions of some R2R3MYB transcription factors. This analysis found 44 functional subgroups, seven of which were unique for Rosaceae. In addition, our results showed a highly collinearity among some genes revealing the existence of conserved gene models between the three species studied. Although some gene models in these species have been validated under several approaches, more research in the Rosaceae family is necessary to determine gene expression patterns in specific tissues and development stages to facilitate understanding of the regulatory and biochemical mechanism in this botanical family.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications	
Organism	*Prunus persica*, *Malus domestica* and *Fragaria vesca*
Sex	Not applicable
Sequencer or array type	Genome sequences
Experimental factors	R2R3MYB gene models of three Rosaceae genomes
Experimental features	Gene models of R2R3MYB transcription factors were obtained from functional annotation of each genome project for apple, strawberry and peach. Sequences identified were checked for R2R3 domains through InterProscan. Intron – exon structure was determined using the information available for each genome project. To determine the relationship across Rosaceae genomes and predicted R2R3MYB we analyzed the synteny and collinearity of these genes between the genomes. A Phylogenetic analysis and functional diversity study of R2R3MYB genes in Rosaceae were realized.
Consent	Not applicable
Sample source location	https://phytozome.jgi.doe.gov/pz/portal.html https://www.arabidopsis.org/

* Corresponding author.
E-mail address: mpgonza2@uc.cl (M. González).

1. Introduction

Global sequencing of genomes is an invaluable tool for plant research. Many genomes have been sequenced, facilitating the understanding of the physiology, biochemistry, genetics and evolution of a great number of species [1–3]. Sequencing of *Arabidopsis* genome (The *Arabidopsis* Genome Initiative, 2000) marked the beginning of plant genomics. However, this provided only partial support for the study of phylogenetically distant species. This led to the sequencing of agronomic important species such as rice [4], tomato [3], potato [5] and grape [6], among others. In recent years there has been a general interest to sequence different genomes of the botanical family Rosaceae; the sequences of apple [7], strawberry [8] and peach [9] are currently available. Each of these species is a specific genome model for the subfamilies Malaideae, Rosoideae and Amygdeloideae, respectively, which allows more precise understanding of the biochemical and physiological processes which occur in each of these taxa [1].

The genomes of *Fragaria* (x = 7), *Prunus* (x = 8) and *Malus* (x = 17) evolved from a common ancestor with x = 9 [10]. The evolutionary mechanisms involved in the speciation have been studied with molecular markers [11] and later, more precisely, by the comparative study of

the sequenced genomes [2]. These studies determined that the genus *Malus* underwent a whole duplication event (WGD), and that *Prunus* is the genus with the most conserved karyotype compared to the ancestral karyotype configuration. In addition, high co-linearity between *Pyrus* and *Malus*, and high macrosynteny between *Fragaria* and *Prunus* was described [11].

In spite of this knowledge about genome evolution, there are no analyses which explain the physiological and biochemical differences among these genera. In recent years, studies in grape [12], maize [13], soybean [14] and tomato [15] have attempted to identify the physiological effects of acquisition, duplication, conservation and deletions of the R2R3MYB family of transcription factors. This family has been associated with the control of secondary metabolites [16–19], development of structures [20], opening and closing of stomata [21], cold tolerance [22] phosphorous capture [23], hormone synthesis and cell signaling biotic [24–25] and abiotic stress [22], among others.

The MYB transcription factors are characterized by the presence of the MYB domain, which is composed of four imperfectly repeated sequences (R). Each R domain has about 52 amino acids which may form three α-helix structures [26]. The MYB family is classified into four subgroups according to the number of R domains present in the protein; so far R1, R2R3, R3 and R4 have been described. The R2R3MYB subfamily is the largest in plants; it is characterized by two types of R domain in the N-terminal end and usually by a transcription activator or repressor in the C-terminal end [26].

To improve our understanding of the gene subfamily of R2R3MYB transcription factors in the sequenced genomes of the three most important species of the Rosaceae family: *Malus domestica*, *Prunus persica* and *Fragaria vesca*, we describe the expansion and conservation of clades of the R2R3MYB gene subfamily in each taxon. We also incorporate a functional analysis based on the clustering of R2R3MYB genes of *Arabidopsis*, evolutionarily more recent than Rosaceae family. We determine the chromosome distribution of the predicted models and putative paralog gene models. Our results provide the first comprehensive analysis of the gene structure, gene distribution, gene duplication and putative function of the R2R3MYB family encoded in the Rosaceae genome.

2. Results and discussion

2.1. Identification of R2R3MYB genes in Rosaceae

Gene models of R2R3MYB transcription factors were obtained from functional annotation of each genome project for apple, strawberry and peach. Sequences identified were checked for R2R3 domains through InterProscan (V.5.14), removing sequences with incomplete R2 or R3 domains. Previously, 244; 157; 126; 121; 109 and 108 R2R3MYB genes have been described in maize [13], soybean [14], *Arabidopsis* [16], tomato [15], rice [27] and grape [12], respectively. Our study identified 186, 98 and 86 non-redundant gene models for apple, peach and strawberry, respectively. Both the largest number of predicted genes in sequenced genomes of plants and the greatest number of predicted transcription factors have been reported in apple; more than 400 transcription factors of the MYB family have been identified [7]. Cao et al. mentioned that 222 and 121 models correspond to apple and peach R2R3MYB subfamily, respectively [28]. Our study reduced these values to 16% for apple and 30% for peach updating the number of R2R3MYB transcription factors available in each genome. By contrast, in strawberry only 187 MYB models have been identified [8], and we identified the smallest R2R3MYB subfamily (86 models) described in Rosacea family up to now.

To determine both homologous domain sequences and amino acid frequency in each position within R2 and R3 domains, we performed a multiple alignment analysis for each domain in each species under study. In Fig. 1 we can observe different amino acids frequency for each position of the R2 and R3 domains. Our results confirm the conserved nature of each of these domains. All gene families have three conserved tryptophans in the R2 domain, while only two are present in the R3 domain, where the first tryptophan is always replaced by a hydrophobic amino acid, which is consistent with that was observed in other studies of this gene family such as cucumber [29] and tomato [15].

2.2. Intron – exon structure of the Rosaceae R2R3MYB family and genomic distribution

We found 4, 7.5 and 8% of single exon genes for peach, apple and strawberry, respectively. Exon 1 and 2 usually encode almost the entire R2R3 domain, presenting both exons a uniform size range through the botanical family. In peach, the first exon has an average size of 202 bp, while it reaches 246 bp and 253 bp for strawberry and apple, respectively. Furthermore, exon 2 achieved average size values of 241, 261 and 271 bp for peach, apple and strawberry, respectively (Fig. 2a). Regardless of the species analyzed, exon 3 is the one that has the greatest diversity in sequence length (strawberry, 44–1062 bp; apple, 52–1176 bp; peach, 56–1159 bp). Diversity in sequence length leads into a source of functional divergence due to amino acid changes in motifs and domains. Considering length of the first three exons, we have identified a common pattern of gene diversity in Rosaceae (Fig. 2b) with similar results to those obtained in tomato [15], *Arabidopsis* and grape [12].

Using available genome position reference for each gene model, it was possible to determine that R2R3MYB genes are distributed across all chromosomes (Fig. 3 and Fig. S1). Using this information, we proceeded to obtain specific evolutionary relationships between both species and gene models through synteny analysis. Our results, considering gene model position and sequence similarity, determined that 105 genes are collinear between species (29.09%), and ten syntenic blocks were found (Table S1). Apple chromosomes 9 and 17, as well as 5 and 10, have gene models highly collinear coming from a segmental duplication event. In addition, apple chromosomes 9 and 17 are highly collinear with peach chromosome 3 and strawberry chromosome 6, respectively. In turn, peach chromosomes 1 and 2 are collinear with apple chromosomes 8 and 1, respectively (Fig. 4), being consistent with what was reported by Illa et al. [10].

2.3. Phylogenetic analysis and functional diversity of R2R3MYB genes in Rosaceae family

To identify new clades, expansions and genic conservation, we constructed phylogenetic tree using the neighbor-joining algorithm. Neighbor-joining tree considered all putative genes identified in each genome and all R2R3MYB genes described in *Arabidopsis* (Fig. 5; Fig. S2). The association in clades and subclades of gene models and described genes, has facilitated the identification of possible functions of candidate genes [12], allowing the recognition of similar genes with cooperative or redundant functions. These analyses allowed us the identification of 44 functional clusters suggested by R2R3MYB *Arabidopsis* sequences and, additionally, this strategy allowed us the identification of seven new clusters, identified as subgroup 14 (14S); subgroup 18 (18S); subgroup 22 (22S); subgroup 28 (28S); subgroup 31 (31S); subgroup 40 (40S) and subgroup 44 (44S). Furthermore, gene models inside of each syntenic block mentioned above (Table S1) were found in different subgroups, suggesting that the evolution of this family was through a mechanism different from tandem duplication.

2.3.1. Regulation of plant development and cell destiny

By RNA-seq, gene models proposed in the apple genome have been widely validated, allowing the study of transcriptional profile of primary roots (radicles) [30], auxin-responsive genes [31] and regulatory genes in different stages of fruit development [32]. Specifically, Legay et al. [33] validated by RNA-seq the following R2R3MYB transcription factors: MDP0000655330 (S16), MDP0000787808 (S10), MDP0000149535 (S8), MDP0000320772 and MDP0000228252 (S3), MDP0000197283

a) *Prunus persica*

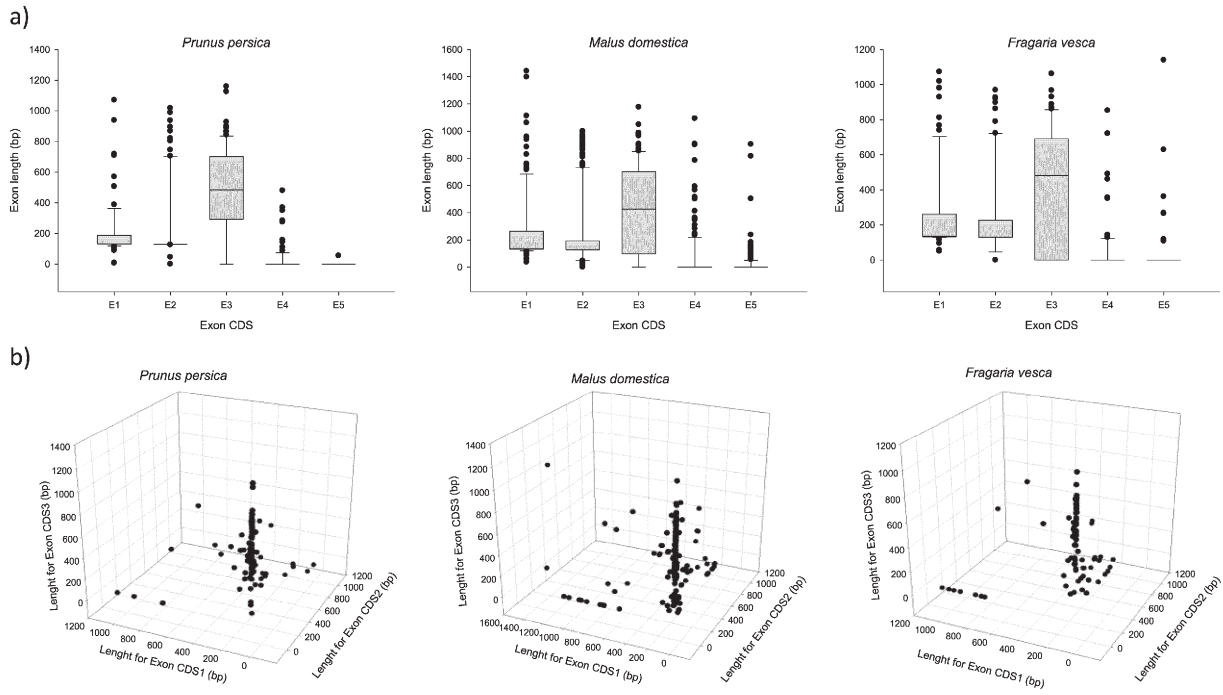

b) *Fragaria vesca*
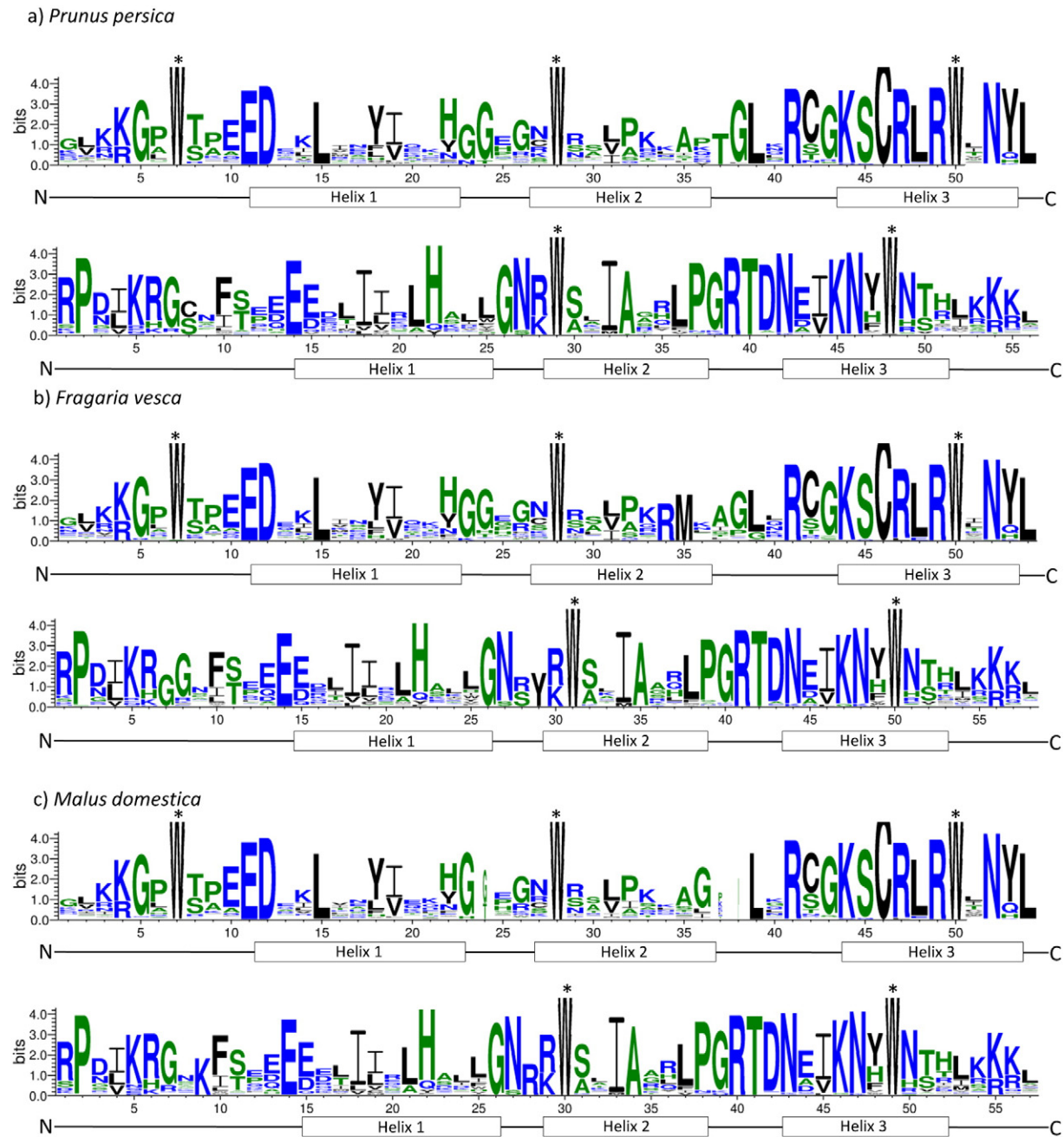
c) *Malus domestica*

Fig. 1. R2 and R3 MYB repeats are highly conserved across all R2R3MYB domain proteins. Sequence logos of the R2 (top) and R3 (bottom) MYB repeats are based on conserved alignments for *Prunus persica* (a), *Fragaria vesca* (b) and *Malus domestica* (c). The overall height of each stack indicates the conservation of the sequence at that position, whereas the height of letters within each stack represents the relative frequency of the corresponding amino acid. Asterisk indicates the position of the conserved amino acid that are identical among R2R3MYB proteins.

(S1), MDP0000291518 and MDP0000852158 (S41), MDP0000322479 (S23), MDP0000157506 (S30). These transcription factors are expressed in apple exocarp, where some of them are expressed differentially in fruit with russet defect characterized by the accumulation of suberin at the inner part of the cell wall of the outer epidermal cell layers. On the other hand, several gene models of R2R3MYB transcription factors have been validated in peach [34–37].

Arabidopsis has many control points in the development of male reproductive structures, which is reflected in the number of MYB factors that participate in this process. It has been reported that AtMYB21, AtMYB24, AtMYB57 (S34) and AtMYB108 (S35) are expressed during pollen and stamen maturation. Factor clustered in subgroup 34 responds to gibberellin to induce jasmonate synthesis [39]; AtMYB24 and AtMYB108 have redundant functions [40] and are downstreamed

from the regulator AtMYB21 [20]. Studies in *Arabidopsis* have shown that AtMYB103 (S25) is closely associated with tapetum, anthers and exine development [41]. Although male structures are highly conserved in angiosperm evolution, we found fewer gene models of subgroup 34 in Rosaceae than the genes described in *Arabidopsis*.

AtMYB61 (S26) has been reported to be a transcription factor associated with photomorphogenic control, mucilage deposition, stomatal aperture, xylem formation and carbon translocation to the roots, among others [42–45]. This group also includes transcription factors AtMYB50, AtMYB55 and AtMYB86, for which no associated functions have been found yet. The number of apple gene models found in subgroup 26 was similar to that described for *Arabidopsis*; however, only one gene was found for peach and strawberry, respectively. In *Arabidopsis*, there was probably an event of genic sub-functionalization

Fig. 2. Exon length distribution analysis of Rosaceae R2R3MYB genes. a) Exon length (Exon 1–Exon 5) values were analyzed using Box plot depicted by SigmaPlot 10.0 on *Prunus persica*, *Malus domestica* and *Fragaria vesca*, respectively. Each box represents the exon size range in which 50% of the values for a particular exon are grouped. The median is represented as a continuous line. b) First, second and third exon lengths in *Prunus persica*, *Malus domestica* and *Fragaria vesca*, respectively.

with respect to peach and strawberry, while apple may have conserved these genes after global genome duplication.

Arabidopsis transcription factor AtMYB5 was found in subgroup 16 (16S). This factor participates in trichome morphogenesis and mucilage synthesis [46]. In this subgroup, we only found one gene model for peach and three for apple. Gene model MDP0000655330 has been previously described in russeted cultivars for apple, characterized by a slightly rough skin phenotype [33].

Subgroup 5 (5S) contains transcription factors AtMYB106/NOK and AtMYB16/MIXTA. These transcription factors participate in trichome development, and the latter determines the form of petals epidermal cells [47]. Recently, ppa023143m has been validated and characterized as a positive regulator of trichome formation in peach fruit, generating the difference among nectarine and peach fruits [37]. Nevertheless, our data indicate that there are a large number of genes in each of the three studied genomes, which reveals strong control over the development of these structures in Rosaceae.

2.3.2. Control of primary and secondary metabolism

Transcription factors AtMYB28, AtMYB29, AtMYB34, AtMYB51, AtMYB76 and AtMYB122 are associated with the synthesis of

glucosinolates [48–51] and always formed a monophyletic clade without interacting directly with any Rosaceae gene model. This clade is related to the evolution of the order Brassicaleae as a function of the adaptive response to herbivory [52].

We obtain consistent and conserved clades with factors AtMYB58 and AtMYB63 (S4), both associated with lignin synthesis during the formation of the secondary cell wall in plant cells [53]. Lignin biosynthesis implies highly complex transcriptional networks [54]. Secondary Wall-Associated NAC Domain Protein 1 (SND1) is a key transcription factor for the formation of the secondary wall, including cellulose and lignin synthesis. AtMYB83 and AtMYB46 are direct targets of SND1, showing functional redundancy in the secondary wall formation cascade [18]. AtMYB46 would also regulate AtMYB52 and AtMYB63 [55], which are transcriptional activators of lignin biosynthesis regulated by homologs of SND1 [53]. AtMYB54 (S41), AtMYB61 (S26) and AtMYB69 (S41) also participate in secondary wall synthesis, in biosynthetic genes activation and in lignification control [40].

The high number of gene models related to these genes may be justified by the structural diversity observed in trees (peach and apple) compared to herbaceous species (strawberry). We must consider the total genome duplication in apple as well as physiological differences;
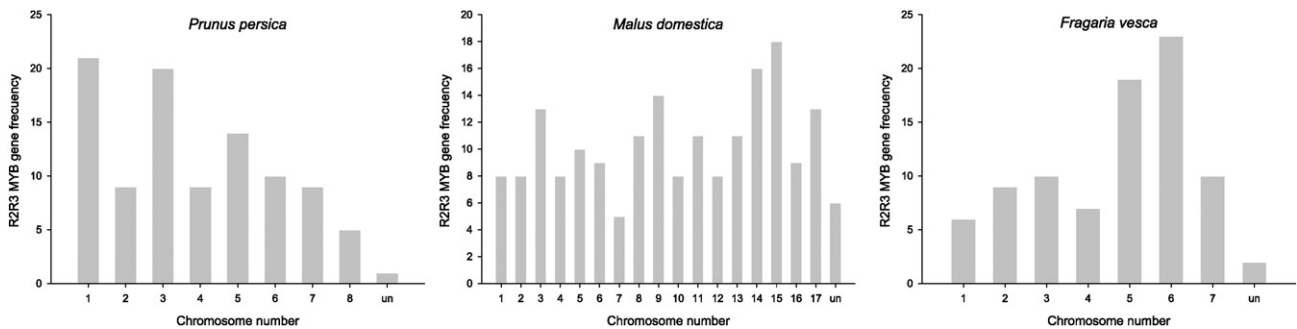
Fig. 3. Chromosome locations of R2R3MYB transcription factors at *Prunus persica*, *Malus domestica* and *Fragaria vesca*. Chromosomal position of each R2R3MYB was mapped according to each genome information.
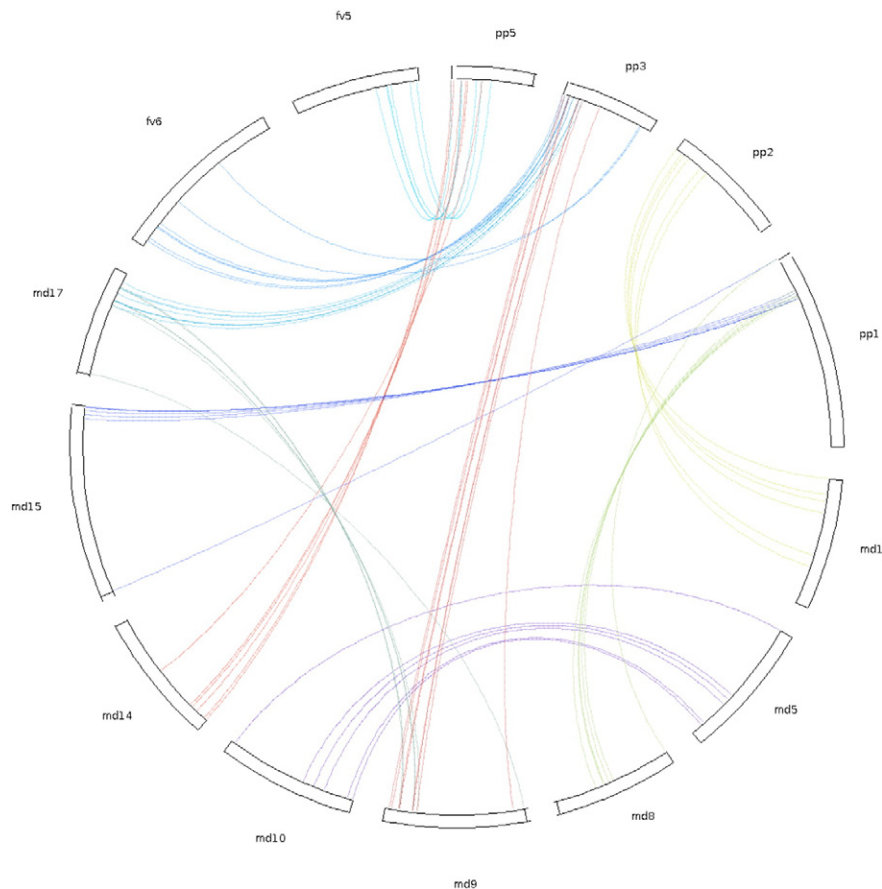
Fig. 4. Schematic representation of R2R3MYB genes synteny in Rosaceae family. The chromosome number is indicated at the top of each chromosome pp.: *Prunus persica*, md: *Malus domestica* and fv: *Fragaria vesca*. Color lines suggest duplicated R2R3MYB gene pairs between chromosomes.

for example, pit formation in peach, where R2R3MYB genes had to change their spatial and temporal expresion patterns to give origin to this particular structure.

Regulation of phenylpropanoid pathway, and specifically, flavonoid pathway, has been well described [26]. The AtMYB11, AtMYB12 and AtMYB111 genes (S12) of *Arabidopsis* are related to the production of flavonol glycosides [56]. This clade is conserved with respect to *Arabidopsis* in the Rosaceae; there were two or three gene models in each analyzed genome (Fig. S2). In peach, ppa004560m (MYB15), ppa010277m (MYB16) and ppa010716m (MYB111) have already been described [34]. While MYB15 is part of the flavonol glycosides clade (S4).
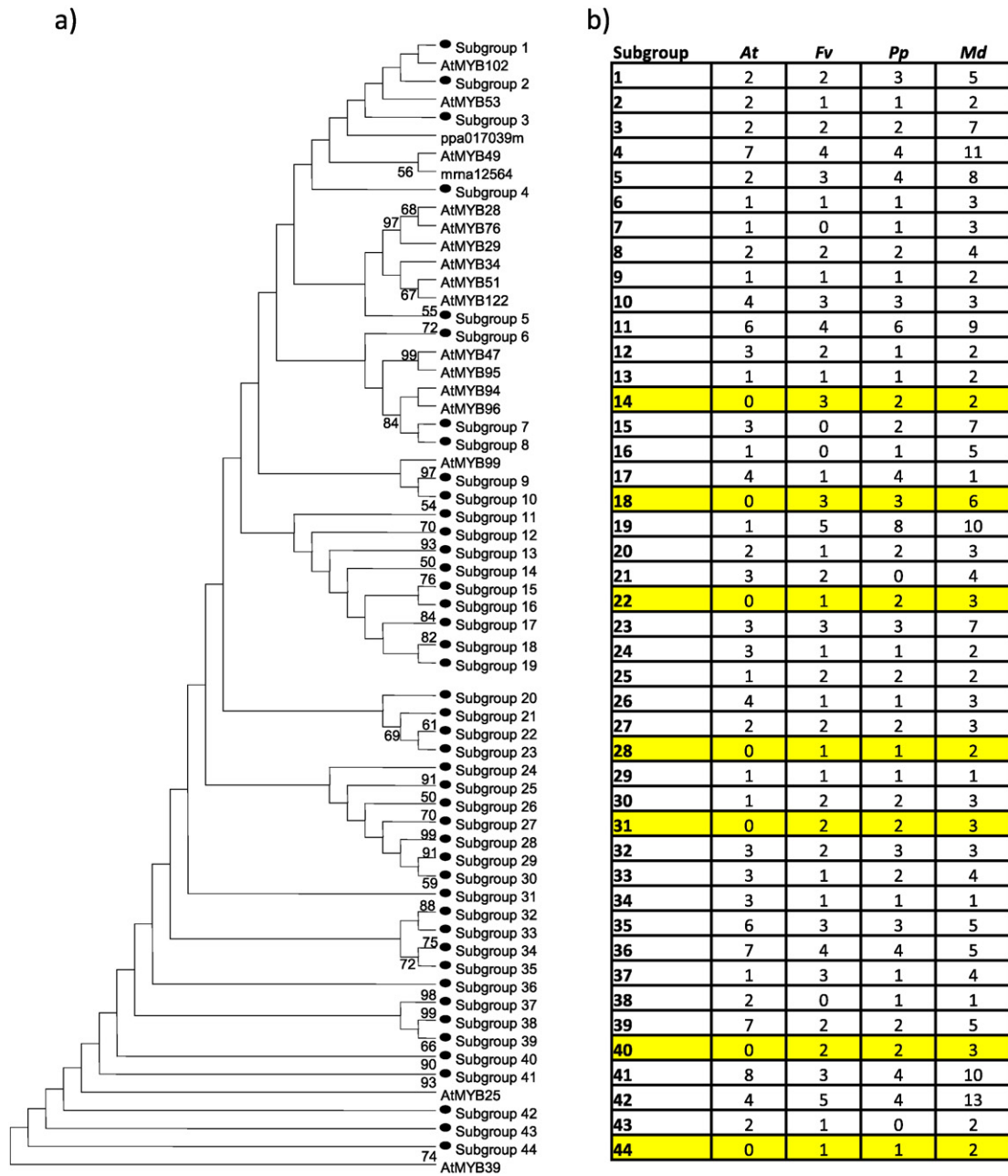
The repressor group (S11) has been extensively described in *Arabidopsis*. Jin et al. [57] and Preston et al. [58] found two protein domains associated with this clade, which were also identified in our alignment. High level of regulation by this subgroup in the phenylpropanoid pathway has been described in *Arabidopsis*. AtMYB32 mutants decrease the expression levels of DFR (dihydrofavonol-4 reductase) and LDOX (leucoanthocianidine dioxygenase) enzymes [57], while AtMYB4 has been associated with the production of cinnamate-4 hydroxylase and photoprotector compounds in response to UV radiation [59-60]. In Rosaceae, studies of repressor MYB transcription factors are scarce. However, FaMYB1 is strongly implicated in the suppression of anthocyanins and the accumulation of flavonols, probably by the regulation of the UFGT (UDP glucose: flavonoid-3-O-glucosyltransferase) and LDOX enzymes [61].

Many studies have been performed to determine genes homologous to AtMYB75/PAP1 and AtMYB90/PAP2 in Rosaceae. These are specific activators of the flavonoid pathway which, when overexpressed, stimulate the production of anthocyanins [62]. Several studies in apple

determined the allelic variants and the regulatory zone of gene MYB10 of Rosaceae [17,19,63,64,65]. This gene is strongly involved in the regulation of structural genes of the metabolic pathway associated with anthocyanin synthesis and thus directly determines their accumulation. In our alignments, genes AtMYB75/PAP1, AtMYB90/PAP2, AtMYB113 and AtMYB114 always formed a specific clade for *Arabidopsis*. In peach, the regulation of flavonoids pathway have been extensively characterized revealing tandem duplication of MYB10 (linkage group 3): ppa026640m (MYB10.1), ppa016711m (MYB10.2) and ppa020385m (MYB10. 3) [36]. Zhou et al. [38] have also validated these gene models-integrating to the analysis- the three transcription factors belonging to the linkage group 6. However, these models do not correspond to R2R3MYB; therefore they have been excluded from our analysis.

Genes AtMYB5 (S16) and AtMYB123/TT2 (S19) showed partially redundant functions in tannin synthesis in *Arabidopsis* seeds [66]. Ravaglia et al. [34] have described ppa023768m (MYB123) and ppa009439m (MYBPA1/subgroup 14). Our study proposes a total of 5, 8 and 10 gene models for MYB123-like, and 3, 2, and 2 for MYBPA1-like in strawberry, peach and apple, respectively. In the genome sequence of strawberry there is an expansion of the clade associated with AtMYB123 [8]. We determined that this expansion occurs in the botanical family probably due, among other reasons, to a response in the formation of fleshy fruit structures which accumulate large amounts of these compounds in the first development stages. It should be noted that neither apples nor strawberries are true fruits, thus the events of gene duplication, which occurred in these subfamilies, would also have facilitated the accumulation of anthocyanins and tannins in other structures.

In conclusion, 186, 98 and 86 non-redundant gene models were identified for apple, peach and strawberry, respectively. Phylogenetic

a)

b)

Subgroup	At	Fv	Pp	Md
1	2	2	3	5
2	2	1	1	2
3	2	2	2	7
4	7	4	4	11
5	2	3	4	8
6	1	1	1	3
7	1	0	1	3
8	2	2	2	4
9	1	1	1	2
10	4	3	3	3
11	6	4	6	9
12	3	2	1	2
13	1	1	1	2
14	0	3	2	2
15	3	0	2	7
16	1	0	1	5
17	4	1	4	1
18	0	3	3	6
19	1	5	8	10
20	2	1	2	3
21	3	2	0	4
22	0	1	2	3
23	3	3	3	7
24	3	1	1	2
25	1	2	2	2
26	4	1	1	3
27	2	2	2	3
28	0	1	1	2
29	1	1	1	1
30	1	2	2	3
31	0	2	2	3
32	3	2	3	3
33	3	1	2	4
34	3	1	1	1
35	6	3	3	5
36	7	4	4	5
37	1	3	1	4
38	2	0	1	1
39	7	2	2	5
40	0	2	2	3
41	8	3	4	10
42	4	5	4	13
43	2	1	0	2
44	0	1	1	2

Fig. 5. a) Phylogenetic relationships and subgroup designations in R2R3MYB proteins from Rosaceae. 44 subgroups were identified in the neighbor-joining tree analysis. 125 protein sequences of *Arabidopsis* (*At*), 86 from strawberry (*Fv*), 98 from peach (*Pp*) and 186 from apple (*Md*) were used. Not condensed neighbor-joining tree is showed in Supplementary Fig. 2. b) Membership of each subgroup is described in the table. Highlighted subgroups are conformed only by Rosaceae sequences.

comparisons grouped gene models into 44 functional subgroups, seven of which were unique for Rosaceae and syntenic analysis determinted 10 syntenic block and 105 collinear gene models across the studied species. Although some gene models in peach and apple have been validated under several approaches, more research in the Rosaceae family is necessary to determine gene expression patterns in specific tissues and development stages for each gene model identified in this study.

3. Materials and methods

3.1. Search for MYBR2R3 transcription factors and sequence conservation analysis

R2R3MYB transcription factors annotated (predicted) for each genome project were downloaded from Phytozome database (www.phytozome.net) [67]. In order to confirm the presence of R2R3MYB domain and verify the reliability of the candidate sequences, all sequences were analyzed using InterProScan V.5.14 [68]. Only those which had complete R2R3 domains were used in order to make the search more precise and obtain models with greater probability of being expressed, thus avoiding pseudogenes. Sequences which share the same chromosome position but are given different names were considered as redundant in the counts of total numbers of MYB transcription factors. On the other hand, 125 *Arabidopsis* R2R3MYB protein sequences were downloaded from TAIR *Arabidopsis* genome (http://www.Arabidopsis.org/). All IDs of selected sequences from Phytozome and TAIR databases are shown in the Supplementary Table 2.

To analyze the features of the R2R3MYB domain of each genome, the sequences of R2 and R3 MYB repeats were aligned with clustalW. Then,

multiple alignment sequences were submitted to WEBLOGO website (http://weblogo.berkeley.edu/logo.cgi).

3.2. Intron – exon structure of the Rosaceae R2R3MYB family and genomic distribution

Intron – exon structure was determined using the information available for each genome project. GFF files (General Feature Format files) of each genome were used to extract the exact position of all intron and exons for each predicted gene model. Then, number and length of introns and exons were calculated. Linkage groups were constructed using the software Mapchart 2.2; while all the graphs were constructed using SigmaPlot version 10.0.

To determine the relationship across Rosaceae genomes and predicted R2R3MYB we analyzed the synteny and collinearity of these genes between the genomes. For this, we used the MCScan software [69]. This software is an algorithm to scan multiple genomes or subgenomes to identify putative homologous chromosomal regions, and then align these regions using genes as anchors, in our case, R2R3MYB transcription factors.

3.3. Phylogenetic analysis and functional diversity of R2R3MYB genes in Rosaceae

Using the dataset generated for each species we performed a phylogenetic analysis along with the 125 R2R3MYB genes described for *Arabidopsis* (www.arabidopsis.org) [70]. Alignments were performed with a BLOSUM matrix (gap opening penalty 25 and gap extension penalty 1) using the ClustalW algorithm (based on the AlignX module) in the MEGA6 program. Phylogenetic analyses used the neighbor-joining algorithm. Confidence in the nodes was estimated by bootstrap sampling with 1000 repetitions. Classification of the R2R3MYB genes was performed according to their phylogenetic relationships with their corresponding *Arabidopsis* R2R3MYB genes.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.06.004.

Conflict of interest

All authors read the manuscript and declare no conflict of interests.

References

[1] V. Shulaev, S.S. Korban, B. Sosinski, A.G. Abbott, H.S. Aldwinckle, K.M. Folta, A. Iezzoni, D. Main, P. Arús, A.M. Dandekar, K. Lewers, S.K. Brown, T.M. Davis, S.E. Gardiner, D. Potter, R.E. Veilleux, Multiple models for Rosaceae genomics. Plant Physiol. 147 (2008) 985–1003.

[2] S. Jung, A. Cestaro, M. Troggio, D. Main, P. Zheng, I. Cho, K.M. Folta, B. Sosinski, A. Abbott, J.M. Celton, P. Arús, V. Shulaev, I. Verde, M. Morgante, D. Rokhsar, R. Velasco, D.J. Sargent, Whole genome comparisons of *Fragaria*, *Prunus* and *Malus* reveal different modes of evolution between Rosaceous subfamilies. BMC Genomics 13 (2012) 129.

[3] M. Zovine, A. Latché, C. Rousseau, F. Regad, J.C. Pech, M. Philippot, et al., The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485 (2012) 635–641.

[4] J. Yu, S. Hu, J. Wang, G.K.S. Wong, S. Li, B. Liu, et al., A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296 (2002) 79–92.

[5] X. Xu, S. Pan, S. Cheng, B. Zhang, D. Mu, P. Ni, et al., Genome sequence and analysis of the tuber crop potato. Nature 475 (2011) 189–195.

[6] O. Jaillon, J.M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449 (2007) 463–467.

[7] R. Velasco, A. Zharkikh, J. Affourtit, A. Dhingra, A. Cestaro, A. Kalyanaraman, et al., The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nat. Genet. 42 (2010) 833–839.

[8] V. Shulaev, D.J. Sargent, R.N. Crowhurst, T.C. Mockler, O. Folkerts, A.L. Delcher, et al., The genome of woodland strawberry (*Fragaria vesca*). Nat. Genet. 43 (2011) 109–116.

[9] R. Ahmad, D. Parfitt, J. Fass, E. Ogundiwin, A. Dhingra, et al., The International Peach Genome Initiative. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat. Genet. 45 (2013) 487–494, http://dx.doi.org/10.1038/ng.2586.

[10] E. Illa, D.J. Sargent, E.L. Girona, J. Bushakra, A. Cestaro, R. Crowhurst, M. Pindo, A. Cabrera, E. van der Knaap, A. Iezzoni, S. Gardiner, R. Velasco, P. Arús, D. Chagné, M. Troggio, Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. BMC Evol. Biol. 11 (2011) 9.

[11] S. Vilanova, D. Sargent, P. Arús, A. Monfort, Synteny conservation between two distantly-related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry). BMC Plant Biol. 8 (2008) 67.

[12] J.T. Matus, F. Aquea, P. Arce-Johnson, Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes. BMC Plant Biol. 8 (2008) 83.

[13] H. Du, B.R. Fen, S.S. Yang, Y.B. Huang, Y.X. Tang, The R2R3-MYB transcription factor gene family in maize. PLoS One 7 (2012), e37463.

[14] H. Du, S.S. Yang, Z. Liang, B.R. Feng, L. Liu, Y.B. Huang, Y.X. Tang, Genome-wide analysis of the MYB transcription factor superfamily in soybean. BMC Plant Biol. 12 (2012) 106.

[15] P. Zhao, Q. Li, J. Li, L. Wang, Z. Ren, Genome-wide identification and characterization of R2R3MYB family in *Solanum lycopersicum*. Mol. Gen. Genomics. 289 (2014) 1183–1207.

[16] R. Stracke, M. Werber, B. Weisshaar, The R2R3-MYB gene family in *Arabidopsis thaliana*. Curr. Opin. Plant Biol. 4 (2001) 447–456.

[17] A.M. Takos, F.W. Jaffé, S.R. Jacob, J. Bogs, S.P. Robinson, A.R. Walker, Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiol. 142 (2006) 1216–1232.

[18] R.L. McCarthy, R. Zhong, Z.H. Ye, MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. Plant Cell Physiol. 50 (2009) 1950–1964.

[19] R.V. Espley, C. Brendolise, D. Chagné, S. Kutty-Amma, S. Green, Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. Plant Cell 21 (2009) 168–183.

[20] A. Mandaokar, J. Browse, MYB108 acts together with MYB24 to regulate jasmonate-mediated stamen maturation in *Arabidopsis*. Plant Physiol. 149 (2009) 851–862.

[21] E. Cominelli, M. Galbiati, A. Vavasseur, L. Conti, T. Sala, M. Vuylsteke, N. Leonhardt, S.L. Dellaporta, C. Tonelli, A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. Curr. Biol. 15 (2005) 1196–1200.

[22] L. Zhang, G. Zhao, J. Jia, X. Liu, X. Kong, Molecular characterization of 60 isolated wheat MYB genes and analysis of their expression during abiotic stress. J. Exp. Bot. 63 (2012) 203–214.

[23] X. Dai, Y. Wang, A. Yang, W.H. Zhang, OsMYB2P-1, an R2R3 MYB transcription factor, is involved in the regulation of phosphate-starvation responses and root architecture in rice. Plant Physiol. 159 (2012) 169–183.

[24] F. Vailleau, X. Danie, M. Tronchet, J.L. Montillet, C. Triantaphylidès, D. Roby, A R2R3-MYB gene, AtMYB30, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack. PNAS 99 (2002) 10179–10184.

[25] H. Shi, R. Cui, B. Hu, X. Wang, S. Zhang, R. Liu, H. Dong, Overexpression of transcription factor AtMYB44 facilitates Botrytis infection in *Arabidopsis*. Physiol. Mol. Plant Pathol. 76 (2011) 90–95.

[26] C. Dubos, R. Stracke, E. Grotewold, B. Weisshaar, C. Martin, L. Lepiniec, MYB transcription factors in *Arabidopsis*. Trends Plant Sci. 15 (2010) 573–581.

[27] C. Yanhui, Y. Xiaoyuan, H. Kun, L. Meihua, L. Jigang, G. Zhaofeng, et al., The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol. Biol. 60 (2006) 107–124.

[28] Z.H. Cao, S.Z. Zhang, R.K. Wang, R.F. Zhang, Y.J. Hao, Genome wide analysis of the apple MYB transcription factor family allows the identification of MdoMYB121 gene confering abiotic stress tolerance in plants. PLoS One 8 (2013), e69955, http://dx.doi.org/10.1371/journal.pone.0069955.

[29] Q. Li, C. Zhang, J. Li, L. Wang, Z. Ren, Genome-wide identification and characterization of R2R3MYB family in *Cucumis sativus*. PLoS One 7 (2012), e47576.

[30] R. Petersen, H. Djozgic, B. Rieger, S. Rapp, E.R. Schmidt, Columnar apple primary roots share some features of the columnar-specific gene expression profile of aerial plant parts as evidenced by RNA-Seq analysis. BMC Plant Biol. 15 (2015) 34.

[31] F. Devoghalaere, T. Doucen, B. Guitton, J. Keeling, W. Payne, T.J. Ling, J.J. Ross, I.C. Hallett, K. Gunaseelan, G. Dayatilake, R. Diak, K.C. Breen, D.S. Tustin, E. Costes, D. Chagné, R.J. Schaffer, K.M. David, A genomics approach to understanding the role of auxin in apple (*Malus* × *domestica*) fruit size control. BMC Plant Biol. 12 (2012) 7.

[32] Y. Bai, L. Dougherty, L. Cheng, G.Y. Zhong, K. Xu, Uncovering co-expression gene network modules regulating fruit acidity in diverse apples. BMC Genomics 16 (2015) 612.

[33] S. Legay, G. Guerriero, A. Deleruelle, M. Lateur, D. Evers, C.M. André, J.F. Hausman, Apple russeting as seen through the RNA-seq lens: strong alterations in the exocarp cell wall. Plant Mol. Biol. 88 (2015) 21–40.

[34] D. Ravaglia, R.V. Espley, R. Henry-Kirk, C. Andreotti, V. Ziosi, R.P. Hellens, G. Costa, A.C. Allan, Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. BMC Plant Biol. 13 (2013) 68.

[35] Y. Chen, Y. Mao, H. Liu, F. Yu, S. Li, T.M. Yin, Transcriptome analysis of differentially expressed genes relevant to variegation in peach flowers. PLoS One 9 (2014), e90842.

[36] M.A. Rahim, N. Busatto, L. Trainotti, Regulation of anthocyanin biosynthesis in peach fruits. Planta 240 (2014) 913–929.

[37] E. Vendramin, G. Pea, L. Dondini, I. Pacheco, M.T. Dettori, L. Gazza, S. Scalabrin, F. Strozzi, S. Tartarini, D. Bassi, I. Verde, L. Rossini, A unique mutation in a MYB gene cosegregates with the nectarine phenotype in peach. PLoS One 9 (2014), e90574.

[38] Y. Zhou, H. Zhou, K. Lin-Wang, S. Vimolmangkang, R.V. Espley, L. Wang, A.C. Allan, Y. Han, Transcriptome analysis and transient transformation suggest an ancient duplicated MYB transcription factor as a candidate gene for leaf red coloration in peach. BMC Plant Biol. 14 (2014) 388.

[39] H. Cheng, S. Song, L. Xiao, H.M. Soo, Z. Cheng, D. Xie, J. Peng, Gibberellin acts through jasmonate to control the expression of MYB21, MYB24, and MYB57 to promote stamen filament growth in *Arabidopsis*. PLoS Genet. 5 (2009), e1000440.

[40] F. Lippold, D.H. Sanchez, M. Musialak, A. Schlereth, W.R. Scheible, D.K. Hincha, M.K. Udvardi, AtMyb41 regulates transcriptional and metabolic responses to osmotic stress in *Arabidopsis*. Plant Physiol. 149 (2009) 1761–1772.

[41] M.D. Gomez, C. Urbez, M.A. Perez-Amador, J. Carbonell, Characterization of constricted fruit (ctf) mutant uncovers a role for AtMYB117/LOF1 in ovule and fruit development in *Arabidopsis thaliana*. PLoS One 6 (2011), e18760.

[42] S. Penfield, R.C. Meissner, D.A. Shoue, N.C. Carpita, M.W. Bevan, MYB61 is required for mucilage deposition and extrusion in the *Arabidopsis* seed coat. Plant 13 (2001) 2777–2791.

[43] L.J. Newman, D.E. Perazza, L. Juda, M.M. Campbell, Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and darphotomorphogenic components of the det3 mutant phenotype. Plant J. 37 (2004) 239–250.

[44] Y.K. Liang, C. Dubos, I.C. Dodd, G.H. Holroyd, A.M. Hetherington, M.M. Campbell, AtMYB61, an R2R3-MYB transcription factor controlling stomatal aperture in *Arabidopsis thaliana*. Curr. Biol. 15 (2005) 1201–1206.

[45] J.M. Romano, C. Dubos, M.B. Prouse, O. Wilkins, H. Hong, M. Poole, K.Y. Kang, E. Li, C.J. Douglas, T.L. Western, S.D. Mansfield, M.M. Campbell, AtMYB61, an R2R3-MYB transcription factor, functions as a pleiotropic regulator via a small gene network. New Phytol. 195 (2012) 774–786.

[46] Y. Sun, Q. Zhou, W. Zhang, Y. Fu, H. Huang, ASYMMETRIC LEAVES1, an *Arabidopsis* gene that is involved in the control of cell differentiation in leaves. Planta 214 (2002) 694–702.

[47] K. Baumann, M. Perez-Rodriguez, D. Bradley, J. Venail, P. Bailey, H. Jin, R. Koes, K. Roberts, C. Martin, Control of cell and petal morphogenesis by R2R3 MYB transcription factors. Development 134 (2007) 1691–1701.

[48] T. Gigolashvili, B. Berger, H.P. Mock, C. Müller, B. Weisshaar, U.I. Flügge, The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. Plant J. 50 (2007) 886–901.

[49] T. Gigolashvili, R. Yatusevich, B. Berger, C. Müller, U.I. Flügge, The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. Plant J. 51 (2007) 247–261.

[50] T. Gigolashvili, M. Engqvist, R. Yatusevich, C. Müller, U.I. Flügge HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. New Phytol. 177 (2008) 627–642.

[51] Y. Chen, X. Yan, S. Chen, Bioinformatic analysis of molecular network of glucosinolate biosynthesis. Comput. Biol. Chem. 35 (2011) 10–18.

[52] C.D. Grubb, S. Abel, Glucosinolate metabolism and its control. Trends Plant Sci. 11 (2006) 89–100.

[53] J. Zhou, C. Lee, R. Zhong, Z.H. Ye, MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. Plant Cell 21 (1) (2009) 248–266.

[54] Q. Zhao, R.A. Dixon, Transcriptional networks for lignin biosynthesis: more complex than we thought? Trends Plant Sci. 16 (2011) 227–233.

[55] J.H. Ko, W.C. Kim, K.H. Han, Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. Plant J. 60 (2009) 649–665.

[56] R. Stracke, H. Ishihara, G. Huep, A. Barsch, F. Mehrtens, K. Niehaus, B. Weisshaar, Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J. 50 (2007) 660–677.

[57] H. Jin, E. Cominelli, P. Bailey, A. Parr, F. Mehrtens, J. Jones, C. Tonelli, B. Weissharr, C. Martin, Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. EMBO J. 19 (2000) 6150–6161.

[58] J. Preston, J. Wheeler, J. Heazlewood, S.F. Li, R.W. Parish, AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. Plant J. 40 (2004) 979–995.

[59] M.R. Hemm, K.M. Herrmann, C. Chapple, AtMYB4: a transcription factor general in the battle against UV. Trends Plant Sci. 6 (2001) 135–136.

[60] M. Todesco, I. Rubio-Somoza, J. Paz-Ares, D. Weigel, A collection of target mimics for comprehensive analysis of microRNA function in *Arabidopsis thaliana*. PLoS Genet. 6 (2010) e1001031.

[61] A. Aharoni, C.H. De Vos, M. Wein, Z. Sun, R. Greco, et al., The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant J. 28 (3) (2001) 319–332.

[62] J.O. Borevitz, Y. Xia, J. Blount, R.A. Dixon, C. Lamb, Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12 (2000) 2383–2393.

[63] Y. Ban, C. Honda, Y. Hatsuyama, M. Igarashi, H. Bessho, T. Moriguchi, Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. Plant Cell Physiol. 48 (2007) 958–970.

[64] R.V. Espley, R.P. Hellens, J. Putterill, D.E. Stevenson, S. Kutty-Amma, A.C. Allan, Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. Plant J. 49 (2007) 414–427.

[65] K. Lin-Wang, K. Bolitho, K. Grafton, A. Kortstee, S. Karunairetnam, T.K. McGhie, R.V. Espley, R.P. Hellens, A.C. Allan, An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. BMC Plant Biol. 10 (2010) 50.

[66] A. Gonzalez, J. Mendenhall, Y. Huo, A. Lloyd, TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. Dev. Biol. 325 (2009) 412–421.

[67] D.M. Goodstein, S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Miltros, W. Dirks, U. Hellsten, N. Putnam, D.S. Rokhsar, Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40 (2012) D1178–D1186.

[68] P. Jones, D. Binns, H.Y. Chang, M. Fraser, W. Li, C. McAnulla, et al., InterProScan 5: genome-scale protein function classification. Bioinformatics 30 (2014) 1236–1240.

[69] Y. Wang, H. Tang, J.D. DeBarry, X. Tan, J. Li, X. Wang, T.H. Lee, H. Jin, B. Marler, H. Guo, J.C. Kissinger, A.H. Paterson, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40 (2012), e49.

[70] E. Huala, A.W. Dickerman, M. Garcia-Hernandez, D. Weems, L. Reiser, F. LaFond, et al., The *Arabidopsis* information resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res. 29 (2001) 102–105.