The *R2R3-MYB* gene family in banana (*Musa acuminata*): genome-wide identification, classification and expression patterns.

Boas Pucker[1], Ashutosh Pandey[1,2], Bernd Weisshaar[1], Ralf Stracke[1,*]

[1] Bielefeld University, Faculty of Biology, Genetics and Genomics of Plants, Sequenz 1, 33615 Bielefeld, Germany

[2] National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi, Delhi 110067, India

[*] communicating author

Abstract

The *R2R3-MYB* genes comprise one of the largest transcription factor gene families in plants, playing regulatory roles in plant-specific developmental processes, defense responses and metabolite accumulation. To date MYB family genes have not yet been comprehensively identified in the major staple fruit crop banana. In this study, we present a comprehensive, genome-wide analysis of the *MYB* genes from *Musa acuminata* DH-Pahang (A genome). A total of 286 *R2R3-MYB* genes as well as genes encoding three other classes of MYB proteins containing multiple MYB repeats were identified and characterised with respect to structure and chromosomal organisation. Organ- an delopment-specific expression patterns were determined from RNA-seq data. For 279 *M. acuminata MYB* genes for which expression was found in at least one of the analysed samples, a variety of expression patterns were detected. The *M. acuminata R2R3-MYB* genes were functionally categorised which led to the identification of seven clades containing only *M. acuminata* R2R3-MYBs. The encoded proteins may have specialised functions that were acquired or expanded in *Musa* during genome evolution. The presented functional classification and expression analyses of the *MYB* gene family in banana establish a solid foundation for future comprehensive functional analysis of MaMYBs and can be utilized in banana improvement programmes.

Keywords: *Musa acuminata*, Zingiberales, R2R3-MYB, transcription factor, gene family

Introduction

Banana (*Musa* spp.), including dessert and cooking types, is a staple fruit crop for a major world population, especially in developing countries. The crop is grown in more than 100 countries throughout the tropics and sub-tropics, mainly in the African, Asia-Pacific, and Latin American and Caribbean regions (Frison and Sharrock 1999). Bananas provide an excellent source of energy and are rich in certain minerals and in vitamins A, C and B6. Furthermore, this perennial, monocotyledonous plant provides an important source of fibre, sugar, starch and cellulose (used for paper, textiles). Bananas have also been considered as a useful tool to deliver edible vaccines (Langridge 2006). Certain agronomic traits, such as stress and pest resistance as well as fruit quality, are of great importance in this plant species. Banana improvement through breeding exercises has been challenging for various reasons. Therefore, genetic engineering based optimisations hold great promise for improvement of this crop plant species. For this purpose, qualified target genes needs to be identified. The release of a high quality banana genome sequence (D'Hont et al. 2012) provides an useful resource to understand functional genomics of important agronomic traits and to identify candidate genes to be utilized in banana improvement programmes.

As almost all biological processes in eukaryotic cells or organisms are influenced by transcriptional control of gene expression, this regulatory level is a good starting point for genetic engineering. Regulatory proteins perform in transcriptional control, alone or complexed with other proteins, by activating or repressing (or both) the recruitment of RNA polymerase to promoters of specific genes (Cooper 2000). These proteins are designated as transcription factors. Connected to their substantial regulatory complexity, transcription factors are numerous and diverse (Riechmann et al. 2000). By binding to specific DNA sequence motifs and regulating gene expression, transcription factors control various regulatory and signaling networks involved in the development, growth and stress response in an organism.

One of the widest distributed transcription factor families in all eukaryotes is the MYB (myeloblastosis) protein family. In the plant kingdom, MYB proteins constitute one of the largest transcription factor families. MYB proteins are defined by a highly conserved MYB DNA-binding domain, mostly located at the N-terminus of the protein. The MYB domain generally consists of up to four imperfect amino acid sequence repeats (R) of about 50-53 amino acids, each forming three alpha–helices (summarised in Dubos et al. 2010). The second and third helices of each repeat build a helix–turn–helix (HTH) structure with three regularly spaced tryptophan (or hydrophobic) residues, forming a hydrophobic core (Ogata et al. 1996). The third helix of each repeat has been identified as the DNA recognition helix that makes

direct contact with DNA (Jia et al. 2004). During DNA contact, two MYB repeats are closely packed in the major groove, so that the two recognition helices bind cooperatively to the specific DNA recognition sequence motif. In contrast to vertebrates, which solely have MYB transcription factors with three repeats, plants have different MYB domain organisations, comprising one to four repeats (Stracke et al. 2001; Dubos et al. 2010). Particularly the number of MYB proteins with two repeats, named R2R3-MYBs according to repeat numbering in vertebrate MYBs, is extremely extended. Gene numbers are ranging from 45 unique *R2R3-MYBs* in *Ginkgo biloba* (Liu et al. 2017) to 360 in Mexican cotton (*Gossypium hirsutum*) (Salih et al. 2016). The expansion of the R2R3-MYB family was coupled with widening in the functional diversity of R2R3-MYBs, considered to regulate mainly plant-specific processes including secondary metabolisms, stress responses and development (Dubos et al. 2010). In this course R2R3-MYBs are involved in several biological traits, such as for example wine quality in grape, fruit color, cotton fibre length, pollinator preferences and nodulation in legumes.

The identification and functional characterization of the *R2R3-MYB* gene family from banana will provide an insight into the regulatory aspects of different biochemical and physiological processes, as those operating during fruit ripening as well as response to various environmental stresses. In this study, we have used genomic resources to systematically identify members of the *M. acuminata* (A genome) *R2R3-MYB* gene family. We used knowledge from other plant species, including the model plant *A. thaliana*, leading to a functional classification of the banana *R2R3-MYB* genes on the basis of phylogenetic analyses. Furthermore, RNA-seq data was used to analyse expression in different *M. acuminata* organs and developmental stages and to compare expression patterns of closely grouped co-orthologs.

Our findings offer the first step towards further investigations on the biological and molecular functions of MYB transcription factors with the selection of genes responsible for economically important traits in *Musa*, which can be utilized in banana improvement programmes.

Material and Methods

Search for MYB protein coding genes in the *M. acuminata* genome

A consensus R2R3-MYB DNA binding domain sequence (Matus et al. 2008) was used as protein query in tBLASTn (Altschul et al. 1990) searches on the *M. acuminata* DH-Pahang

genome sequence (version 2) (http://banana-genome-hub.southgreen.fr/) in an initial search for MYB protein coding genes. To confirm the obtained amino acid sequences, the putative MYB sequences were manually analysed for the presence of an intact MYB domain. All *M. acuminata* MYB candidates from the initial BLAST were inspected to ensure that the putative gene models encode two or more (multiple) MYB repeats. The identified gene models were analysed to map them individually to unique loci in the genome and redundant sequences were discarded from the data set to obtain unique *MaMYB* genes. The identified *MaMYB* genes were matched to the automatically annotated genes from the Banana Genome Hub database (Droc et al. 2013). The open reading frames of the identified *MaMYB* were manually inspected and, if available, verified by mapping of RNA-seq data to the genomic sequence.

Genomic distribution of *MaMYB* genes

The *MaMYB* genes were located on the corresponding chromosomes by the MapGene2chromosome web v2 (MG2C) software tool (http://mg2c.iask.in/mg2c_v2.0/) according to their position information of a physical map, available from the DH-Pahang genome annotation (V2).

Phylogenetic analyses

Protein sequences of 133 *A. thaliana* MYBs were obtained from TAIR (http://www.arabidopsis.org/). We also considered other multiple MYB-repeat proteins from *A. thaliana* in the phylogenetic analysis to determine orthologs in the *M. acuminata* genome: five AtMYB3R, AtMYB4R1 and AtCDC5. Additionally, 43 well-known, functionally characterised landmark plant R2R3-MYB protein sequences were collected from GenBank at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). Phylogenetic trees were constructed from ClustalΩ aligned MYB domain sequences (294 MaMYBs, 132 AtMYBs and 43 plant landmark MYBs) using MEGA7 (Kumar et al. 2016) with default settings. A Maximum Likelihood (ML) consensus tree inferred from 1000 bootstrap replications was calculated. Classification of the *M. acuminata* MYB proteins was performed according to their phylogenetic relationships with corresponding *A. thaliana* and landmark MYB proteins.

Expression analysis from RNA-seq data

136 RNA-Seq data sets were retrieved from the Sequence Read Archive
137 (https://www.ncbi.nlm.nih.gov/sra) via fastq-dump v2.9.6 (https://github.com/ncbi/sra-tools)
138 (Additional Table 2). STAR v2.5.1b (Dobin et al. 2013) was applied for the mapping of reads
139 to the Pahang v2 reference genome sequence (Martin et al. 2016) using previously described
140 parameters (Haak et al. 2018). featureCounts (Liao et al. 2014) was applied for quantification
141 of the mapped reads per gene based on the Pahang v2 annotation. Previously developed
142 Python scripts were deployed for the calculation of normalized expression values
143 (https://github.com/bpucker/bananaMYB) (Haak et al. 2018).
144
145
146 **Results and Discussion**
147 The first annotated reference genome sequence of *M. acuminata* (A genome) has become
148 available in 2012 (D'Hont et al. 2012). It has been obtained from a double haploid (DH) plant
149 of the Pahang cultivar (named after the Malaysia's Pahang province where it was collected),
150 derived through haploid pollen and spontaneous chromosomes doubling from the wild
151 subspecies *Musa acuminata* ssp. *malaccensis* (D'Hont et al. 2012). This wild subspecies was
152 involved in the domestication of the vast majority of cultivated bananas and its genetic
153 signature is commonly found in dessert and cooking bananas (ProMusa,
154 http://www.promusa.org). An improved version (DH-Pahang version 2) of the genome
155 assembly and annotation has been presented in 2016 comprising 450.7 Mb (86% of the
156 estimated size) from which 89.5% are assigned to one of the 11 chromosomes, predicted to
157 contain 35,276 protein encoding genes (Martin et al. 2016). This *M. acuminata* DH-Pahang
158 version 2 genome sequence provides the platform of this study.
159
160 **Identification and genomic distribution of *M. acuminata MYB* genes**
161 A consensus R2R3-MYB DNA binding domain sequence (deduced from Arabidopsis, grape
162 and sugarbeet R2R3-MYBs, Additional Table1) was used as protein query in tBLASTn
163 searches on the DH-Pahang version 2 genome sequence to comprehensively identify MYB
164 protein coding genes in *M. acuminata*. The resulting putative MYB sequences were proven to
165 map to unique loci in the genome and were confirmed to contain an intact MYB domain. This
166 ensured that the gene models contained two or more (multiple) MYB repeats. We identified a
167 set of 286 R2R3-MYB proteins and nine multiple repeat MYB proteins distantly related to the
168 typical R2R3-MYB proteins: six R1R2R3-MYB (MYB3R) proteins, two MYB4R proteins
169 and one CDC5-like protein from the *M. acuminata* genome sequence (Table 1).

170

171 **Table 1: List of annotated *MYB* genes in the *Musa acuminata* (DH Pahang) genome**
172 **sequence.**
173 The genes are ordered by DH Pahang version 2 (Martin et al. 2016) pseudochromosomes,
174 from north to south. The annotation-version specific gene code describing the chromosomal
175 assignment and position on pseudochromosomes is given in the first column. "Ma00"
176 indicates genes located in sequences without chromosomal assignment. Functional
177 assignment is based on the Maximum Likelihood tree presented in Figure 2.
178 str.: strand, *: paralogs, ASR: abiotic stress response, PP: phenylpropanoid, [1] Tak et al. 2017,
179 [2] Negi et al. 2015, [3] Fan et al. 2018, [4] Yang et al. 2015.
180

gene ID	synonym	pseudochr. position from	to	str.	BHLH motif	coding exons	peptide length	MYB type	functional assignment
Ma01_g00440		313923	315583	-		1	297	R2R3	stress response, hormone signaling
Ma01_g02850	MusaMYB31 [1]	1866276	1867435	+	+	2	200	R2R3	repressors PP, sinapate, lignin
Ma01_g04470		3004971	3006821	+		3	230	R2R3	ASR, flower morphogenesis, stilbene
Ma01_g10440		7520585	7526901	+		3	586	R2R3	anther development, stress response
Ma01_g10750		7722630	7723879	-		3	273	R2R3	photomorphogenesis
Ma01_g11890		8618960	8620447	+		2	234	R2R3	secondary cell wall, lignin
Ma01_g12250		8869343	8871737	-		3	334	R2R3	
Ma01_g14370		10499468	10501015	+		3	195	R2R3	defense, stress response
Ma01_g15800		11477431	11479333	+		3	197	R2R3	secondary cell wall, lignin
Ma01_g16960		12408973	12410679	+		3	293	R2R3	
Ma01_g17260		12621562	12624265	-		3	360	R2R3	
Ma01_g17450		12782326	12783615	+		3	270	R2R3	defense, stress response
Ma01_g18470		13704376	13705557	-		3	330	R2R3	suberin
Ma01_g19610	MYB32 [2]	15270989	15272451	-	+	2	241	R2R3	repressors PP, sinapate, lignin
Ma01_g19960		15935559	15937046	-		1	282	R2R3	stress response, hormone signaling
Ma01_g21340		20977295	20978433	-		3	196	R2R3	stamen development
Ma02_g00280		2701936	2703155	+		3	281	R2R3	defense, stress response
Ma02_g00290		4086418	4089233	+		3	320	R2R3	flavonols, phlobaphene
Ma02_g03780		15236704	15238133	+		3	307	R2R3	mucillage, lignin, stomatal closure
Ma02_g04860		16268774	16270113	+		3	325	R2R3	
Ma02_g05880		17024271	17029698	-		7	489	3R	cell cycle control
Ma02_g06190		17224931	17226388	+		3	301	R2R3	axillary meristem, root growth
Ma02_g06670		17595310	17597055	+		3	317	R2R3	suberin
Ma02_g09720	MYB46 [2]	19581503	19582589	+		2	297	R2R3	secondary wall, lignin
Ma02_g09870		19644772	19647000	-		3	381	R2R3	
Ma02_g13370		21804465	21806215	+		3	308	R2R3	photomorphogenesis
Ma02_g15770		23338151	23340844	-		3	444	R2R3	anther development, stress response
Ma02_g16570		23869343	23870392	+		3	297	R2R3	root development

Ma02_g17950	MYB48 [2]	24668074	24669369	-		3	205	R2R3	
Ma02_g19650		25862200	25864721	+	+	3	258	R2R3	
Ma02_g19770	MYB63 [2]	25962423	25963930	-		3	303	R2R3	PP, lignin
Ma02_g20270		26305416	26308011	+		3	365	R2R3	
Ma02_g21230		26925164	26926280	+		2	244	R2R3	general flavonoid, trichome
Ma02_g21760		27303352	27318151	-		11	1076	3R	cell cycle control
Ma02_g22540		27850957	27852253	+		3	175	R2R3	repressors PP, sinapate, lignin
Ma02_g23870		28706468	28707735	+		3	253	R2R3	defense, stress response
Ma02_g24520		29095023	29096192	-		3	292	R2R3	defense, stress response
Ma03_g01260		944639	946045	+		3	124	R2R3	defense, stress response
Ma03_g06410		4438151	4439467	-		3	287	R2R3	root development
Ma03_g07620		5357785	5358916	-		3	205	R2R3	
Ma03_g07840 *		5566056	5567272	-	+	3	282	R2R3	proanthocyanidins
Ma03_g07850 *		5573615	5574872	+	+	3	238	R2R3	proanthocyanidins
Ma03_g08300		5984884	5986166	-		3	287	R2R3	defense, stress response
Ma03_g08930		6570843	6573022	+		3	360	R2R3	embryogenesis, seed maturation
Ma03_g09310		6861251	6862614	+		3	289	R2R3	defense, stress response
Ma03_g09340		6900527	6902243	+		3	319	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma03_g09840		7294462	7296176	+		3	320	R2R3	anther-, tapetum development
Ma03_g11910		9246252	9248138	+		3	360	R2R3	
Ma03_g12480		9622470	9625578	-		4	175	R2R3	defense, stress response
Ma03_g12720		9782618	9784261	-		1	290	R2R3	stress response, hormone signaling
Ma03_g14020		11197595	11199013	-		3	313	R2R3	suberin
Ma03_g18410		23945756	23946927	+		2	168	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma03_g19810		25073643	25074748	+		2	259	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma03_g20390		25556441	25558664	-		3	382	R2R3	embryogenesis, seed maturation
Ma03_g21920		26789930	26791072	+	+	3	254	R2R3	flavonoid repressor
Ma03_g21970		26827699	26828820	+		3	303	R2R3	
Ma03_g23170		27799296	27800913	-	+	3	217	R2R3	repressors PP, sinapate, lignin
Ma03_g25780		29742482	29744122	+		3	344	R2R3	mucillage, lignin, stomatal closure
Ma03_g28720		31825467	31827667	+	+	3	274	R2R3	proanthocyanidins
Ma03_g29070		32109705	32110872	-		3	325	R2R3	
Ma03_g29510		32402433	32403599	-		1	288	R2R3	stress response, hormone signaling
Ma03_g29770		32613488	32614916	+		3	307	R2R3	
Ma04_g00460		405117	406515	+		3	265	R2R3	defense, stress response
Ma04_g01010		896344	897266	-		3	226	R2R3	
Ma04_g05460		4086523	4089539	+		4	290	R2R3	photomorphogenesis
Ma04_g06410		4737137	4738569	+		3	306	R2R3	

Ma04_g09430		6706206	6707827	+		1	302	R2R3	stress response, hormone signaling
Ma04_g11930		8517172	8518919	+		3	246	R2R3	photomorphogenesis
Ma04_g12940		9788762	9790485	+		1	266	R2R3	stress response, hormone signaling
Ma04_g13260		10044518	10045599	+		3	308	R2R3	
Ma04_g16770		16592801	16594124	-		3	348	R2R3	suberin
Ma04_g18740		20887851	20889334	+		3	146	R2R3	ASR, flower morphogenesis, stilbene
Ma04_g19500		22140557	22142266	+		3	347	R2R3	suberin
Ma04_g20120		22833446	22834622	+		3	322	R2R3	anther development, stress response
Ma04_g22200		24575736	24576706	-		2	195	R2R3	repressors PP, sinapate, lignin
Ma04_g22930		25119938	25121516	-		3	254	R2R3	axillary meristem, root growth
Ma04_g23220		25400377	25401463	-		3	243	R2R3	PP, lignin
Ma04_g24670		26638956	26640374	+		3	243	R2R3	
Ma04_g26220		27753147	27754781	+		3	283	R2R3	
Ma04_g26550	MYB85 [2]	27973129	27974181	-		2	249	R2R3	secondary cell wall, lignin
Ma04_g26660		28040878	28042092	-		3	128	R2R3	defense, stress response
Ma04_g26810		28140431	28141878	-		1	269	R2R3	stress response, hormone signaling
Ma04_g28300		29374641	29376644	-	+	3	232	R2R3	
Ma04_g28510		29563724	29565208	+		3	219	R2R3	ASR, flower morphogenesis, stilbene
Ma04_g30160		30889022	30891484	-		3	368	R2R3	anther development, stress response
Ma04_g31800		32024735	32025860	+		3	279	R2R3	axillary meristem, root growth
Ma04_g31880		32082474	32084373	-		3	142	R2R3	photomorphogenesis
Ma04_g32240		32306362	32307006	+		2	192	R2R3	repressors PP, sinapate, lignin
Ma04_g33920	MYB72 [2]	33328174	33329487	+		3	300	R2R3	PP, lignin
Ma04_g34300		33574944	33576372	+		3	382	R2R3	anther development, stress response
Ma04_g34660		33734710	33735657	+		3	226	R2R3	axillary meristem, root growth
Ma04_g35350		34164151	34165124	-		3	281	R2R3	PP, lignin
Ma04_g35730		34354754	34362279	+		4	204	R2R3	PP, lignin
Ma04_g35890		34456195	34457741	-		3	279	R2R3	photomorphogenesis
Ma04_g38740		36138963	36140410	+		3	279	R2R3	axillary meristem, root growth
Ma05_g01100		651244	652185	+		3	206	R2R3	repressors PP, sinapate, lignin
Ma05_g01880		1152994	1154327	-		3	281	R2R3	axillary meristem, root growth
Ma05_g03340		2404252	2405715	+		3	303	R2R3	PP, lignin
Ma05_g03690		2718181	2719844	-		3	240	R2R3	ASR, flower morphogenesis, stilbene
Ma05_g05670		4316913	4318043	-		3	246	R2R3	secondary cell wall, lignin
Ma05_g06310		4706038	4707394	-	+	2	375	R2R3	general flavonoid, trichome
Ma05_g07140		5206984	5207790	-		1	269	R2R3	stress tolerance
Ma05_g07450		5428289	5429962	-		3	295	R2R3	defense, stress response
Ma05_g08960		6598602	6600986	+		3	192	R2R3	

Ma05_g10430		7526267	7527710	+		3	279	R2R3	axillary meristem, root growth
Ma05_g12030		8749406	8751293	+		3	255	R2R3	flower meristem identity
Ma05_g14510		10595097	10596227	-		3	229	R2R3	ASR, flower morphogenesis, stilbene
Ma05_g17720		21202263	21203407	-		3	234	R2R3	
Ma05_g18420		23864457	23865647	+		3	241	R2R3	repressors PP, sinapate, lignin
Ma05_g18710		24607776	24609429	-		3	288	R2R3	defense, stress response
Ma05_g19630		28141902	28143134	-		3	295	R2R3	root development
Ma05_g20320		31975488	31976981	+		3	307	R2R3	suberin
Ma05_g20740		32405856	32407834	+		3	375	R2R3	
Ma05_g20940		32642123	32643636	+		3	333	R2R3	
Ma05_g23480		35547417	35548389	+		3	182	R2R3	defense, stress response
Ma05_g23640		35794588	35797244	-		3	321	R2R3	flavonols, phlobaphene
Ma05_g24200		36482627	36483579	+		3	207	R2R3	axillary meristem, root growth
Ma05_g24840		36981509	36982629	+		3	323	R2R3	anther-, tapetum development
Ma05_g25150		37166044	37167618	+		3	316	R2R3	axillary meristem, root growth
Ma05_g25490		37422954	37423941	-		3	226	R2R3	secondary cell wall, lignin
Ma05_g25630		37499737	37501266	-		3	290	R2R3	trichome branching, petal morphogenesis
Ma05_g25680		37532195	37533810	+		3	302	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma05_g28370		39401272	39403071	-	+	3	210	R2R3	repressors PP, sinapate, lignin
Ma05_g30120		40637050	40638724	+		3	252	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma05_g30720		40993150	40994479	+		3	237	R2R3	cell cycle regulation
Ma05_g31160		41198403	41202707	-		2	397	R2R3	stress tolerance
Ma05_g31440		41356205	41357951	+		1	344	R2R3	leaf-, shoot-, germ morphogenesis
Ma06_g00910		745494	747844	+		4	394	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma06_g03570		2604323	2605964	+		3	333	R2R3	axillary meristem, root growth
Ma06_g04210		3057847	3059452	-	+	3	361	R2R3	proanthocyanidins
Ma06_g04240		3073024	3074684	-		3	345	R2R3	trichome branching, petal morphogenesis
Ma06_g04270		3094694	3096881	+		3	354	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma06_g04370		3152384	3153807	+		3	250	R2R3	defense, stress response
Ma06_g05680		4227417	4228614	+		3	305	R2R3	root development
Ma06_g05960		4396733	4397886	-	+	3	275	R2R3	anthocyanins
Ma06_g06660		4800702	4808542	-		7	517	3R	cell cycle control
Ma06_g08100		5748553	5749986	+		3	276	R2R3	embryogenesis, seed maturation
Ma06_g08440		5976662	5977609	-		3	243	R2R3	
Ma06_g08910		6239004	6240845	+	+	3	199	R2R3	repressors PP, sinapate, lignin
Ma06_g11140	MaMYB3 [3]	7827947	7829034	+	+	3	200	R2R3	starch degradation, flavonoid repressor
Ma06_g11270		7905151	7906898	+		3	308	R2R3	defense, stress response
Ma06_g12110		8407734	8411775	-		12	469	R2R3	guard cell division, root gravitropism

Ma06_g12160	8449752	8451062	-		3	256	R2R3	defense, stress response
Ma06_g14470	9916014	9917324	-	+	3	286	R2R3	repressors PP, sinapate, lignin
Ma06_g16350	11053372	11054799	+	+	2	290	R2R3	general flavonoid, trichome
Ma06_g16920	11470718	11471708	+		3	220	R2R3	cell cycle regulation
Ma06_g17440	11851409	11852643	-		3	275	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma06_g19030	13020068	13021561	+		2	263	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma06_g27210	29243106	29244308	-		3	237	R2R3	repressors PP, sinapate, lignin
Ma06_g29060	30553182	30554508	+		3	195	R2R3	stamen development
Ma06_g31020	32236972	32238507	+		3	351	R2R3	defense, stress response
Ma06_g32530	33447706	33449189	+		3	315	R2R3	axillary meristem, root growth
Ma06_g33100	33858221	33872663	+		12	838	4R	SNAP complex
Ma06_g33190	33914323	33915299	+		3	244	R2R3	
Ma06_g33430 *	34072591	34074225	+		3	399	R2R3	axillary meristem, root growth
Ma06_g33440 *	34079035	34083709	+		2	335	R2R3	axillary meristem, root growth
Ma06_g33920	34371133	34372942	+		3	346	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma06_g35430	35258465	35260108	+		1	225	R2R3	stress tolerance
Ma06_g35620	35399156	35400631	+		4	400	R2R3	embryogenesis, seed maturation
Ma06_g37660	36662931	36664139	+		3	299	R2R3	anther-, trichome development
Ma06_g38880	37506759	37508514	-		3	311	R2R3	anther-, trichome development
Ma07_g00270	249167	250371	+		2	326	R2R3	
Ma07_g02470	1968475	1973718	-		3	600	R2R3	anther development, stress response
Ma07_g05660	4115493	4116032	+		1	272	R2R3	defense, stress response
Ma07_g05780	4199983	4201888	+		2	251	R2R3	secondary cell wall, lignin
Ma07_g08110	6059901	6061028	+		3	256	R2R3	axillary meristem, root growth
Ma07_g10330	7698294	7710182	+		7	568	3R	cell cycle control
Ma07_g10340	7710102	7712664	-		1	332	R2R3	leaf-, shoot-, germ morphogenesis
Ma07_g11110	8262980	8264078	-		3	1286	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma07_g12330	9212596	9214045	+		3	272	R2R3	secondary cell wall, lignin
Ma07_g13590	10215552	10217192	-	+	2	307	R2R3	general flavonoid, trichome
Ma07_g17600	20758827	20760547	+		3	343	R2R3	
Ma07_g19350	27376391	27377767	-		3	270	R2R3	defense, stress response
Ma07_g19470	27474531	27475523	+		3	241	R2R3	cell cycle regulation
Ma07_g19700	27655370	27656957	-		3	470	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma07_g19720	27682847	27684927	+		3	357	R2R3	trichome branching, petal morphogenesis
Ma07_g19880 *	27793461	27794847	+	+	3	264	R2R3	proanthocyanidins
Ma07_g19890 *	27798123	27799369	-	+	3	269	R2R3	proanthocyanidins
Ma07_g20020	27925622	27926695	+		4	270	R2R3	secondary cell wall, lignin
Ma07_g20990	28973981	28975550	-		3	341	R2R3	axillary meristem, root growth

Ma07_g22540		30450165	30451931	-		3	269	R2R3	defense, stress response
Ma07_g23060		30808440	30823676	+		4	551	R2R3	anther development, stress response
Ma07_g23180		30931923	30933477	+		3	289	R2R3	defense, stress response
Ma07_g23230 *		30951474	30953174	+		3	274	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma07_g23240 *		30953852	30959687	+		5	1124	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma07_g26530		33294249	33295626	-		3	292	R2R3	
Ma08_g01300		1210739	1212318	-		3	263	R2R3	defense, stress response
Ma08_g01760		1466320	1467246	+		1	249	R2R3	stress tolerance
Ma08_g02100		1707379	1708993	+		3	421	R2R3	embryogenesis, seed maturation
Ma08_g02450		1904208	1906264	+		3	241	R2R3	
Ma08_g03420		2489605	2490995	+		3	319	R2R3	anther-, trichome development
Ma08_g10260		7475666	7478250	-		3	371	R2R3	flavonols, phlobaphene
Ma08_g10600		7758560	7759645	-		3	256	R2R3	defense, stress response
Ma08_g11120		8204459	8205747	+		3	278	R2R3	photomorphogenesis
Ma08_g12510		9469637	9470606	+		3	231	R2R3	PP, lignin
Ma08_g13070		10389004	10390770	-		3	313	R2R3	axillary meristem, root growth
Ma08_g14720		14652239	14654273	+		3	370	R2R3	
Ma08_g15820		16034463	16037767	-	+	3	260	R2R3	
Ma08_g15960		16625932	16636910	-		3	572	R2R3	
Ma08_g16760		20724394	20725390	-	+	3	194	R2R3	flavonoid repressor
Ma08_g17860		27301469	27302427	-		3	249	R2R3	defense, stress response
Ma08_g18540		32070622	32071809	-		4	276	R2R3	defense, stress response
Ma08_g23390		36792752	36794121	+		3	290	R2R3	proanthocyanidins
Ma08_g25570		38350574	38352398	+		3	307	R2R3	axillary meristem, root growth
Ma08_g25960	MYBS3 [4]	38629094	38631076	+		1	296	R2R3	stress response, hormone signaling
Ma08_g26720		39209905	39211361	-		3	541	R2R3	
Ma08_g30360		41652638	41653604	+		3	269	R2R3	defense, stress response
Ma08_g31720	MYB83 [2]	42548860	42549971	+		2	279	R2R3	secondary wall, lignin
Ma08_g32760		43364665	43366044	+		3	200	R2R3	repressors PP, sinapate, lignin
Ma08_g34230		44310373	44311394	-	+	2	203	R2R3	repressors PP, sinapate, lignin
Ma08_g34710		44706406	44707610	+		3	245	R2R3	
Ma09_g03310		2240424	2242599	-		1	342	R2R3	leaf-, shoot-, germ morphogenesis
Ma09_g03740		2480875	2484428	+		2	322	R2R3	stress tolerance
Ma09_g04930		3159017	3160626	-		3	302	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma09_g06730		4303577	4304528	+		3	204	R2R3	stamen development
Ma09_g08140		5353765	5354763	+		2	256	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma09_g08260		5451761	5453265	-	+	3	272	R2R3	repressors PP, sinapate, lignin
Ma09_g09400		6192674	6193933	-		3	263	R2R3	

Ma09_g09720		6397365	6410516	-		13	798	4R	SNAP complex
Ma09_g10800		7342785	7344298	-		3	327	R2R3	anther-, trichome development
Ma09_g11770		8000781	8002111	-		3	262	R2R3	defense, stress response
Ma09_g13170		8908572	8910010	-		3	310	R2R3	axillary meristem, root growth
Ma09_g14260		9744056	9745210	-		3	187	R2R3	root development
Ma09_g15050		10356702	10357805	-		3	248	R2R3	secondary cell wall, lignin
Ma09_g15130		10456782	10469060	+		5	211	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma09_g15440		10764111	10765364	-		3	257	R2R3	proanthocyanidins
Ma09_g15940		11297144	11298342	-		3	260	R2R3	defense, stress response
Ma09_g16940		12437966	12439652	+		3	279	R2R3	defense, stress response
Ma09_g16980		12476466	12478547	+		3	304	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma09_g20280		28997934	28999432	+		3	273	R2R3	axillary meristem, root growth
Ma09_g22730		34631674	34632781	-		3	315	R2R3	
Ma09_g23100		35024297	35025651	+		1	288	R2R3	stress response, hormone signaling
Ma09_g23770		35511400	35518711	+		5	1120	2R	CDC5
Ma09_g24640	MYB31 [2]	36291635	36292658	-		2	173	R2R3	repressors PP, sinapate, lignin
Ma09_g25010		36615281	36616548	+		3	301	R2R3	defense, stress response
Ma09_g25590		36999254	37000706	-		3	315	R2R3	anther-, trichome development
Ma09_g27990		38852455	38853644	-	+	3	260	R2R3	anthocyanins
Ma09_g28970		39592459	39593487	+		1	343	R2R3	leaf-, shoot-, germ morphogenesis
Ma09_g29010		39617312	39618287	-		3	214	R2R3	
Ma09_g29660	MYB52 [2]	40061201	40062117	-		3	210	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma10_g01730		5102443	5104125	-		3	286	R2R3	defense, stress response
Ma10_g01750		5158925	5160589	+		3	288	R2R3	defense, stress response
Ma10_g04420		15095580	15096722	-		3	284	R2R3	
Ma10_g04920		15562002	15563726	-		3	321	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma10_g05260		15989100	15989831	+		1	243	R2R3	stress tolerance
Ma10_g05680		17016063	17017845	+		3	499	R2R3	embryogenesis, seed maturation
Ma10_g06140		17599985	17600776	+		3	212	R2R3	
Ma10_g09100		23305076	23306673	-		3	297	R2R3	anther-, trichome development
Ma10_g09370		23565723	23566809	-		3	268	R2R3	axillary meristem, root growth
Ma10_g10820		24568694	24569501	+		3	176	R2R3	repressors PP, sinapate, lignin
Ma10_g11100		24705798	24706830	-		2	260	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma10_g13000		25941801	25943526	+		3	304	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma10_g13640		26378268	26379297	-		3	236	R2R3	cell cycle regulation
Ma10_g14150		26690021	26691729	-	+	2	298	R2R3	general flavonoid, trichome
Ma10_g14950		27210210	27223227	+		11	869	3R	cell cycle control
Ma10_g16050		27917256	27918773	+	+	3	217	R2R3	repressors PP, sinapate, lignin

Ma10_g17650	28964249	28965577	+	+	3	278	R2R3	anthocyanins
Ma10_g18840	29620788	29624706	+		12	467	R2R3	guard cell division, root gravitropism
Ma10_g19130	29806295	29807639	-		3	297	R2R3	root development
Ma10_g19820	30233785	30235049	-		3	175	R2R3	defense, stress response
Ma10_g19970	30311359	30312430	-	+	3	206	R2R3	flavonoid repressor
Ma10_g24510	33076381	33078033	-		3	305	R2R3	root development
Ma10_g25660	33694709	33695874	+		3	319	R2R3	root development
Ma10_g26540	34187286	34196507	+		7	566	3R	cell cycle control
Ma10_g26660	34247740	34248867	+		2	226	R2R3	
Ma10_g29230	35877570	35878873	-		3	315	R2R3	defense, stress response
Ma10_g29290	35933157	35937337	-		3	386	R2R3	anther development, stress response
Ma10_g29660	36170110	36171732	-		3	289	R2R3	trichome branching, petal morphogenesis
Ma10_g29900	36334717	36336097	+		3	285	R2R3	secondary cell wall, lignin
Ma11_g00330	235199	236300	-		3	235	R2R3	
Ma11_g00350	255253	257212	+		3	375	R2R3	
Ma11_g02310	1659706	1660645	-		3	213	R2R3	PP, lignin
Ma11_g03860	2954679	2956968	-		3	294	R2R3	flower meristem identity
Ma11_g04680	3650617	3651708	+		3	264	R2R3	defense, stress response
Ma11_g06880	5504819	5505826	-		2	167	R2R3	repressors PP, sinapate, lignin
Ma11_g07330	5826847	5827933	+		3	261	R2R3	axillary meristem, root growth
Ma11_g07530	6014287	6015836	-		3	319	R2R3	suberin
Ma11_g08730	6941295	6943798	-		3	275	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma11_g10680	10271323	10272648	-		3	275	R2R3	defense, stress response
Ma11_g10710	10304401	10305535	+		3	275	R2R3	
Ma11_g11300	12744754	12745994	+		3	251	R2R3	defense, stress response
Ma11_g11940	15469297	15470702	-		3	315	R2R3	defense, stress response
Ma11_g14670	20385599	20386631	-		3	262	R2R3	defense, stress response
Ma11_g15740	21412142	21413141	-		3	230	R2R3	PP, lignin
Ma11_g16150	21712282	21713968	-		3	233	R2R3	
Ma11_g16430	21944276	21945346	+		3	287	R2R3	PP, lignin
Ma11_g19220	24164773	24166212	-		3	344	R2R3	defense, stress response
Ma11_g21160	25411481	25412732	-		3	303	R2R3	
Ma11_g21730	25754193	25755462	-		3	326	R2R3	suberin
Ma11_g21820	25812612	25817014	-	+	3	235	R2R3	
Ma11_g23010	26544673	26545555	-		2	241	R2R3	cell cycle regulation
Ma11_g23420	26767577	26769029	+		3	183	R2R3	cell wall, lignin, seed oil, axillary meristem
Ma00_g01590	9744102	9745946	+		1	283	R2R3	stress response, hormone signaling
Ma00_g04340	36355056	36356431	+		2	283	R2R3	secondary wall, lignin

Ma00_g04960	43098831	43099818	-	3	267	R2R3	defense, stress response

The number of R2R3-MYB genes is one of the highest among the species that have been studied to date, ranging from 45 in *Ginkgo biloba* (Liu et al. 2017) over 157 in *Zea mays* (Du et al. 2012a) and 249 in *Brassica napus* (Hajiebrahimi et al. 2017) to 360 in *Gossypium hirsutum* (Salih et al. 2016). This is probably due to three whole-genome duplications (γ 100 Myr ago and α, β 65 Myr ago) that occured during *Musa* genome evolution (D'Hont et al. 2012; Wu et al. 2016). The number of atypical multiple repeat *MYB* genes identified in *M. acuminata* is in the same range as those reported for most other plant species, with for example up to six *MYB3R* and up to two *MYB4R* and *CDC5*-like genes.

The identified 286 *MaR2R3-MYB* genes, constituting approximately 0.81 % of the 35,276 predicted protein-coding *M. acuminata* genes and 9.1 % of the 3,155 putative *M. acuminata* transcription factor genes (Martin et al. 2016), were subjected to further analyses. The identified *MaR2R3-MYB* genes were named following the nomenclature of the locus tags provided in the DH-Pahang version 2 genome annotation (Table 1). A keyword search in the NCBI database (http://www.ncbi.nlm.nih.gov/) revealed no evidence for *M. acuminata MYB* genes not present in Table 1.

Six publications dealing with *M. acuminata MYB* genes were identified: one study describes the elevated expression of nine *MaMYB* genes in transgenic banana plants overexpressing the NAC domain transcription factor MusaVND1 (vascular related NAC domain) indicating a role of these MaMYBs in the regulation of secondary wall deposition (Negi et al. 2015). MYBS3 *(Ma08_25960)* was found to be differentially expressed between cold-sensitive and cold-tolerant bananas (Yang et al. 2015). Another publication described the *M. acuminata R2R3-MYB* gene *Ma05_03690* being upregulated in the early response to the endoparasitic root-knot nematode *Meloidogyne incognita* in roots (Castañeda et al. 2017). MusaMYB31 *(Ma01_02850)* was identified as a negative regulator of lignin biosynthesis and the general phenylpropanoid biosynthesis pathway (Tak et al. 2017). MaMYB3 *(Ma06_11140)* was found to repress starch degradation in fruit ripening (Fan et al. 2018) and MaMYB4 *(Ma01_19610)* was recently described to control fatty acid desaturation (Song et al. 2019).

On the basis of the DH-Pahang version 2 annotation, 292 of the 295 *MaMYB* genes could be assigned to the eleven chromosomes. The chromosomal distribution of *MaMYB* genes on the pseudochromosomes is shown in Figure 1 and revealed that *M. acuminata MYB* genes are distributed across all chromosomes.

Gene structure analysis revealed that most *MaR2R3-MYB* genes (226 of 286; 79 %) follow the previously reported rule of having two introns and three exons, and display the highly conserved splicing arrangement that has also been reported for other plant species (Du et al. 2012a; Du et al. 2012b; Stracke et al. 2014). A total of 29 (10.1 %) *MaR2R3-MYB* genes have two exons and 19 (6.6 %) were one exon genes. Eight (2.8 %) *MaR2R3-MYB* genes have four exons, and there are two of each kind with five or twelve exons, respectivily. The complex exon-intron structure of Ma06_12110 and Ma10_18840 is known from their *A. thaliana* orthologs AtMYB88 and AtMYB124/FOUR LIPS (FLP) containing ten and eleven exons, respectively. This supports their close evolutionary relationship, but also indicates the conservation of this intron pattern in evolution since the monocot-dicot split 140-150 Myr ago (Chaw et al. 2004).

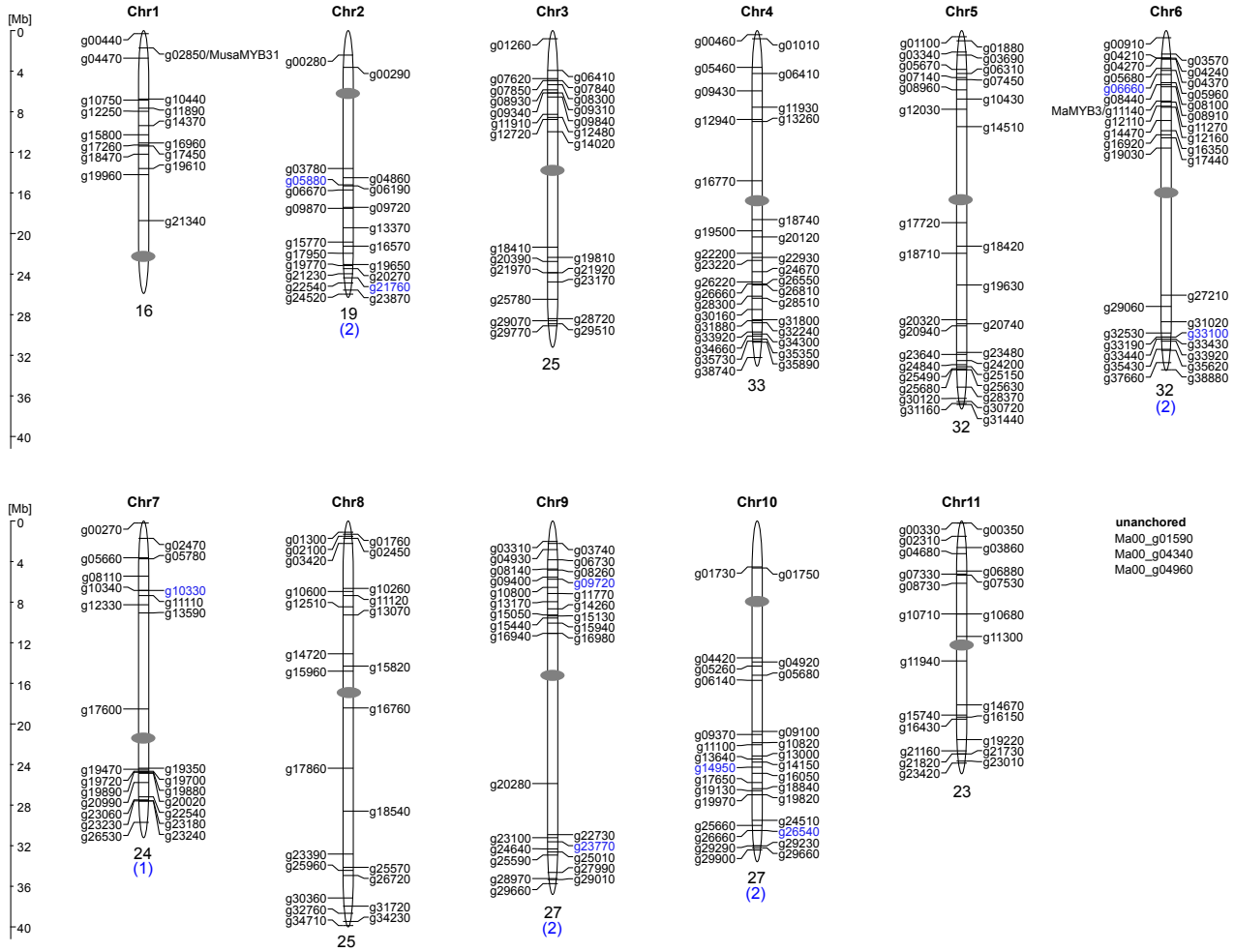

Figure 1: Distribution of *MaMYB* genes on the eleven *M. acuminata* chromosomes. Chromosomes are drawn to scale. The positions of centromeres (grey ovals) are roughly estimated from repeat distribution data. The chromosomal positions of the *MaMYB* genes (given in DH Pahang version 2 annotation ID) are indicated. *R2R3-MYB* genes are given in

230 black letters, *MYB* genes with more than two MYB repeats are given in blue letters. The
231 number of *MYB* genes on each chromosome is given below the respective chromosome.
232
233
234 **Phylogenetic analysis of the *M. acuminata* MYB family**
235 With the aim to explore the putative function of the predicted *M. acuminata* MYBs, we
236 assigned them to plant MYB proteins with known function. For this, we chose primarily data
237 from *A. thaliana*, as most of our knowledge about plant MYB genes has been obtained from
238 studies of this major plant model. As known from comparable studies, most MYB proteins
239 with similar functions cluster in the same phylogenetic clades, suggesting that closely related
240 MYBs recognise similar/same target genes and possess cooperative, overlapping or redundant
241 functions.
242 To unravel the relationships, we constructed a phylogenetic tree with 469 MYB domain
243 amino acid sequences of MYB proteins. We used 294 MaMYBs (omitting the CDC5-like
244 MaMYB), the complete *A. thaliana* MYB family (132 members, including 126 R2R3-MYB,
245 five MYB3R and one MYB4R) and 43 functionally well characterised landmark R2R3-
246 MYBs from other plant species. The phylogenetic tree topology allowed us to classify the
247 analysed MYBs into one MYB3R clade, one MYB4R clade and 42 R2R3-MYB protein
248 clades (Figure 2).
249 Most R2R3-MYB clades (31 of 42) include variable numbers of MYB proteins from
250 Arabidopsis and banana, indicating that the appearance of most *MYB* genes in these two
251 species predates the monocot-dicot split as observed in other studies (Du et al. 2015). Several
252 of these clades also contain landmark MYBs from other plant species. The two clades 4 and
253 28 only contain Arabidopsis R2R3-MYB members, while seven clades (9, 13, 19, 21, 26, 29
254 and 32, displayed with white background in Figure 2) only contain banana R2R3-MYB
255 members. Additionally, two clades, 24 with monocot anthocyanin biosynthesis regulators and
256 27 with the strawberry landmark flavonoid biosynthesis repressor FaMYB1 (Aharoni et al.
257 2001), do not contain any *A. thaliana* MYB (Figure 2).
258

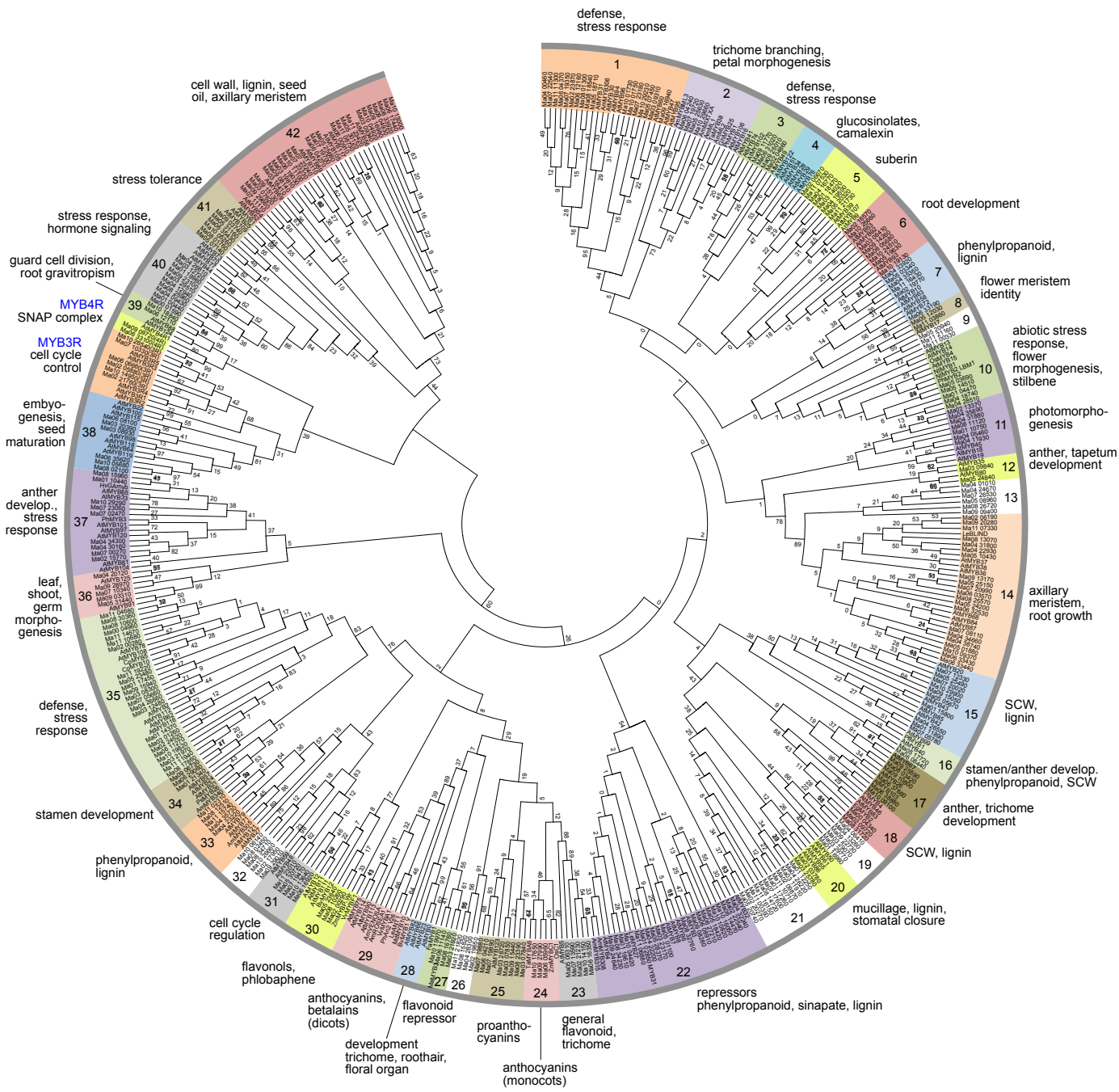

Figure 2: Phylogenetic Maximum Likelihood (ML) tree (consensus tree inferred from 1000 bootstraps) with 466 MYB domain amino acid sequences of MYB proteins from *Musa accuminata* (Ma), *Arabidopsis thaliana* (At) and landmark MYBs from other plant species built with MEGA7. Subgroups are labeled with different colors and functional annotations are given. The numbers at the branches give bootstrap support values from 1000 replications. SCW, secondary cell wall

The observation of "species/order-specific" clades could be taken as a hint that *MYB* genes in these clades may have been acquired or been lost in a single order or species, during the following divergence from the most recent common ancestor, as indicated also by other studies (Baum 2008; Zhong et al. 2015; Mehta et al. 2016). One MYB clade without

271 *M. acuminata* orthologs was clade 28, that include the *A. thaliana* R2R3-MYBs
272 AtMYB0/GLABRA1 and AtMYB66/WEREWOLF, which have been identified being
273 involved in the formation of trichomes and root hairs from epidermal cells (Oppenheimer et
274 al. 1991; Lee and Schiefelbein 1999). Similar observations have been made in maize
275 (monocot) and sugarbeet (eudicot, caryophyllales) (Du et al. 2012b; Stracke et al. 2014), both
276 not containing clade 28 orthologs, while grape (eudicot, rosid) and poplar (eudicot, rosid) do
277 (Matus et al. 2008; Wilkins et al. 2009). It has been hypothesized, that GLABRA1-like *MYB*
278 genes have been acquired in rosids after the rosid-asterids split (Brockington et al. 2013). The
279 absence of MaMYBs in clade 28 is consistent with this hypothesis, since monocots branched
280 off before the separation of asterids and rosids in eudicots. Another clade without banana
281 R2R3-MYBs is clade 4, covering the glucosinolate biosynthesis regulators AtMYB28,
282 AtMYB34 and AtMYB51 (Sonderby et al. 2007; Gigolashvili et al. 2007). The absence of
283 MaMYBs in this clade is in good accordance with the fact that glucosilates are only present in
284 the Brassicaceae family. This clade is thought to have originated from a duplication event
285 before the divergence of Arabidopsis from Brassica (Yanhui et al. 2006).
286 The seven clades containing only MaMYBs were manually inspected by applying BLAST
287 searches at the NCBI protein database in order to identify high homology to functionally
288 characterized landmark plant R2R3-MYBs. In no case landmark MYB could be identified.
289 Consequently, these clades could be described as a lineage-specific expansion in
290 *M. acuminata*, reflecting a species-, genius- or order-specific adaptation. These MaMYB
291 proteins may have specialised functions that were acquired or expanded in *M. acuminata*
292 during genome evolution. Further research will be needed to decipher the biological roles of
293 these *MaMYB* genes.
294 It is well established, that R2R3-MYBs could interact with basic helix-turn-helix (BHLH)-
295 type transcription factors, together with WD-repeat (WDR) proteins forming a so called
296 trimeric MBW complex. These R2R3-MYBs are defined by a BHLH-binding consensus
297 motif [D/E]Lx2[R/K]x3Lx6Lx3R (Zimmermann et al. 2004) found in all BHLH-interacting
298 R2R3-MYBs. A search in the MaMYB proteins for the mentioned BHLH-interaction motif
299 identified 30 MaMYBs containing this motif (Table 1) and thus putatively interacting with
300 BHLH proteins. 36 of these 40 MaMYBs were all functionally assigned to clades containing
301 (potentially) known BHLH-interacting R2R3-MYBs: nine in clade 22 (repressors
302 phenylpropanoid, sinapate-, lignin regulators), six in clade 25 (proanthocyanin regulators),
303 four in clade 23 (general flavonoid, trichome regulators), four in clade 27 (flavonoid
304 repressors) and three in clade 24 (monocot anthocyanin regulators). Four MaMYBs,

Ma02_19650, Ma04_28300, Ma08_15820 and Ma11_21820 were all found in the *M. acuminata*-specific clade 26. We suppose that these MaMYBs interact with BHLH-type transcription factors. The close relation to flavonoid biosynthesis regulators, suggest that they act similar like R2R2-MYB proanthocyanin regulators. If they regulate a Musa-specific biosynthesis pathway needs further research.

Expression profiles for *M. acuminata MYB* genes in different organs and developmental stages

Since it is not unusual for large transcription factor families in higher organisms, that some members have redundant function, a particular transcription factor needs to be studied and characterised in the context of the whole family. In this regard, the gene expression pattern can provide a hint to gene function. In this study we used public available RNA-seq reads available at the Sequence Read Archive (Additional Table 2) to analyse the expression of the 294 *MaMYB* genes in different organs and developmental stages: embryonic cell suspension, seedling, root, young, adult and old leaf, pulp stages S1-S4, peel stages S1-S4. Filtered RNA-seq reads were aligned to the genome reference sequence and the number of mapped reads per annotated transcript were quantified and compared across the analysed samples giving normalised RNA-seq read values which are given in Table 2.

Table 2: The expression profiles of *MaR2R3-MYB* genes in different organs and developmental stages based on RNA-seq data. The intensity of green background is correlated to expression level given in FPKM.

gene	embryogenic	seedling	root	leaf	young_leaf	adult_leaf	old_leaf	pulp	pulpS1	pulpS2	pulpS3	pulpS4	peel	peelS1	peelS2	peelS3	peelS4
Ma00_g01590	50	35	6	151	77	175	348	6	5	5	8	5	8	2	19	10	2
Ma00_g04340	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Ma00_g04960	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma01_g00440	181	49	38	129	53	175	276	32	31	58	34	7	30	9	63	35	11
Ma01_g02850	9	2	14	27	21	31	38	20	24	23	23	10	11	7	9	17	12
Ma01_g04470	1	0	12	4	1	6	2	9	13	9	7	8	6	6	3	7	7
Ma01_g10440	3	16	4	5	7	5	5	4	5	5	3	3	3	3	2	3	4
Ma01_g10750	0	0	2	2	0	0	4	1	1	1	1	2	4	5	4	4	5
Ma01_g11890	0	0	3	3	10	1	0	0	0	0	0	0	0	0	0	0	0
Ma01_g12250	1	0	4	5	17	2	0	3	4	3	2	2	4	4	4	3	3
Ma01_g14370	0	17	11	0	0	0	0	5	7	7	8	0	8	3	7	18	4
Ma01_g15800	1	0	4	1	0	0	0	4	2	5	7	3	6	8	5	2	9
Ma01_g16960	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Ma01_g17260	1	0	11	1	0	0	0	3	3	3	2	3	4	5	3	4	5
Ma01_g17450	0	0	6	2	0	1	6	5	13	2	2	1	1	1	2	2	0
Ma01_g18470	0	0	0	0	0	0	0	0	1	0	0	0	1	2	1	0	0
Ma01_g19610	13	0	6	39	56	40	57	29	78	22	15	2	12	7	36	2	5
Ma01_g19960	35	1	6	97	94	148	145	5	7	6	5	4	19	5	58	10	4

Ma01_g21340	0	0	0	0	0	0	0	2	8	0	0	0	0	0	0	0	0
Ma02_g00280	1	1	21	4	1	1	5	6	8	10	4	2	5	5	4	5	7
Ma02_g00290	0	0	3	9	5	26	1	8	25	0	6	2	3	2	3	2	5
Ma02_g03780	0	0	29	18	26	0	0	13	14	9	18	12	30	39	31	17	33
Ma02_g04860	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Ma02_g05880	4	2	2	3	2	5	3	2	4	2	2	1	2	1	1	1	3
Ma02_g06190	0	0	19	8	0	0	0	16	30	12	13	11	28	42	18	18	32
Ma02_g06670	1	1	4	1	0	0	0	4	3	7	4	4	5	6	2	6	5
Ma02_g09720	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Ma02_g09870	29	1	23	37	99	33	11	19	15	28	27	6	11	11	10	9	13
Ma02_g13370	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma02_g15770	6	0	1	0	0	1	0	1	2	0	1	0	1	1	1	0	1
Ma02_g16570	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
Ma02_g17950	8	0	46	14	4	6	13	33	53	35	16	26	19	19	14	20	22
Ma02_g19650	0	0	3	3	5	1	0	7	10	5	3	11	9	15	8	6	5
Ma02_g19770	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma02_g20270	3	0	3	16	57	4	0	1	1	2	2	1	4	3	4	3	7
Ma02_g21230	2	2	1	5	14	3	1	3	8	3	0	0	1	0	3	2	0
Ma02_g21760	0	3	7	2	2	0	0	8	11	7	8	7	8	11	5	5	12
Ma02_g22540	1	0	7	4	2	6	7	1	1	1	2	0	1	1	1	1	1
Ma02_g23870	3	0	0	7	4	20	3	2	8	0	0	0	0	0	0	0	0
Ma02_g24520	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Ma03_g01260	11	0	6	16	0	0	61	9	31	3	1	2	2	1	3	4	1
Ma03_g06410	0	0	11	1	0	0	0	1	0	3	0	1	1	2	0	2	1
Ma03_g07620	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma03_g07840	0	0	2	1	2	0	0	10	24	9	3	5	2	1	3	1	2
Ma03_g07850	0	0	1	1	1	0	0	1	2	1	0	1	2	2	2	3	1
Ma03_g08300	0	1	3	0	0	0	1	1	5	0	0	0	1	0	0	6	0
Ma03_g08930	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma03_g09310	0	0	0	12	42	7	1	0	0	0	0	0	0	0	0	0	0
Ma03_g09340	2	0	7	1	1	0	1	5	8	4	5	2	3	4	2	3	5
Ma03_g09840	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma03_g11910	5	8	25	35	76	25	6	12	10	13	12	13	24	26	30	18	22
Ma03_g12480	0	0	4	0	0	0	0	1	1	1	0	0	0	0	0	1	0
Ma03_g12720	16	10	4	45	59	49	70	2	2	2	2	3	18	2	32	35	3
Ma03_g14020	0	0	1	0	0	0	0	2	1	4	1	3	2	3	1	3	2
Ma03_g18410	0	0	0	1	2	0	0	4	15	0	0	0	0	0	0	0	0
Ma03_g19810	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma03_g20390	5	1	1	2	2	0	4	2	1	3	2	2	0	1	0	0	1
Ma03_g21920	5	0	2	17	13	32	22	4	12	2	2	0	2	0	6	3	0
Ma03_g21970	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma03_g23170	0	0	3	4	8	2	0	2	2	3	1	2	4	6	2	3	4
Ma03_g25780	3	2	60	18	14	3	0	37	28	81	18	22	40	57	29	26	48
Ma03_g28720	3	3	1	2	0	1	5	1	3	1	0	1	0	0	0	0	0
Ma03_g29070	0	0	1	2	4	2	0	1	1	0	3	0	2	1	2	1	2
Ma03_g29510	14	4	5	44	49	65	55	9	25	5	5	2	6	3	7	11	3
Ma03_g29770	1	0	3	1	1	0	0	2	3	4	0	2	1	2	1	1	2
Ma04_g00460	2	0	1	9	14	20	2	0	0	0	0	0	1	0	2	1	0
Ma04_g01010	0	0	7	0	0	0	1	1	3	1	1	0	1	1	1	0	1
Ma04_g05460	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1
Ma04_g06410	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
Ma04_g09430	4	1	9	12	17	19	2	4	4	2	6	5	8	11	7	3	11
Ma04_g11930	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma04_g12940	4	0	2	3	6	2	4	2	3	1	2	0	1	0	1	0	0
Ma04_g13260	0	2	2	1	1	1	2	2	3	2	3	1	0	0	0	0	0
Ma04_g16770	0	1	1	1	0	1	0	1	0	1	0	3	1	1	0	1	0
Ma04_g18740	11	1	7	2	1	1	5	2	0	0	6	0	1	0	1	2	0
Ma04_g19500	0	0	4	1	0	0	0	2	4	3	0	1	1	1	1	1	0
Ma04_g20120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma04_g22200	0	0	3	0	0	0	0	3	2	8	1	3	1	1	0	1	0
Ma04_g22930	0	0	4	1	0	0	0	2	2	3	0	2	4	6	2	3	4
Ma04_g23220	0	0	5	3	0	0	1	1	3	1	1	1	4	3	6	1	6
Ma04_g24670	0	0	33	6	0	0	0	24	32	24	19	19	17	22	9	16	19
Ma04_g26220	0	1	1	2	1	1	0	1	3	1	1	0	4	2	11	2	1
Ma04_g26550	0	0	5	1	1	0	0	2	3	3	1	1	1	1	1	1	3
Ma04_g26660	0	2	22	1	0	1	3	12	27	11	5	4	1	0	2	1	2
Ma04_g26810	41	78	15	153	258	145	185	39	35	74	32	17	33	13	54	50	14
Ma04_g28300	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma04_g28510	8	1	3	11	2	22	10	9	13	8	14	2	3	2	5	3	3
Ma04_g30160	1	0	0	0	1	1	0	2	5	1	0	0	0	0	0	0	0
Ma04_g31800	0	0	0	1	0	0	3	0	1	1	0	0	2	0	1	6	0

Ma04_g31880	0	0	2	0	0	0	0	1	0	1	1	0	1	1	2	0	1
Ma04_g32240	0	0	1	0	0	0	0	2	2	4	2	0	0	1	0	0	1
Ma04_g33920	3	7	2	2	4	1	1	0	1	1	0	0	1	1	1	3	1
Ma04_g34300	1	0	0	1	2	1	0	2	4	1	1	2	1	1	0	0	1
Ma04_g34660	0	1	0	0	0	0	0	1	0	0	0	3	0	0	0	0	0
Ma04_g35350	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma04_g35730	0	0	4	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Ma04_g35890	1	0	0	0	0	0	0	1	2	1	0	0	0	1	0	1	0
Ma04_g38740	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1
Ma05_g01100	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Ma05_g01880	0	0	2	1	0	0	0	2	3	2	0	2	2	1	1	2	2
Ma05_g03340	0	0	2	1	1	0	1	1	0	2	0	0	0	0	0	0	0
Ma05_g03690	10	0	1	14	2	25	28	1	0	1	1	1	2	0	8	1	0
Ma05_g05670	1	0	1	0	0	0	0	1	2	0	0	0	0	0	0	1	0
Ma05_g06310	4	1	1	4	4	4	8	1	3	1	1	0	0	0	1	0	0
Ma05_g07140	4	0	0	0	0	0	0	0	1	1	1	0	0	0	0	2	0
Ma05_g07450	4	0	0	2	4	3	1	1	2	1	1	0	0	0	0	0	0
Ma05_g08960	27	1	9	34	47	53	23	17	38	11	8	10	9	10	3	9	12
Ma05_g10430	0	0	2	1	0	0	0	2	1	1	2	3	3	6	2	1	4
Ma05_g12030	4	1	2	1	1	1	2	3	8	2	1	2	2	2	2	1	1
Ma05_g14510	1	0	4	48	0	1	190	1	0	1	2	0	1	1	1	3	0
Ma05_g17720	0	0	1	0	0	0	0	0	1	0	0	0	4	6	4	2	4
Ma05_g18420	0	0	0	0	0	1	0	1	4	0	0	0	0	0	0	0	0
Ma05_g18710	4	0	3	3	4	4	5	3	5	2	3	0	2	1	4	2	1
Ma05_g19630	0	0	6	0	0	0	0	1	2	3	0	1	1	1	0	2	0
Ma05_g20320	1	0	2	10	5	14	19	0	0	1	0	0	1	1	0	1	0
Ma05_g20740	0	4	1	2	9	0	0	2	5	1	2	0	1	0	2	1	0
Ma05_g20940	1	0	5	11	0	0	44	1	1	0	1	0	0	1	0	0	1
Ma05_g23480	0	0	0	0	0	0	0	2	3	5	0	0	0	0	0	0	1
Ma05_g23640	0	0	1	1	0	3	0	2	7	0	2	0	2	1	3	0	2
Ma05_g24200	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Ma05_g24840	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma05_g25150	0	0	6	2	0	0	0	3	4	4	5	1	3	3	2	2	4
Ma05_g25490	0	1	2	2	1	3	1	1	2	2	1	1	1	2	1	1	2
Ma05_g25630	0	0	0	4	15	0	1	1	6	0	0	0	0	0	0	0	0
Ma05_g25680	1	0	1	2	2	1	3	7	24	2	3	1	0	0	1	1	0
Ma05_g28370	0	0	4	4	9	4	1	1	0	1	0	1	2	2	1	1	2
Ma05_g30120	0	3	16	1	0	1	0	4	7	4	2	2	2	2	1	2	2
Ma05_g30720	0	0	19	1	0	0	0	7	12	10	2	4	3	3	3	1	6
Ma05_g31160	9	3	5	6	7	7	7	5	7	4	4	3	2	2	3	2	4
Ma05_g31440	0	3	2	68	266	2	0	43	12	88	71	2	26	2	89	9	2
Ma06_g00910	0	7	1	5	1	1	15	0	0	0	0	0	1	1	1	3	0
Ma06_g03570	0	0	4	2	0	0	0	3	5	3	3	4	5	7	4	3	7
Ma06_g04210	0	0	5	2	1	0	0	5	5	2	7	4	3	4	3	2	5
Ma06_g04240	1	0	1	1	4	0	0	7	29	0	0	0	0	0	0	0	0
Ma06_g04270	11	1	1	1	0	1	2	3	6	3	2	1	0	0	1	0	0
Ma06_g04370	7	0	10	13	5	29	15	1	3	0	0	0	2	1	6	1	0
Ma06_g05680	0	0	7	4	0	0	0	20	15	34	15	16	8	9	2	11	10
Ma06_g05960	0	0	6	4	4	0	0	13	9	19	13	9	4	4	3	9	2
Ma06_g06660	3	2	2	3	2	3	6	2	3	1	2	2	1	1	1	1	1
Ma06_g08100	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma06_g08440	3	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Ma06_g08910	4	3	5	28	68	23	16	8	6	11	9	5	10	4	20	14	3
Ma06_g11140	4	0	1	14	6	33	16	2	4	2	0	1	1	0	4	1	0
Ma06_g11270	0	1	10	0	0	0	0	0	1	0	0	0	1	0	1	3	0
Ma06_g12110	0	1	1	1	0	1	1	1	2	1	2	1	1	1	1	1	2
Ma06_g12160	1	1	1	1	1	2	2	1	2	0	1	1	1	0	0	2	1
Ma06_g14470	0	0	3	1	1	1	2	2	1	2	2	2	1	0	1	3	0
Ma06_g16350	0	1	0	3	4	2	5	1	2	1	1	2	1	0	0	1	0
Ma06_g16920	3	1	40	38	30	13	13	60	108	51	34	46	54	72	36	45	61
Ma06_g17440	0	0	0	0	0	0	0	1	1	1	0	1	1	0	1	1	2
Ma06_g19030	0	0	2	0	0	0	1	3	8	1	1	1	1	1	0	1	0
Ma06_g27210	2	2	8	6	2	6	16	3	4	5	0	2	2	2	1	2	2
Ma06_g29060	9	2	0	0	0	0	1	2	7	1	0	0	0	0	0	0	0
Ma06_g31020	7	3	7	6	3	9	8	1	2	2	2	0	2	1	7	1	1
Ma06_g32530	7	0	0	2	1	4	2	0	0	0	0	0	0	0	0	0	1
Ma06_g33100	4	6	3	3	2	3	3	4	5	3	2	6	3	4	2	2	5
Ma06_g33190	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Ma06_g33430	0	1	11	2	1	1	0	8	20	7	2	4	3	5	3	4	3
Ma06_g33440	38	9	21	26	12	27	53	24	36	21	23	15	14	14	9	11	20
Ma06_g33920	0	7	10	4	3	0	1	5	5	4	3	6	10	13	8	9	9

Ma06_g35430	6	15	28	31	49	14	31	20	37	22	13	8	33	24	63	25	17
Ma06_g35620	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Ma06_g37660	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma06_g38880	1	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0
Ma07_g00270	0	0	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0
Ma07_g02470	5	1	4	5	5	5	7	5	6	7	2	3	3	3	2	2	4
Ma07_g05660	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0
Ma07_g05780	0	1	1	3	6	2	1	2	3	2	2	1	2	1	1	2	2
Ma07_g08110	0	1	5	1	0	0	4	0	1	0	0	0	0	0	0	1	0
Ma07_g10330	2	4	4	1	1	1	2	2	1	2	1	2	2	2	1	1	3
Ma07_g10340	2	11	6	23	82	3	2	11	22	10	8	5	14	8	15	28	5
Ma07_g11110	0	0	0	3	5	5	1	0	0	0	0	0	0	0	0	0	0
Ma07_g12330	0	0	1	7	5	18	3	1	2	1	0	0	0	0	1	0	0
Ma07_g13590	0	0	1	6	4	16	3	0	1	0	0	0	0	0	0	0	0
Ma07_g17600	1	1	8	2	4	0	1	1	1	1	0	1	2	2	2	2	2
Ma07_g19350	2	0	1	22	18	44	23	4	15	0	0	1	9	1	26	11	0
Ma07_g19470	3	0	15	7	5	6	6	54	80	62	33	42	10	14	6	9	10
Ma07_g19700	0	0	0	0	0	0	0	1	4	0	0	0	0	0	0	0	0
Ma07_g19720	0	0	7	7	26	1	0	5	11	6	3	1	4	2	12	1	1
Ma07_g19880	1	1	5	4	0	0	2	10	14	7	11	8	9	12	3	8	12
Ma07_g19890	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
Ma07_g20020	2	0	1	0	0	1	1	0	0	1	0	0	0	0	1	1	0
Ma07_g20990	0	0	2	0	0	0	0	2	2	1	3	0	2	2	1	1	1
Ma07_g22540	5	1	0	8	13	12	7	1	3	0	0	0	1	0	3	1	0
Ma07_g23060	0	1	2	1	2	0	0	2	4	2	3	1	1	2	1	1	1
Ma07_g23180	3	1	1	3	8	1	0	4	14	1	1	0	3	4	2	1	5
Ma07_g23230	0	171	9	2	2	1	3	34	6	10	12	106	9	1	2	33	2
Ma07_g23240	16	23	7	9	9	12	11	9	7	10	12	6	5	6	3	3	7
Ma07_g26530	0	1	1	1	0	2	1	1	4	0	0	0	0	0	0	0	0
Ma08_g01300	6	0	3	6	12	3	7	0	0	0	0	0	1	0	0	2	0
Ma08_g01760	6	0	0	2	2	2	4	0	1	0	0	0	1	0	1	1	1
Ma08_g02100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma08_g02450	1	3	5	2	2	1	3	7	7	10	9	1	5	3	13	2	3
Ma08_g03420	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma08_g10260	1	0	3	7	16	3	1	4	12	1	1	3	13	18	6	7	19
Ma08_g10600	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0
Ma08_g11120	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
Ma08_g12510	0	0	3	1	0	0	0	2	3	2	2	0	2	2	1	1	5
Ma08_g13070	0	0	10	4	0	0	0	17	12	30	18	8	15	19	7	12	21
Ma08_g14720	0	1	5	5	14	1	0	2	4	1	3	2	6	8	6	2	7
Ma08_g15820	0	1	4	2	1	1	2	6	7	6	6	4	4	5	4	5	2
Ma08_g15960	2	6	3	1	2	1	1	4	5	1	2	5	2	2	1	1	2
Ma08_g16760	2	0	2	10	5	13	21	3	5	3	0	3	3	1	4	7	0
Ma08_g17860	0	0	1	2	0	0	8	0	0	0	0	0	2	0	0	7	0
Ma08_g18540	0	0	0	1	0	1	2	0	0	0	0	0	0	0	0	2	0
Ma08_g23390	0	0	0	0	0	0	0	2	6	0	0	0	0	0	1	0	0
Ma08_g25570	7	1	16	11	7	6	7	20	20	26	25	10	19	16	19	16	23
Ma08_g25960	9	11	9	21	22	32	15	13	18	17	11	5	9	9	9	5	13
Ma08_g26720	1	0	7	9	19	4	13	1	1	2	1	1	1	0	1	1	1
Ma08_g30360	0	0	5	0	0	0	0	1	4	0	0	0	1	0	0	3	0
Ma08_g31720	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0
Ma08_g32760	1	3	11	4	4	1	1	5	5	4	9	3	9	8	16	5	7
Ma08_g34230	0	0	2	1	0	1	0	2	4	1	1	1	2	1	1	2	2
Ma08_g34710	0	0	15	1	0	0	0	16	17	20	8	20	5	5	4	6	7
Ma09_g03310	2	1	11	4	8	3	1	2	5	1	1	2	3	3	5	1	3
Ma09_g03740	8	1	3	4	5	5	4	4	6	3	3	4	2	2	2	2	2
Ma09_g04930	0	2	12	2	0	0	0	10	7	22	7	3	5	5	9	4	3
Ma09_g06730	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
Ma09_g08140	0	0	2	0	1	0	0	3	10	0	0	0	0	0	0	0	0
Ma09_g08260	15	1	9	6	8	4	3	9	9	13	10	4	8	9	8	6	8
Ma09_g09400	0	0	7	1	0	0	1	1	2	2	0	1	1	0	1	1	0
Ma09_g09720	1	0	3	1	0	1	0	2	2	2	2	1	1	1	0	1	1
Ma09_g10800	0	0	5	0	0	0	0	1	1	2	0	2	1	1	0	1	0
Ma09_g11770	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma09_g13170	0	0	5	3	0	0	0	6	7	4	10	4	9	12	4	5	13
Ma09_g14260	0	0	1	0	0	0	0	0	0	0	0	0	1	3	0	2	1
Ma09_g15050	1	0	0	0	0	0	0	1	0	0	1	1	1	1	0	0	0
Ma09_g15130	1	0	1	0	0	0	0	1	0	0	3	0	1	1	0	1	1
Ma09_g15440	0	0	1	2	5	1	0	5	9	3	2	4	1	1	2	2	1
Ma09_g15940	2	0	2	3	0	0	13	0	0	0	0	0	0	0	0	0	0
Ma09_g16940	3	0	0	4	5	8	2	0	1	0	0	0	1	2	1	0	1

Ma09_g16980	0	1	7	0	0	0	0	1	4	1	0	1	1	1	1	1	1
Ma09_g20280	0	0	62	8	0	0	0	44	31	73	45	27	26	30	21	29	23
Ma09_g22730	0	0	2	0	0	0	0	1	1	0	0	1	1	1	1	1	1
Ma09_g23100	32	2	3	36	45	65	32	10	9	21	9	1	5	2	12	4	3
Ma09_g24640	0	0	8	4	3	0	8	7	12	6	3	6	4	5	2	6	4
Ma09_g25010	4	0	2	4	3	4	7	1	1	1	2	0	2	1	2	1	2
Ma09_g25590	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma09_g27990	0	0	1	1	2	0	0	1	2	2	0	1	2	2	1	1	3
Ma09_g28970	1	0	0	7	26	2	0	1	4	0	0	0	0	0	0	0	0
Ma09_g29010	0	2	32	6	0	1	3	25	26	32	15	25	15	16	7	17	22
Ma09_g29660	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
Ma10_g01730	0	0	0	1	4	1	0	0	0	0	0	0	0	0	0	0	0
Ma10_g01750	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Ma10_g04420	0	0	1	4	10	2	2	1	2	1	1	0	1	1	3	0	0
Ma10_g04920	3	0	0	1	3	2	0	0	1	0	0	0	0	0	1	0	0
Ma10_g05260	0	0	0	1	1	1	3	0	0	0	0	0	0	0	0	0	0
Ma10_g05680	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Ma10_g06140	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
Ma10_g09100	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0
Ma10_g09370	0	0	4	3	1	0	5	2	2	3	3	1	3	3	2	4	2
Ma10_g10820	0	0	2	0	0	0	0	4	14	1	0	1	1	2	1	1	1
Ma10_g11100	0	1	6	2	0	0	6	0	0	0	0	0	2	0	0	7	0
Ma10_g13000	3	0	5	10	2	21	16	1	1	1	0	2	1	1	1	1	1
Ma10_g13640	2	1	40	31	30	34	12	22	25	19	20	25	35	48	21	28	44
Ma10_g14150	0	1	1	2	2	3	4	1	3	1	0	1	1	0	1	1	0
Ma10_g14950	0	0	3	2	0	0	0	4	8	1	2	3	4	4	2	3	7
Ma10_g16050	5	1	10	28	74	23	13	9	14	16	4	3	7	4	16	3	4
Ma10_g17650	0	30	0	0	0	0	0	4	13	1	0	0	2	2	2	4	1
Ma10_g18840	3	1	5	2	1	2	2	4	6	3	6	2	3	5	2	2	5
Ma10_g19130	0	0	2	0	0	0	0	1	0	0	1	1	1	0	1	0	1
Ma10_g19820	1	0	2	6	0	0	22	0	0	0	0	0	5	0	12	9	0
Ma10_g19970	1	0	0	2	6	3	1	1	3	1	0	0	0	0	0	0	0
Ma10_g24510	9	4	2	7	16	6	4	3	4	2	4	2	1	1	1	1	2
Ma10_g25660	1	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0
Ma10_g26540	5	7	6	4	3	4	4	6	8	4	4	6	4	5	2	3	6
Ma10_g26660	2	0	6	3	1	2	2	10	14	11	8	5	4	3	3	5	6
Ma10_g29230	6	0	1	4	6	9	3	0	1	1	0	0	1	0	2	0	0
Ma10_g29290	4	1	1	2	2	2	2	1	1	0	1	1	1	1	0	0	1
Ma10_g29660	2	0	1	15	60	0	0	1	3	0	0	0	0	0	0	0	0
Ma10_g29900	5	1	9	10	7	11	11	8	8	7	9	8	9	11	10	10	7
Ma11_g00330	0	0	4	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Ma11_g00350	0	0	6	5	13	1	0	3	5	2	2	2	5	7	5	3	6
Ma11_g02310	0	0	2	1	0	0	0	2	1	1	4	0	2	2	1	1	4
Ma11_g03860	0	0	2	13	47	1	0	5	14	4	1	0	2	1	5	1	1
Ma11_g04680	20	2	0	32	6	57	64	0	1	1	0	0	4	0	7	10	0
Ma11_g06880	22	4	15	13	7	19	12	31	41	39	30	14	7	6	4	9	11
Ma11_g07330	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma11_g07530	0	0	5	2	1	1	1	6	3	14	4	2	3	5	1	4	2
Ma11_g08730	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1
Ma11_g10680	1	2	5	2	1	0	3	12	21	20	6	3	6	7	5	8	6
Ma11_g10710	0	1	2	1	4	0	0	2	5	2	1	2	3	1	6	2	1
Ma11_g11300	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0
Ma11_g11940	0	0	5	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Ma11_g14670	0	1	2	0	0	0	0	0	0	0	0	0	2	0	2	5	0
Ma11_g15740	4	0	2	5	7	4	6	6	9	4	6	5	2	1	3	2	1
Ma11_g16150	1	2	3	10	37	2	1	4	7	4	6	1	3	3	3	2	4
Ma11_g16430	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ma11_g19220	1	0	1	0	0	0	0	0	1	1	0	0	2	0	0	7	0
Ma11_g21160	1	0	9	3	0	1	12	2	4	0	4	0	0	1	0	0	0
Ma11_g21730	0	0	2	1	0	1	0	2	2	3	0	1	2	4	0	2	2
Ma11_g21820	0	0	0	2	2	4	0	0	0	0	0	0	0	0	1	0	0
Ma11_g23010	0	0	6	1	0	0	0	3	3	5	2	2	2	3	1	2	3
Ma11_g23420	1	0	2	0	0	0	0	1	2	0	0	0	0	0	0	0	0

Our expression analyses revealed that *M. acuminata* MYBs have diverse expression patterns in different organs. Many of the *MaMYBs* exhibited low transcript abundance levels with

332 expression in only one or a few organs. This is in good accordance with other transcription
333 factor genes, typically known to be expressed in this manner due to functional specificity and
334 diversity. The highest number of expressed *MaMYB* genes (222; 75.5%) is observed in roots,
335 followed by pulp (210; 71.4%), leaf (204; 69.4%), peel (197; 67%) and embryonic cell
336 suspension (130; 44.2%). The fewest *MaMYB* genes are expressed in seedlings (99; 33.7%) in
337 the considered dataset. 41 *MaMYB* genes (14%) were expressed in all samples analysed
338 (albeit with varying expression levels), which suggested that these MaMYBs play regulatory
339 roles at multiple developmental stages. 14 *MaMYB* genes (4.8%) lacked expression
340 information in any of the analysed samples, possibly indicating that these genes are expressed
341 in other organs (e.g. pseudostem, flower, bract), specific cells, at specific developmental
342 stages, under special conditions or being pseudogenes. 279 MaMYBs (94.8%) are expressed
343 in at least one of the analysed samples, although the transcript abundance of some genes was
344 very low. Some *MaMYB* genes were expressed in all analysed RNA-seq samples at similar
345 levels (e.g. R2R3-MYBs *Ma06_33440* and *Ma11_06880*) while others show variance in
346 transcript abundance with low (no) levels in one or several organs and high levels in others
347 (or vice versa). For example, *Ma02_16570, Ma03_09310, Ma07_11110, Ma10_01730,*
348 *Ma10_05260* and *Ma10_09100* show organ-specific expression, as their transcripts were
349 exclusively detected in leaves, which hints to leaf-specific functionality. *Ma03_07840* and
350 *Ma07_19470* were found to be predominantly expressed in pulp, showing expression also in
351 other analysed organs, but not in the seedling. Overall, this results suggests that the
352 corresponding MaMYB regulators are limited to distinct organs, tissues, cells or conditions.
353 Some clustered *MaMYB* genes (Table 1), which are considered to be created by duplication
354 events within the genome and termed as paralogs, showed different expression profiles, while
355 other clustered *MaMYB* genes did not. For example, the clustered *R2R3-MYB* genes
356 *Ma06_33430* and *Ma06_33440* (both in the axillary meristem/root growth cluster 14) are both
357 expressed in pulps, peels and roots, but *Ma06_33440* is also highly expressed in leaves and
358 embyonic cells, while *Ma06_33430* is not. The expression pattern of *Ma03_07840* and
359 *Ma03_07850* (both in the proanthocyanin-related cluster 25) is, in contrast, very similar, with
360 low expression in root, leaf, pulp and peel, but no expression in embyonic cells and roots.
361 These results could point to functional redundancy of the genes *Ma03_07840* and
362 *Ma03_07850*, while *Ma06_33430* and *Ma06_33440* could be (partly) involved in distinct
363 aspects of growth processes.
364
365

366 **Conclusions**

367 The given genome-wide identification, chromosomal organisation, functional classification
368 and expression analyses of *M. acuminata MYB* genes provide a first step towards cloning and
369 functional dissection to decode the role of *MYB* genes in this economically interesting species.
370
371

372 **Authors' contributions**

373 AP and RS conceived and designed research. BP and RS conducted experiments. BP analyzed
374 and visualized transcriptome data. RS and BW interpreted the data. RS wrote the manuscript.
375 All authors read and approved the final manuscript.
376

377 **Acknowledgements**

378 We are grateful to Melanie Kuhlmann for excellent technical assistance. The research stay of
379 Ashutosh Pandey at the Weisshaar lab was made possible by the Alexander von Humbold
380 Foundation (AvH).
381

382 **References**

383 Aharoni A, De Vos CHR, Wein M, Sun Z, Greco R, Kroon A, Mol JNM, O'Connell AP
384 (2001) The strawberry FaMYB1 transcription factor suppresses anthocyanin and
385 flavonol accumulation in transgenic tobacco. The Plant Journal 28 (3):319-332
386 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search
387 tool. Journal of Molecular Biology 215 (3):403-410
388 Baum D (2008) Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups. Nature
389 Education 1 (1):190
390 Brockington SF, Alvarez-Fernandez R, Landis JB, Alcorn K, Walker RH, Thomas MM,
391 Hileman LC, Glover BJ (2013) Evolutionary analysis of the MIXTA gene family
392 highlights potential targets for the study of cellular differentiation. Molecular Biology
393 and Evolution 30 (3):526-540
394 Castañeda NEN, Alves GSC, Almeida RM, Amorim EP, Fortes Ferreira C, Togawa RC,
395 Costa MMDC, Grynberg P, Santos JRP, Cares JE, Miller RNG (2017) Gene
396 expression analysis in Musa acuminata during compatible interactions with
397 Meloidogyne incognita. Annals of Botany 119 (5):915-930
398 Chaw SM, Chang CC, Chen HL, Li WH (2004) Dating the monocot-dicot divergence and the
399 origin of core eudicots using whole chloroplast genomes. Journal of Molecular
400 Evolution 58 (4):424-441
401 Cooper GM (2000) Regulation of transcription in eukaryotes. In: The cell: A molecular
402 approach. Sinauer Associates, Sunderland, MA,
403 D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc
404 G, Rouard M, Da Silva C, Jabbari K, Cardi C, Poulain J, Souquet M, Labadie K,
405 Jourda C, Lengellé J, Rodier-Goud M, Alberti A, Bernard M, Correa M,
406 Ayyampalayam S, Mckain MR, Leebens-Mack J, Burgess D, Freeling M, Mbéguié-A-
407 Mbéguié D, Chabannes M, Wicker T, Panaud O, Barbosa J, Hribova E, Heslop-
408 Harrison P, Habas R, Rivallan R, Francois P, Poiron C, Kilian A, Burthia D, Jenny C,

409 Bakry F, Brown S, Guignon V, Kema G, Dita M, Waalwijk C, Joseph S, Dievart A,
410 Jaillon O, Leclercq J, Argout X, Lyons E, Almeida A, Jeridi M, Dolezel J, Roux N,
411 Risterucci AM, Weissenbach J, Ruiz M, Glaszmann JC, Quétier F, Yahiaoui N,
412 Wincker P (2012) The banana (Musa acuminata) genome and the evolution of
413 monocotyledonous plants. Nature 488 (7410):213-217
414 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
415 Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29
416 (1):15-21
417 Droc G, Larivière D, Guignon V, Yahiaoui N, This D, Garsmeur O, Dereeper A, Hamelin C,
418 Argout X, Dufayard JF, Lengelle J, Baurens FC, Cenci A, Pitollat B, D'Hont A, Ruiz
419 M, Rouard M, Bocs S (2013) The banana genome hub. Database (Oxford) 2013
420 (bat035)
421 Du H, Feng BR, Yang SS, Huang YB, Tang YX (2012a) The R2R3-MYB transcription factor
422 gene family in maize. PLoS One 7 (6):e37463.
423 Du H, Liang Z, Zhao S, Nan MG, Phan Tran LS, Lu K, Huang YB, Li JN (2015) The
424 Evolutionary History of R2R3-MYB Proteins Across 50 Eukaryotes: New Insights
425 Into Subfamily Classification and Expansion. Scientific Reports 5 (11037)
426 Du H, Yang SS, Liang Z, Feng BR, Liu L, Huang YB, Tang YX (2012b) Genome-wide
427 analysis of the MYB transcription factor superfamily in soybean. BMC Plant Biology
428 12 (106):106
429 Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB
430 transcription factors in Arabidopsis. Trends in Plant Science 15:573-581
431 Fan ZQ, Ba LJ, Shan W, Yun-Yi X, Wang-Jin L, Jian-Fei K, Jian-Ye C (2018) A banana
432 R2R3-MYB transcription factor MaMYB3 is involved in fruit ripening through
433 modulation of starch degradation by repressing starch degradation-related genes and
434 MabHLH6. The Plant Journal 96(1): 1191-1205
435 Frison EA, Sharrock SL (1999) The economic, nutritional and social importance of bananas
436 in the world. In: Picq Cea (ed) Bananas and Food Security. INIBAP, Montpellier,
437 France, pp 21-35
438 Gigolashvili T, Berger B, Mock HP, Muller C, Weisshaar B, Flugge UI (2007) The
439 transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in
440 Arabidopsis thaliana. The Plant Journal 50 (5):886-901
441 Haak M, Vinke S, Keller W, Droste J, Rückert C, Kalinowski J, Pucker B (2018) High
442 Quality de Novo Transcriptome Assembly of Croton tiglium. Frontiers in Molecular
443 Biosciences 5:62
444 Hajiebrahimi A, Owji H, Hemmati S (2017) Genome-wide identification, functional
445 prediction, and evolutionary analysis of the R2R3-MYB superfamily in Brassica
446 napus. Genome 60 (10):797-814
447 Jia L, Clegg MT, Jiang T (2004) Evolutionary dynamics of the DNA-binding domains in
448 putative R2R3-MYB genes identified from rice subspecies indica and japonica
449 genomes. Plant Physiology 134 (2):575-585
450 Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis
451 Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33 (7):1870-1874
452 Langridge WHR (2006) Edible Vaccines. Scientific American Sp 16:46-53
453 Lee MM, Schiefelbein J (1999) WEREWOLF, a MYB-related protein in Arabidopsis, is a
454 position-dependent regulator of epidermal cell patterning. Cell 24 (5):473-483
455 Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for
456 assigning sequence reads to genomic features. Bioinformatics 30 (7):923-930
457 Liu X, Yu W, Zhang X, Wang G, Cao F, Cheng H (2017) Identification and expression
458 analysis under abiotic stress of the R2R3-MYB genes in Ginkgo biloba L. Physical
459 and Molecular Biology of Plants 23 (3):503-516

460 Martin G, Baurens FC, Droc G, Rouard M, Cenci A, Kilian A, Hastie A, Doležel J, Aury JM,
461 Alberti A, Carreel F, D'Hont A (2016) Improvement of the banana "Musa acuminata"
462 reference sequence using NGS data and semi-automated bioinformatics methods.
463 BMC Genetics 17 (243)
464 Matus JT, Aquea F, Arce-Johnson P (2008) Analysis of the grape MYB R2R3 subfamily
465 reveals expanded wine quality-related clades and conserved gene structure
466 organization across Vitis and Arabidopsis genomes. BMC Plant Biology 8 (83)
467 Mehta RS, Bryant D, Rosenberg NA (2016) The probability of monophyly of a sample of
468 gene lineages on a species tree. Proceedings of the National Academy of Sciences U S
469 A 113 (29):8002-8009
470 Negi S, Tak H, Ganapathi TR (2015) Cloning and functional characterization of MusaVND1
471 using transgenic banana plants. Transgenic Research 24 (3):571-585
472 Ogata K, Kanei-Ishii C, Sasaki M, Hatanaka H, Nagadoi A, Enari M, Nakamura H,
473 Nishimura Y, Ishii S, Sarai A (1996) The cavity in the hydrophobic core of Myb
474 DNA-binding domain is reserved for DNA recognition and trans-activation. Nature
475 Structural Biology 3 (2):178-187
476 Oppenheimer DG, Herman PL, Sivakumaran S, Esch J, Marks MD (1991) A *myb* gene
477 required for leaf trichome differentiation in Arabidopsis is expressed in stipules. Cell
478 67:483-493
479 Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O,
480 Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D,
481 Sherman BK, Yu CL (2000) Arabidopsis transcription factors: Genome-wide
482 comparative analysis among eukaryotes. Science 290 (5499):2105-2110
483 Salih H, Gong W, He S, Sun G, Sun J, Du X (2016) Genome-wide characterization and
484 expression analysis of MYB transcription factors in Gossypium hirsutum. BMC
485 Genetics 17 (1):129
486 Sonderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA, Kliebenstein DJ (2007) A
487 systems biology approach identifies a R2R3 MYB gene subfamily with distinct and
488 overlapping functions in regulation of aliphatic glucosinolates. PLoS One 2
489 (12):e1322
490 Song C, Yang Y, Yang T, Ba L, Zhang H, Han Y, Xiao Y, Shan W, Kuang J, Chen J, Lu W
491 (2019) MaMYB4 recruits histone deacetylase MaHDA2 and modulates the expression
492 of omega-3 fatty acid desaturase genes during cold stress response in banana fruit.
493 Plant and Cell Physiology 60 (11):2410-2422
494 Stracke R, Holtgräwe D, Schneider J, Pucker B, Rosleff Sörensen T, Weisshaar B (2014)
495 Genome-wide identification and characterisation of R2R3-MYB genes in sugar beet
496 (Beta vulgaris). BMC Plant Biology 14 (249)
497 Stracke R, Werber M, Weisshaar B (2001) The *R2R3-MYB* gene family in *Arabidopsis*
498 *thaliana*. Current Opinion in Plant Biology 4:447-456
499 Tak H, Negi S, Ganapathi TR (2017) Overexpression of MusaMYB31, a R2R3 type MYB
500 transcription factor gene indicate its role as a negative regulator of lignin biosynthesis
501 in banana. PLoS ONE 12 (2):e0172695
502 Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM (2009) Expansion and
503 diversification of the Populus R2R3-MYB family of transcription factors. Plant
504 Physiology 149 (2):981-993
505 Wu W, Yang YL, He WM, Rouard M, Li WM, Xu M, Roux N, Ge XJ (2016) Whole genome
506 sequencing of a banana wild relative Musa itinerans provides insights into lineage-
507 specific diversification of the Musa genus. Scientific Reports 6 (31586):srep31586
508 Yang QS, Gao J, He WD, Dou TX, Ding LJ, Wu JH, Li CY, Peng XX, Zhang S, Yi GJ
509 (2015) Comparative transcriptomics analysis reveals difference of key gene
510 expression between banana and plantain in response to cold stress. BMC Genomics

16:446

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing-Wang D, Zhangliang C, Hongya G, Li-Jia Q (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Molecular Biology 60 (1):107-124

Zhong Y, Yin H, Sargent DJ, Malnoy M, Cheng ZM (2015) Species-specific duplications driving the recent expansion of NBS-LRR genes in five Rosaceae species. BMC Genomics 16:77

Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF (2004) Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. The Plant Journal 40:22-34