Deciphering the regulatory network controlling flavonoid biosynthesis by MYB-bHLH-WDR complexes in Arabidopsis seed

Wenjia Xu

UNIVERSITE PARIS-SUD

ÉCOLE DOCTORALE : SCIENCES DU VÉGÉTAL
Institut Jean-Pierre Bourgin (IJPB), INRA Versailles

DISCIPLINE: Sciences de la Vie

THÈSE DE DOCTORAT

Soutenance prévue le 15/09/2014

par

Wenjia XU

Deciphering the regulatory network controlling flavonoid biosynthesis by MYB-bHLH-WDR complexes in Arabidopsis seed

Composition du jury :

Directeur de thèse :	Loïc LEPINIEC	DR1 INRA (IJPB, Versailles, SPS)
Co-directeur de thèse :	Christian DUBOS	CR1 INRA (BPMP, Montpellier)
Rapporteurs :	Jean-François BRIAT	DREX CNRS (BPMP, Montpellier)
	Jacqueline GRIMA-PETTENATI	DR1 CNRS (LSV, Toulouse)
Examinateurs :	Dao-Xiu ZHOU	PR1 Paris XI (IBP, Orsay)
	Nathalie GUIVARC'H	PR (BBV, Tours)

Acknowledgements

I would like to express my gratitude to various individuals. It could never have been completed without the support of a lot of people.

First and foremost, I would like to thank my supervisors, Dr Loïc Lepiniec and Dr Dubos Christian, who give me the precious opportunity to pursue science and also have faith in my progress with great patience, support and advice. I have learned a lot from you both and realize how kind and generous you've been teachers and supervisors.

I want to thank all my present and past members in my laboratory that provides warm and comfortable surroundings for study and research. They are always helping me when I bother them for the questions of translation or anything else. Special thanks to Damaris Grain, Sophie Bobet and Zsolt Károly Kelemen who helped me a lot in the same flavonoid-team. I am very grateful to Martine Miquel and Johanne Thévenin for their care and encouragement when I was down in spirits.

I would like to thank Dr Nancy Terrier, Dr Kian Hématy and Dr Jean-Marc Routaboul who gave me an exceptional doctoral committee. I am also very grateful for Dr Richard Sibout and Dr Kian Hematy's coming to my presentation of beginning-thesis and mid-thesis for their time and useful suggestions and comments.

I am greatly thankful to the China government and China Scholarship Cousil, for investing in me to pursue my academic dream and change my life. I wish to thank Institut Jean-Pierre Bourgin, INRA Versailles and University of Paris Sud for hosting me. Special thanks to Dr. Jacqui Shykoff for her generous help of administrative stuff.

Finally, this thesis is dedicated to my family. Many full-heart thanks to my parents, without their boundless love, affection and selfless sacrifice, I would not have been what I am today. Special thanks to my mother-in-low, my uncle and my aunt who have always believed in me and given me unconditional loves. From the bottom of my heart I thank my husband for his discussions, suggestions and accompany to overcome any obstacle I have met throughout my study and life.

Table of contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A, C, G, T	adenine, cytosine, guanine, thymine
Amp	ampicillin
ANR	anthocyanidin reductase (BANYULS)
ATP	adenosine triphosphate
BAN	*BANYULS*
BET	ethidium Bromide
bHLH	basic helix-loop-helix, a protein structural motif
bp	base pair
bZIP	basic leucine zipper domain
CaMV	Cauliflower Mosaic Virus
cDNA	complementary DNA
CDS	coding sequence
CHI	chalcone isomerase
ChIP	chromatin immunoprecipitation
CHS	chalcone synthase
Col-0	Columbia-0 Arabidopsis accession
Ct	cycle threshold
DFR	dihydroflavonol 4-reductase
dH2O	distilled water
Dicot	dicotyledon
DMF	di-methyl-formamide
DNA	deoxyribonucleic acid
dNTP	deoxy-ribonucleotide
DTT	dithiothreitol
EBG	early biosynthetic genes
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetate
EGL3	ENHANCER OF GLABRA3
ER	Endoplasmic reticulum
et al.	**‘**And others’
F3’5’H	flavonoid 3’, 5’-hydroxylase
F3’H	flavonoid 3’-hydroxylase
F3H	flavanone 3-hydroxylase
FLS	flavonol synthase
g, mg, ug, ng	gram, milligram, microgram, nanogram
g.	gravitational force
GFP	green fluorescent protein
GL3	GLABRA3
GUS	*β*-glucuronidase
h, min, s	hour, minute, second
Hyg	hygromycin
HCl	hydrochloric acid

Kan	kanamycin
Kb	kilo base pair
LAR	leucoanthocyanidin reductase
LBG	late biosynthetic gene
l, ml, ul	litre, millilitre, microlitre
LB	lysogeny broth
Ler	Landsberg erecta Arabidopsis accession
M, mM, uM	molar, millimolar, micromolar
MBW	MYB-bHLH-WDR protein complex
mm, nm, um	millimetre, nanometer, micrometer
MADS	an acronym of MCMI, AGAMOUS, DEFICIENS and SRF
Monocot	monocotyledon
mRNA	messenger ribonucleic acid
MS	Murashige and Skoog medium
MYB	myeloblastosis
PAP1	production of anthocyanin pigment 1
PAs	proanthocyanidins
PCR	polymerase chain reaction
p35S	35S RNA promoter of cauliflower mosaic virus
pXXXX	XXXX promoter
qPCR	quantitative polymerase chain reaction
qRT-PCR	quantitative reverse transcriptase polymerase chain reaction
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RT-PCR	reverse transcriptase polymerase chain reaction
SDS	sodium dodecyl sulfate
sec	second(s)
TAE	tris, acetic acid, EDTA
Taq	*Thermus aquaticus*
T-DNA	transfer DNA
TE	tris and EDTA
TFs	transcription factors
Tm	melting temperature
TRY	TRIPTYCHON
TT2, 8, 16	TRANSPARENT TESTA2, 8, 16
TTG1, 2	TRANSPARENT TESTA GLABRA1, 2
U	units
UTR	untranslated region
UV	ultraviolet
V	voltage
v/v	volume per volume
w/v	weight per volume
WER	WEREWOLF

WDR	WD40 repeats
WT	wild type
X-gluc	C14H13BrClNO7, as a reagent to detect *β*-glucuronidase

CHAPTER I

GENERAL INTRODUCTION

1 General introduction

1.1 Flavonoids

1.1.1 Flavonoid structures and functions

Flavonoids are one of the largest groups of plant secondary metabolites that derived from the phenylpropanoid pathway. They are widely distributed in plants, ranging from spermatophytes, especially in Angiosperms, to mosses (Popper & Fry, 2003; Tohge *et al.*, 2013). Over 10.000 flavonoid structures sharing 15C carbon skeleton (C6-C3-C6) have been reported and grouped into different classes depending on the modifications of side-chains (Winkel, 2004; Lepiniec *et al.*, 2006; Hernández *et al.*, 2009; Buer *et al.*, 2010; Hichri *et al.*, 2011a; Cheynier *et al.*, 2013; Tohge *et al.*, 2013). Although the central pathway is conserved, depending on the species, various families of enzymes (e.g. isomerases, reductases, hydroxylases, dioxygenases as well as isoflavone synthases and aureusidin synthases) can modify the core flavonoid skeleton, leading to the wide diversity of known flavonoids (e.g. isoflavones in legumes; phlobaphenes in maize; aurones in snapdragon) (Lepiniec *et al.*, 2006; Martens *et al.*, 2010; Falcone Ferreyra *et al.*, 2012). The main classes include flavonols, anthocyanins and proanthocyanidins (PAs) that are well known for the colour they confer to various plant organs including leafs, flowers, fruits or seeds (Mol *et al.*, 1998; Winkel-Shirley, 2001a; Koes *et al.*, 2005; Grotewold, 2006; Lepiniec *et al.*, 2006; Quattrocchio *et al.*, 2006; Allan *et al.*, 2008). Although proanthocyanidins are colourless polymers of flavan-3-ols, they turn brown after oxidation and confer the typical brown colours to the Arabidopsis mature seed (Pourcel *et al.*, 2007). In *Arabidopsis thaliana*, more than fifty flavonoid compounds belonging to flavonols, anthocyanins, and proanthocyanins (PA) have been characterized (Routaboul *et al.*, 2012; Saito *et al.*, 2013).

Their structural diversity and antioxidant properties may contribute to physiological roles in the interactions of plant with the environment. For instance, it has been shown that flavonoids are involved in the protection against predator attacks, pathogen infections, UV-B radiation, nutrition imbalance, temperature variations, and oxidative stresses (Shirley, 1996; Debeaujon *et al.*, 2000; Winkel-Shirley, 2002; Lepiniec *et al.*, 2006; Hernández *et al.*, 2009;

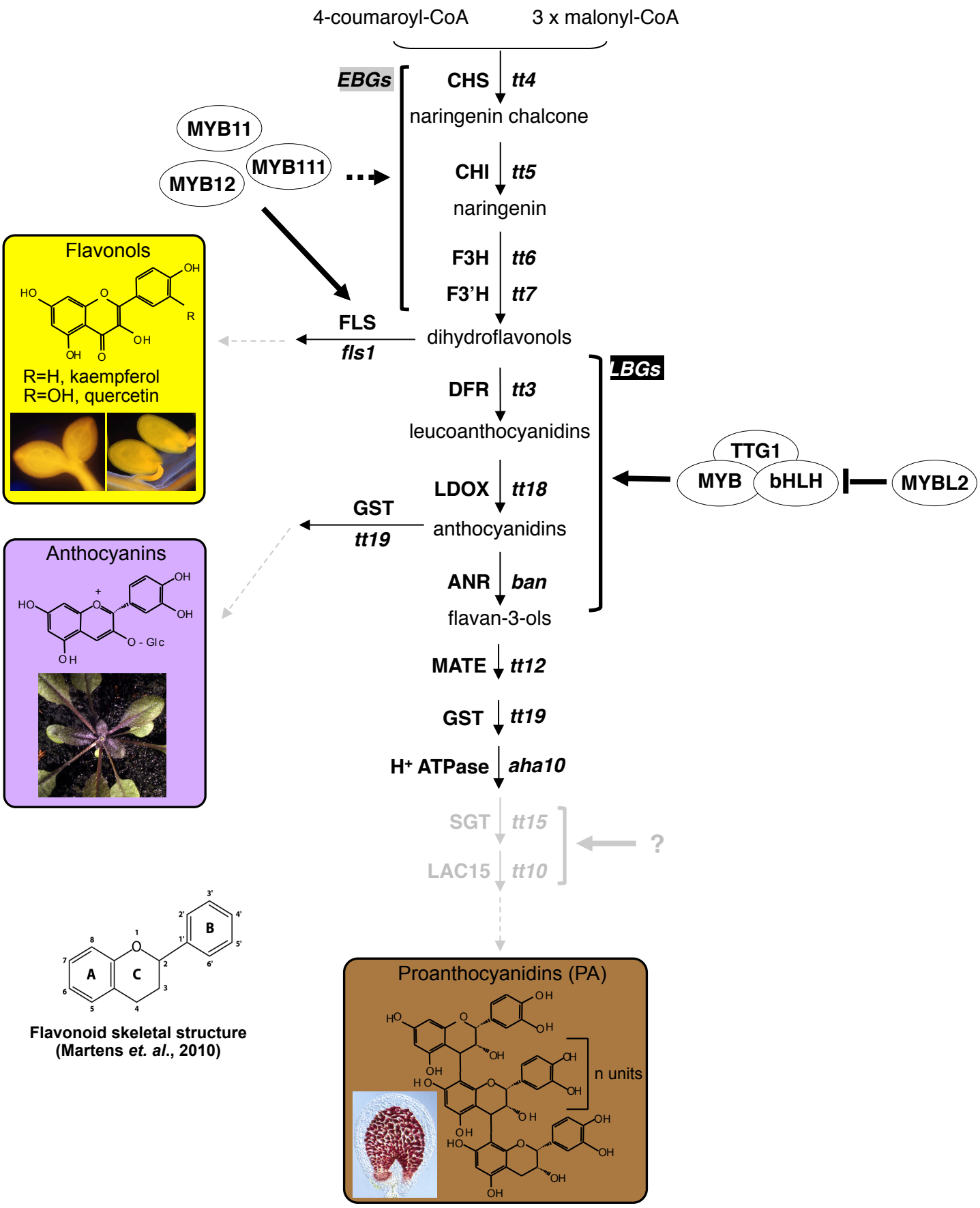


Figure 1-1. Schematic representation of the flavonoid biosynthetic pathway in *Arabidopsis thaliana* (Adapted from Xu *et al*., 2014a and Baudry 2004 PhD thesis). The names of the structural proteins are indicated in capital letters and the corresponding mutants in lower-case italics. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavonol 3-hydroxylase; F3'H, flavonol 3'-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol-4-reductase; LDOX, leucoanthocyanidin dioxygenase; ANR, anthocyanidin reductase; MATE, multidrug and toxic efflux transporter; GST, glutathione-S-transferase; SGT, UDP-glucose:sterol-glucosyltransferase; LAC15, laccase 15. Arrows indicate the different steps leading to the formation of flavonoids in Arabidopsis; dashed lines indicate multiple steps; circles correspond to transcription factors; *EBG*, early biosynthetic gene; *LBG*, late biosynthetic gene. Arrows indicate the different steps leading to the formation and accumulation of flavonoids in Arabidopsis; dashed lines indicate multiple steps; circles correspond to transcription factors.

Olsen *et al.*, 2009; Rowan *et al.*, 2009; Lovdal *et al.*, 2010; Hichri *et al.*, 2011a; Falcone Ferreyra *et al.*, 2012; Fei *et al.*, 2013; Nakabayashi *et al.*, 2014). Moreover, there are growing genetic evidences of cross-talks between flavonoids and various hormones to modulate auxin transport and developmental processes (Peer & Murphy, 2007; Lewis *et al.*, 2011; Buer *et al.*, 2013), including seed size and fertility in some species such as petunia or maize (Lev-Yadun *et al.*, 2002; Falcone Ferreyra *et al.*, 2012; Doughty *et al.*, 2014). Interestingly, in addition to their roles in plant development and physiology, flavonoids also account for agronomic, industrial, nutritional and health values of plant products. For instance, they impact seed quality (e.g. viability and oil extractability), the astringency of various plant products (e.g wine and beer) and health beneficial properties of food (e.g. prevention against cardiovascular and neurodegenerative diseases) (Downey *et al.*, 2006; Butelli *et al.*, 2008; Luceri *et al.*, 2008; Buer *et al.*, 2010). Last, because of their diversity, functions, and ease to use, flavonoids have provided useful models for the study of a wide variety of biological (genetic, epigenetic, cellular, biochemical, and evolutionary) processes and are probably ones of the best-studied metabolic and regulatory pathways in nature (Dixon & Steele, 1999; Clegg & Durbin, 2000; Winkel-Shirley, 2001a; Koes *et al.*, 2005; Lepiniec *et al.*, 2006; Feller *et al.*, 2011; Sobel & Streisfeld, 2013).

1.1.2 Structure and regulation of the biosynthetic pathway

Flavonols, anthocyanins and PA share common precursors (i.e. dihydroflavonols) derived from phenylalanine and malonyl-CoA (Cheynier *et al.*, 2013), through a series of reactions catalyzed by chalcone synthase (CHS), chalcone isomerase (CHI), flavonol 3-hydroxylase (F3H) and flavonol 3'-hydroxylase (F3'H) (Fig. 1-1). The genes encoding these enzymes that are often co-regulated and the first induced in the pathway in response to environmental and developmental cues, have been named *Early Biosynthetic Genes* (*EBGs*) (Fig. 1-1, 2, 3) (Winkel-Shirley, 2001a; Lepiniec *et al.*, 2006; Quattrocchio *et al.*, 2006; Yonekura-Sakakibara *et al.*, 2007; Stracke *et al.*, 2010; Tohge & Fernie, 2012; Lai *et al.*, 2013; Tohge *et al.*, 2013). By extension the other genes of the pathway (as well as *F3H* and *F3'H* in some species) are usually named *Late Biosynthetic Genes* (*LBGs*). Most of the

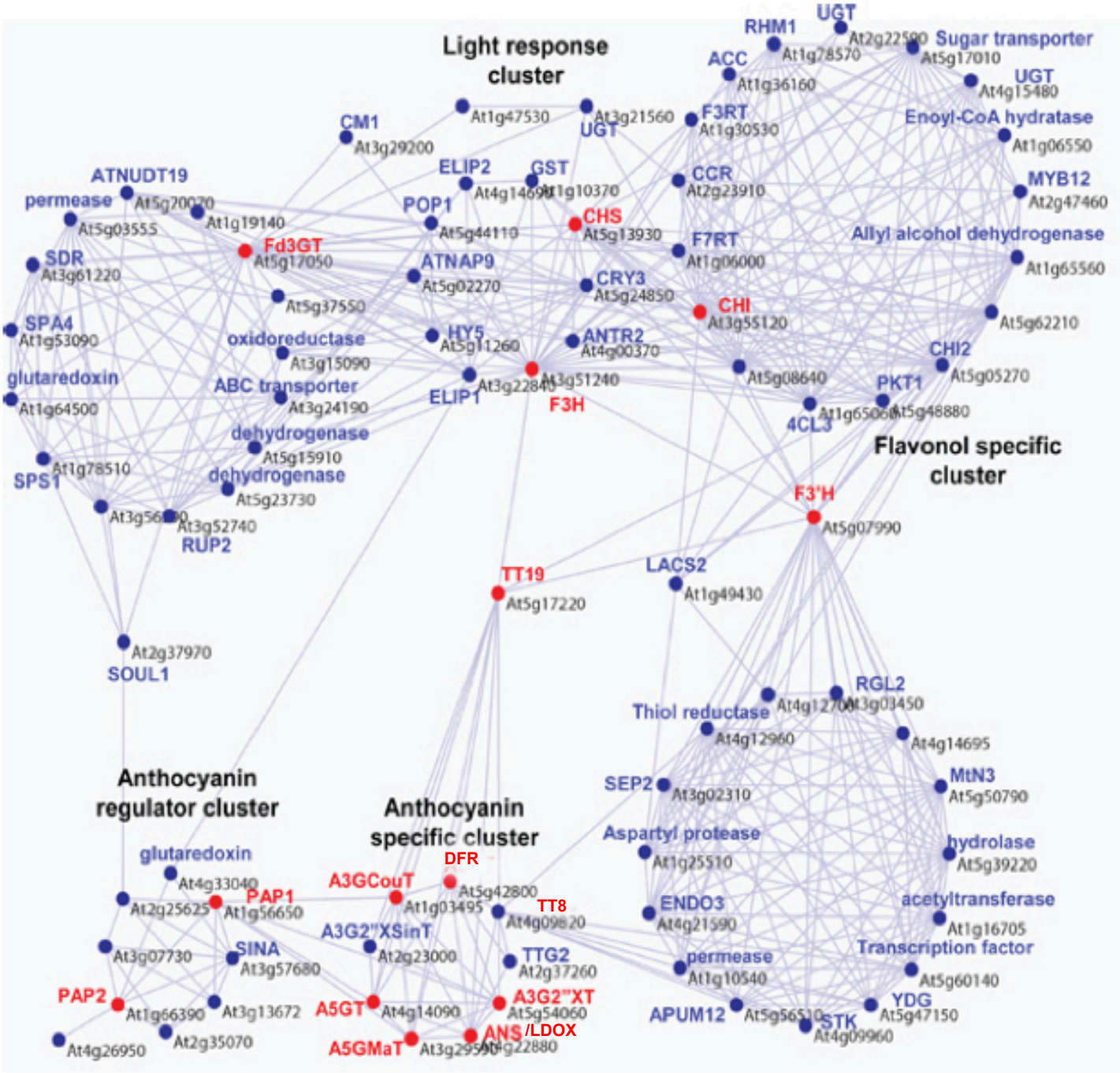


Figure 1-2. Example of co-regulation network analysis of the anthocyanin pathway (Tohge and Fernie, 2012). Coexpression analyses were performed using the PRIMe (http://prime.psc.riken.jp/?action=coexpression_index) based on the data set of ATTEDII version 3 8,2 with the Pajek program (http://vlado.fmf.uni-lj.si/pub/networks/pajek/). Positive correlations (r<0.5) are used to make network connections. Red node: thirteen anthocyanin genes (AT4G09820, **TT8**, regulator; At5g13930, **CHS**, chalcone synthase; At3g55120, **CHI**, chalcone isomerise; At3g51240, **F3H**, flavanone 3-hydroxylase; At5g07990, **F3'H**, flavonoid 3'-hydroxylase; At5g17050, Fd3GT, UGT78D2, flavonoid 3-O-glucosyltransferase; At5g17220, AtGSTF12, **TT19**; At5g42800, **DFR**, dihydroflavonol reductase; At4g22880, **ANS/LDOX**, anthocyanidin synthese; At4g14090, A5GT, anthocyanin 5-O-glucosyltransferase; At5g54060, A3G2"XT, putative anthocyanin 3-O-glucoside 2"-O-xylosyltransferase; At3g29590, A5GMaT, anthocyanin 5-O-glucoside 6"'-O-malonyltransferase; At1g03940, A3GCouT, anthocyanin 3-O-glucoside 6"-O-p-coumaroyltransferase) and two transcription factors for anthocyanin production (At1g56650, **PAP1**; At1g66390, **PAP2**) was used for searching candidate genes. Candidate genes were found by an "intersection of sets" search with a threshold value with a coefficient of r>0.50 queried by intersection of sets by all genes queried (Fourteen anthocyanin biosynthetic genes). A co-expression network, including correlated candidate genes (68 genes) and queried genes (14 genes), was re-constructed by an "interconnection of sets" search with r>0.50 using the PRIMe database. The output files that were formatted with a '.net' file from the PRIMe database and networks were drawn using Pajek software. Blue node indicates candidate genes which correlated with anthocyanin genes.

corresponding loci have been characterized in *Arabidopsis thaliana* on the basis of easily detectable mutations that impaired seed coat pigmentation, the so-called "*transparent testa*" mutations (Fig. 1-4b-d, Table 1-1) (Koornneef, 1990; Lepiniec *et al.*, 2006). Among the proteins encoded by the *LBGs* are TT3 (a dihydroflavonol-4-reductase, DFR), TT18 (a leucoanthocyanidin dioxygenase, LDOX), TT19 (a glutathione-S-transferase, GST), TT12 (a MATE transporter) and AHA10 (a H^+-ATPase) involved in both anthocyanin and PA biosynthesis, whereas BANYULS (an anthocyanidin reductase, BAN/ANR) is specifically dedicated to the biosynthesis of PAs (Fig. 1-8f-i) (Lepiniec *et al.*, 2006). A third group contains structural genes for which the expression pattern appears different from the *EBGs* or *LBGs*, namely *TT15* (encoding a UDP-glucose:sterol-glucosyltransferase, *UGT80B1*) and *TT10* (*laccase-like 15*) (Fig. 1-3) (Pourcel *et al.*, 2005; DeBolt *et al.*, 2009; Routaboul *et al.*, 2012).

In Arabidopsis, flavonols are found in all tissues whereas anthocyanin and PAs specifically accumulate in vegetative tissues and seed coat, respectively (Fig. 1-1). PA-accumulating cells are localized in the innermost cell layer of the integuments (Debeaujon *et al.*, 2003; Lepiniec *et al.*, 2006). A developmental feature of seed development is the formation of five-layer seed coat (Fig. 1-4a) (Schneitz *et al.*, 1995; Debeaujon *et al.*, 2003; Sieber *et al.*, 2004; Haughn & Chaudhury, 2005). Firstly**,** outer (oi1 and oi2) and inner (ii1' and ii2) integuments with four layers initiate from the flanks of the chalaza and develop into the primal seed coat. A few cells belonging to the ii2 layer constitute the micropylar area and remaining cells lying between the vascular bundle and the nucellus make up chalaza. Then ii1' cell layer performs periclinal divisions that give rise to the ii1 cell layer, namely endothelium, on their outer side during ovule and seed development (Debeaujon *et al.*, 2003). PA accumulation begins in the micropylar region of the endothelium, approximately at the two-cell stage of embryo development, progresses in the seed body, and ends in the chalazal bulb at the heart stage of embryo development (Fig. 1-4e-i). Nevertheless**,** *BAN* promoter activity is detected first in the micropyle area, before fertilization, then in the chalaza and endothelium. Seed coat growth and differentiation proceed coordinately with embryo and endosperm development following fertilization (Haughn & Chaudhury, 2005). This mechanism may involve polycomb group proteins whose activities block the formation of the

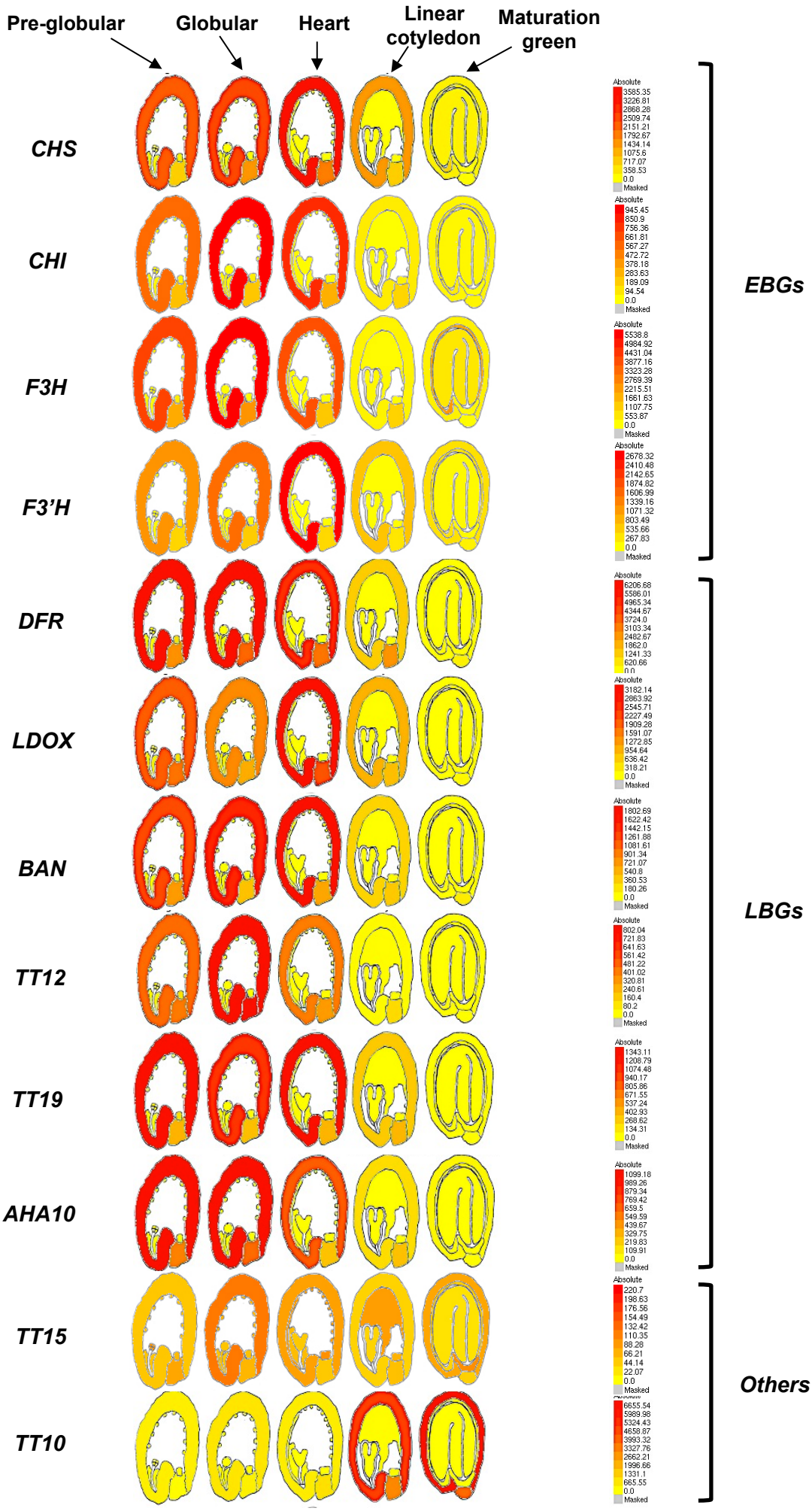


Figure 1-3. Steady state mRNA level of the structural genes involved in PA biosynthesis in developing *Arabidopsis thaliana* seeds. Data were generated by combining laser-capture micro dissection coupled with ATH1 GeneChip analysis (Le *et al*., 2010). Schematic representations are available at the following URL: http://bar.utoronto.ca/efp/cgi-bin/efpWeb.

seed coat and PA synthesis until fertilization (Roszak & Kohler, 2011).

In vegetative tissues, the flavonoid pathway is usually induced in response to physiological and environmental fluctuations, allowing to cope with the oxidative stresses induced by pathogen infections, high-light, UV, temperature, drought, salt, and such as nitrogen, phosphorous or carbon assimilation (Dixon & Paiva, 1995; Teng *et al.*, 2005; Rolland *et al.*, 2006; Solfanelli *et al.*, 2006; Datta *et al.*, 2007; Catala *et al.*, 2011; Fei *et al.*, 2013; Nakabayashi *et al.*, 2014). Various hormones have been shown to be involved in these regulations including abscisic acid (ABA), gibberellic acid (GA), brassinosteroid (BR), ethylene (ET), jasmonate (JA) or cytokinin (CTK) (Loreti *et al.*, 2008; Peng *et al.*, 2011; Das *et al.*, 2012; Ye *et al.*, 2012; Jeong *et al.*, 2010). For instance, GA and ET have negative effect on the induction of flavonoid gene expression by sucrose, whereas JA, ABA and CTK have a synergic effect with sucrose.

These environmental or developmental regulations mostly rely on the coordinated expression of structural genes (Fig. 1-2, 3). Some of the recent discoveries allowed starting to decipher the transcriptional regulatory networks involved, by identifying and characterizing key regulatory proteins. Indeed, besides structural genes, *transparent testa* mutations allowed the characterization of six regulatory genes that encode TT2 (MYB123), TT8 (bHLH042), TTG1 (WDR family), TT16 (ABS/AGL32, MADS box), TT1 (WIP1/Zn finger), and TTG2 (DSL1/WRKY44) proteins (Table 1-1) (Lepiniec *et al.*, 2006). TT1, TT16 and TTG2 are also involved in epidermal cell fates, suggesting that they may control PA accumulation indirectly (Johnson *et al.*, 2002; Nesi *et al.*, 2002; Sagasser *et al.*, 2002; Garcia *et al.*, 2005; Ishida *et al.*, 2007). The regulation of flavonol biosynthesis involves different R2R3-MYB, namely MYB11, 12, and 111 (also known as PFG2, 1 and 3, respectively) targeting *EBGs* and *FLS1*, the first gene specifically committed to the biosynthesis of flavonols (Mehrtens *et al.*, 2005; Stracke *et al.*, 2007; Lai *et al.*, 2013). Here, we will focus on the important progresses made during the last decade on the characterization of the molecular functions and regulations of TT2, TT8, and TTG1, as well as on their homologous MYB, bHLH, and WDR proteins involved in other pathways in Arabidopsis and/or in other plant species, as described below.

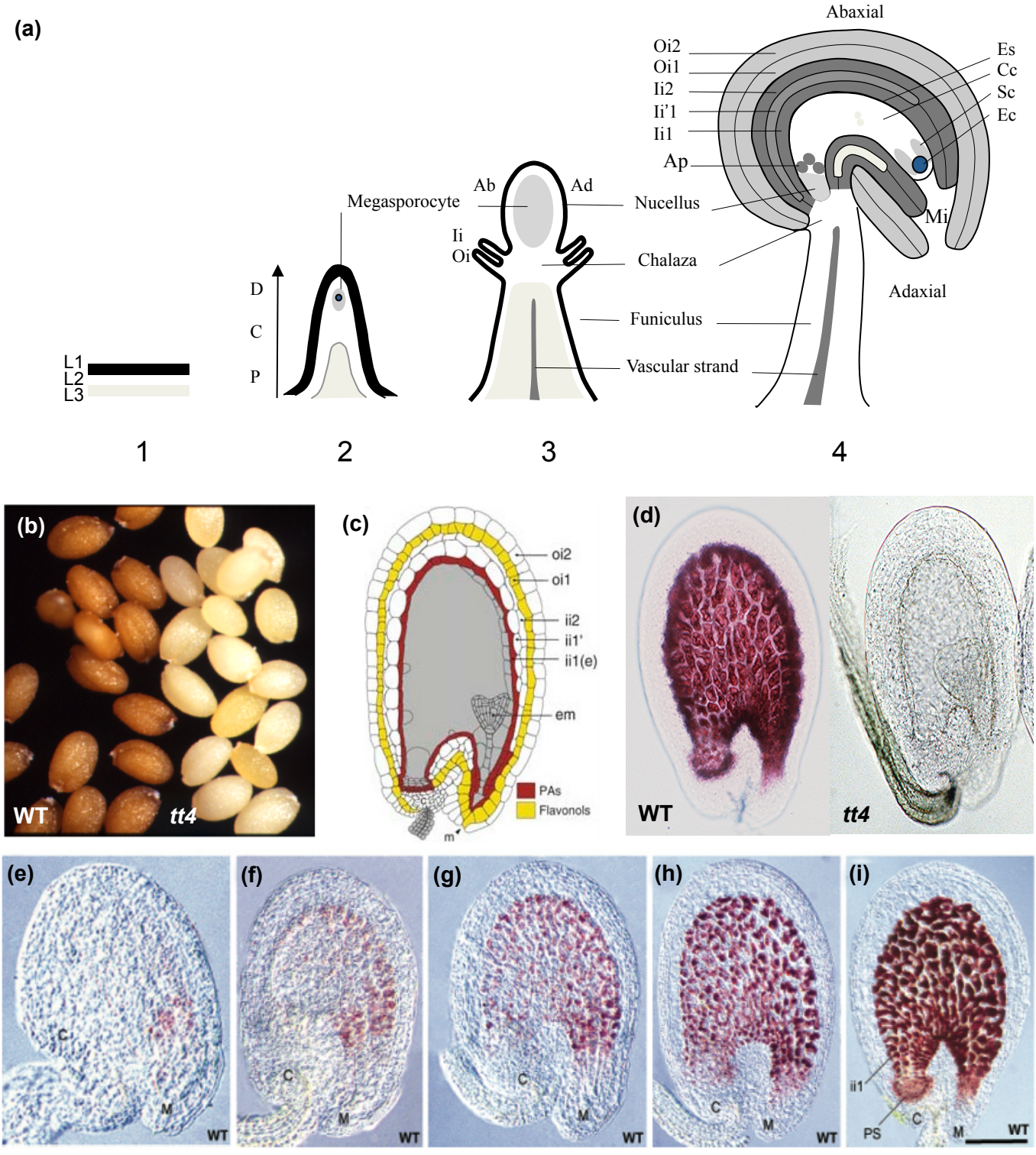


Figure 1-4. Development of the ovule and origin of teguments and PA localization in developing *Arabidopsis thaliana* seeds (Debeaujon *et al*., 2003; Pourcel *et al*., 2005; Lepiniec *et al*., 2005; Lepiniec *et al*., 2006). (a) The ovule is formed from the placenta by periclinal divisions of the epidermal layer L1 and sub-epidermal layers L2 and L3 (1, 2). The primordium of the ovum develops along a proximal-distal axis (PCD). At the distal extremity, the nucellus gives rise to megasporocyte the origin of the embryo sac (Es). From the chalaza, form internal teguments (Ii) and external teguments (Oi) that have epidermal origin (3). The second axis of development of the egg is an adaxial-abaxial axis (Ab-Ad) perpendicular to the proximal-distal axis. At maturity (4), the embryo sac consists of seven cells, oosphere (Ec), two synergids (Sc) , the central cell (Cs) and three poles (Ap). The asymmetric development of teguments with a strong extension pole abaxial and to a lesser extent the adaxial pole causes the approximation of the micropyle (Mi) and funiculus which become adjacent. (b) Arabidopsis seed phenotypes in wild-type and transparent testa mutant. (c) Scheme of Arabidopsis seed anatomy. c, chalaza; e, endothelium; em, embryo; ii, inner integument; m, micropyle; oi, outer integument. (d) Histochemical detection of seed flavonoids at the late globular-heart stage of embryo development stained as whole mounts, as no staining is detected in the tt4 mutant seeds deprived of flavonoids. The vanillin assay is used to detect flavan-3-ols and their proanthocyanidin polymers. (e-i) Accumulation of PAs during seed development (whole-mount vanillin staining). Embryo stages are two cells (e), early globular (f), globular (g), early heart (h) and heart (i).

Locus	Seed coat color[a]	Gene product	Branch[b]
Structural genes			
tt3	Yellow	Dihydroflavonol reductase (DFR)	P, A
tt4	Yellow	Chalcone synthase (CHS)	P, F, A
tt5	Yellow	Chalcone isomerase (CHI)	P, F, A
tt6	Pale brown spotted	Flavanone-3-hydroxylase (F3H)	P, F, A
tt7	Pale brown spotted	Flavanone-3′-hydroxylase (F3′H)	P, F, A
tt10	Dark yellow/brown C[c]	Polyphenol oxydase (PPO)	P, F
tt12	Pale brown	MATE secondary transporter	P
tt15	Pale brown/brown CM	Glycosyltransferase (GT)	P
tt18/ tds4/ tt11	Yellow	Leucocyanidin dioxygenase (LDOX)[d]	P, A
tt19/tt14	Dark yellow[3]	Glutathione S-transferase (GST)	P, A
ban	Pale gray/gray CM	Anthocyanidin reductase (ANR)	P
aba10	Pale brown	Autoinhibited H^+-ATPase isoform 10	P
Regulatory genes			
tt1	Yellow/brown CM	Transcription factor WIP-type Zn-Finger	P
tt2	Yellow	Transcription factor AtMYB123	P
tt8	Yellow	Transcription factor AtbHLH042	P, A
tt16 /abs	Yellow/brown CM	Transcription factor MADS AtAGL32	P
ttg1	Yellow	Regulatory protein ("WD40" or "WDR")	P, A
ttg2	Yellow	Transcription factor AtWRKY44	P
Other loci			
tt9	Pale gray/dark CM	Unknown	?
tt13	Pale brown	Unknown	?
tt17	Pale brown	Unknown	?
tds1,3, 5, 6	Pale brown	Unknown	P
tds2	Pale brown	Unknown	P, A

Table 1-1. *Loci* involved in PA biosynthesis in *Arabidopsis thaliana* seed (Lepiniec *et al.*, 2006). [a]Wild-type seed coat color is brown. [b]Affected metabolic branch: P, proanthocyanidins (only in seed coat); F, flavonols; A, anthocyanins (only in vegetative parts). [c]Seeds brownish with storage time. [d]Also called anthocyanidin synthase (ANS). C, Chalaze; M, Micropyle

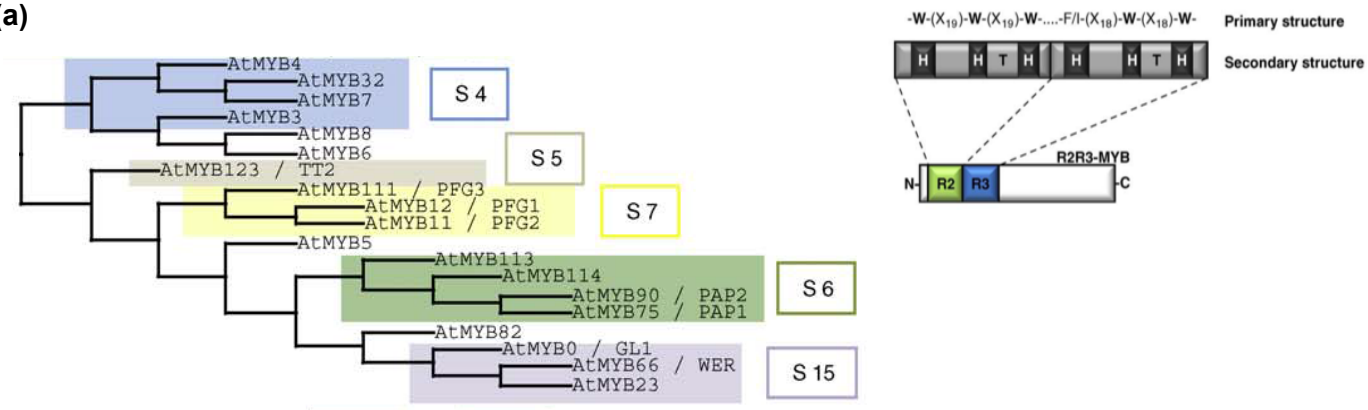
1.2 Regulating flavonoid genes expression at the transcriptional level by MBW complexes

1.2.1 MBW structures and functions

MYB and bHLH transcription factors (TFs) are found in all eukaryotes and are among the largest families of plant TFs with more than hundred genes encoding these proteins in Arabidopsis (Riechmann *et al.*, 2000; Heim *et al.*, 2003; Dubos *et al.*, 2010). MYB proteins are characterized by a highly conserved DNA-binding domain (DBD) that consists of one up to four imperfect amino acid repeats (R) of about 52 amino acids (Dubos *et al.*, 2010). They fall into different classes depending on the number of adjacent repeat(s), the largest class encoding R2R3-MYB, with 126 genes in Arabidopsis. Based on the conservation of the DBD and variability of the C terminal domains, R2R3-MYB proteins have been divided into at least 25 subgroups with various biological functions (Stracke *et al.*, 2001; Dubos *et al.*, 2010). The bHLH proteins, contain a conserved domain which consists of about 60 amino acids including a HLH region involves in dimerization and a 15-amino-acid basic region, located at the N-terminal end of the domain, involved in DNA binding (Heim *et al.*, 2003; Toledo-Ortiz *et al.*, 2003). More than 130 bHLH genes are found in Arabidopsis and have been divided into twelve subgroups. WDR (WD40-Repeats) proteins are found in all eukaryotes and form a large family of more than 200 members in plants (van Nocker & Ludwig, 2003; Li, Q *et al.*, 2014). The WD motif is about 40 amino acids and is usually involved in protein-protein interactions. Among the Arabidopsis WDR proteins, TTG1 has been identified for the pleiotropic phenotypes of the corresponding mutants affected in flavonoid biosynthesis and various epidermal cell fates (Walker *et al.*, 1999).

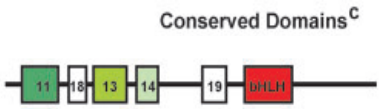
The interactions between the R2R3-MYB from subgroups 5 (TT2), 6 (PAP1/MYB75, PAP2/MYB90, MYB113 and MYB114) and 15 (WER/MYB66, GL1/MYB0, and MYB23) or AtMYB5 and AtMYB82, and the R/B-like bHLH (subgroup IIIf: TT8, GL3, EGL3, and AtMYC1) are probably among the best-described cooperation of TFs in plants (Fig. 1-5a,b) (Payne *et al.*, 2000; Esch *et al.*, 2003; Heim *et al.*, 2003; Zhang *et al.*, 2003; Baudry *et al.*, 2004; Hernandez *et al.*, 2004; Zimmermann *et al.*, 2004; Koes *et al.*, 2005; Quattrocchio *et al.*, 2006; Gonzalez *et al.*, 2008; Feller *et al.*, 2011; Zhao *et al.*, 2012). These interactions

(a)



(b)

Group	Generic name[a]	Synonym[b]	AGI genecode	Accession	Exon Number	5-9-13 Motif
IIIf	AtbHLH042	TT8	At4g09820	AJ277509	7	H-E-R
	AtbHLH012	ATMYC1*	At4g00480	AF251697	8	L-D-R
	AtbHLH001 *	Gl3	At5g41315	AF246291	7	H-E-R
	AtbHLH002 *	EGL1*	At1g63650	AF027732	7	H-E-R



(c)

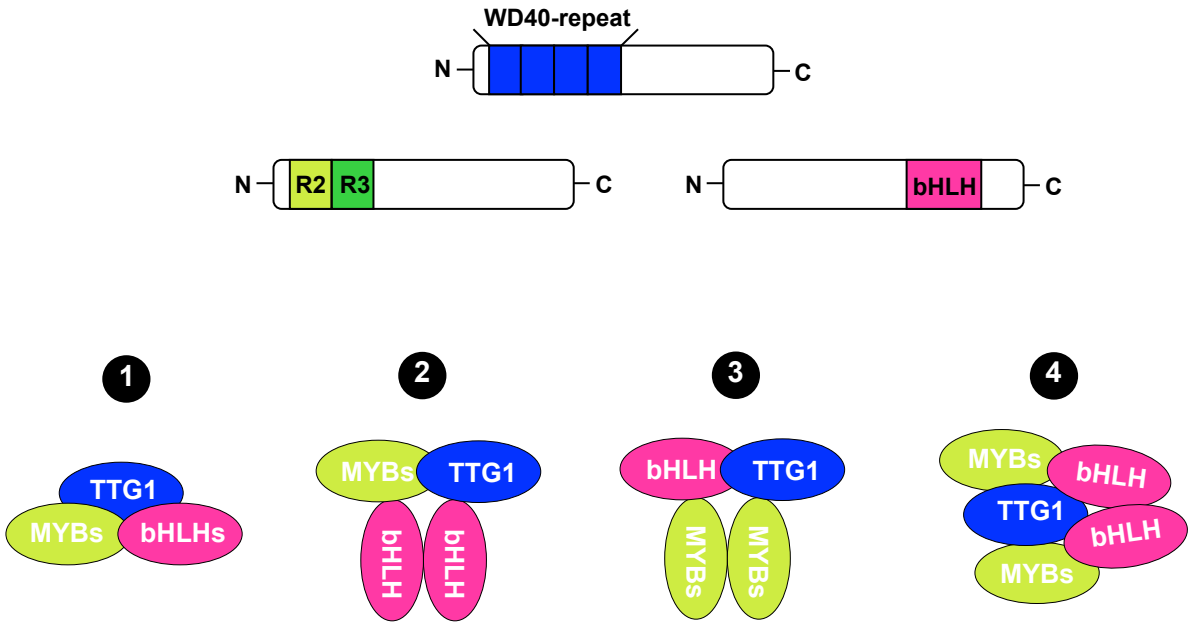


Figure 1-5. MYB-bHLH-WDR protein complexes in *Arabidopsis thaliana*. (a) Phylogenetic S5, S6 and S7 subgroups of *A. thaliana* R2R3-MYBs, The tree was inferred using the neighbor-joining method and 1000 bootstraps with putative amino acid full length MYB sequences with Clustal X2 software (Dubos *et al*., 2010). (b) Phylogenetic IIIf subgroup of *A. thaliana* bHLH gene family, of which subdivision of the groups is based on structural similarities (Heim *et al*., 2003). (c) Putative stereochemistry of MBW complexes.

involved the R3 repeat of the MYB (with a conserved motif $[D/E]L_{X2}[R/K]_{X3}L_{X6}L_{X3}R$) and the N-terminal MYB-interacting region (MIR) of the bHLHs, which contains an arginine residue conserved among the bHLHs of the IIIf subgroup. It was also demonstrated that these MYBs and bHLHs, together with TTG1, could form MBW ternary protein complexes allowing to draw a simple model on the cooperation between the R2R3-MYB and R/B-like bHLH regulators (Fig. 1-8j) (Baudry *et al.*, 2004; Lepiniec *et al.*, 2006; Gonzalez *et al.*, 2008).

Nevertheless, although the bHLHs and some MYBs (e.g. TT2) can form homodimers in yeast, the precise stereochemistry of the complexes remains to be determined *in planta* (Figure 1-5c). The exact molecular role of each of the three partners is also not yet fully understood. For instance, although there are some evidences suggesting that the bHLHs can bind DNA, its direct binding may be dispensable (Quattrocchio *et al.*, 2006; Hernandez *et al.*, 2007). On the contrary, mutations of the MYB-interacting region can induce bHLH transcriptional activity, suggesting that MYB could have regulatory role (Pattanaik *et al.*, 2008). Last, although TTG1 is critical for the activity of the MBW complexes *in planta*, it is dispensable in yeast and in moss protoplasts, although one cannot exclude that some heterologous proteins can replace the WDR partner in these systems (Baudry *et al.*, 2004; Thevenin *et al.*, 2012). One attractive hypothesis is that the WDR protein is necessary to prevent the effect of a negative regulator, allowing stabilization and/or nuclear localization of the MBW complexes, or controlling the interaction with chromatin factors (de Vetten *et al.*, 1997; Payne *et al.*, 2000; Baudry *et al.*, 2004; Broun, 2005; Caro *et al.*, 2007; Hernandez *et al.*, 2007; Feller *et al.*, 2011). Nevertheless, these later hypotheses are difficult to reconciliate with the positive effect of WDR in transient expression in protoplasts from Arabidopsis, maize or moss or in yeast experiments (Baudry *et al.*, 2004; Hernandez *et al.*, 2007; Thevenin *et al.*, 2012). In addition, the ectopic expression of different bHLHs can (partially) complement *ttg1* mutants (Zhang *et al.*, 2003) and TTG1 has a quantitative effect on TT8 activity (Baudry *et al.*, 2004), suggesting that the TTG1 protein plays a role at the post-translational level by regulating its activity in a quantitative manner, which in turn suggests a stabilization role for TTG1.

1.2.2 DNA-binding sites of the MBW complexes

Genetic analyses suggested that MYB partners often determined the involvement of the MBW complexes in specific pathways (Fig. 1-6) (Lepiniec *et al.*, 2006). Supporting this idea, a few amino acid residues of the MYB domain might account for the target gene specificity (e.g. converting the regulatory specificity between TT2 and PAP4) (Heppel *et al.*, 2013; Lai *et al.*, 2013). The MYB binding sites recognised by MBW complexes can be gathered into two major groups, the MYB core elements (*5'-CNGTTR-3'*, also called MBSI) and "*AC*-rich" elements (*5'-[A/C]CC[A/T]A[A/C]-3'*), whereas the bHLHs are known to bind E-box (*5'-CANNTG-3'*) (Koshino-Kimura *et al.*, 2005; Ryu *et al.*, 2005; Ishida *et al.*, 2007; Kang *et al.*, 2009; Song *et al.*, 2011). The specificity of the MBW complexes for different targets most probably relies on the participation of both R2R3-MYB and bHLH partners, as well as other regulators, as described below (Toledo-Ortiz *et al.*, 2003; De Masi *et al.*, 2011; Feller *et al.*, 2011; Prouse & Campbell, 2012).

1.2.3 Post-translational regulation of MBW complexes

During the last years various posttranslational regulations of MBW activities have been reported, which can finely tune the transcriptional activity of the MBW complexes (Fig. 1-7) (Dubos *et al.*, 2010; Feller *et al.*, 2011; Li, 2014). The two bHLH proteins, GL3 and EGL3 are targets of ubiquitin/26S proteasome **(**UPS)-dependent proteolysis, by UPL3/KAKTUS, an HECT E3 ubiquitin ligase (Downes *et al.*, 2003; El Refy *et al.*, 2003; Patra *et al.*, 2013c). Interestingly, the induction of *UPL3* expression by *GL3* triggers a negative regulatory feedback (Patra *et al.*, 2013b). Similarly, the MYB proteins PAP1 and PAP2 are degraded in darkness by UPS proteolysis mediated by the COP1/SPA E3 ligase (Maier *et al.*, 2013). TT8 and TTG1 are also likely to be targeted for proteasomal degradation, but the E3 ligase proteins involved have not yet been identified (Patra *et al.*, 2013a).

Small R3-MYBs that possess a single R3 domain and lack a putative activation domain can inhibit MBW complex (Dubos *et al.*, 2008; Matsui *et al.*, 2008; Zhu *et al.*, 2009; Thevenin *et al.*, 2012; Wang & Chen, 2014). These proteins can interact with different bHLHs of the MBW complexes, as demonstrated for MYBL2 with MYC1, EGL3, GL3, and

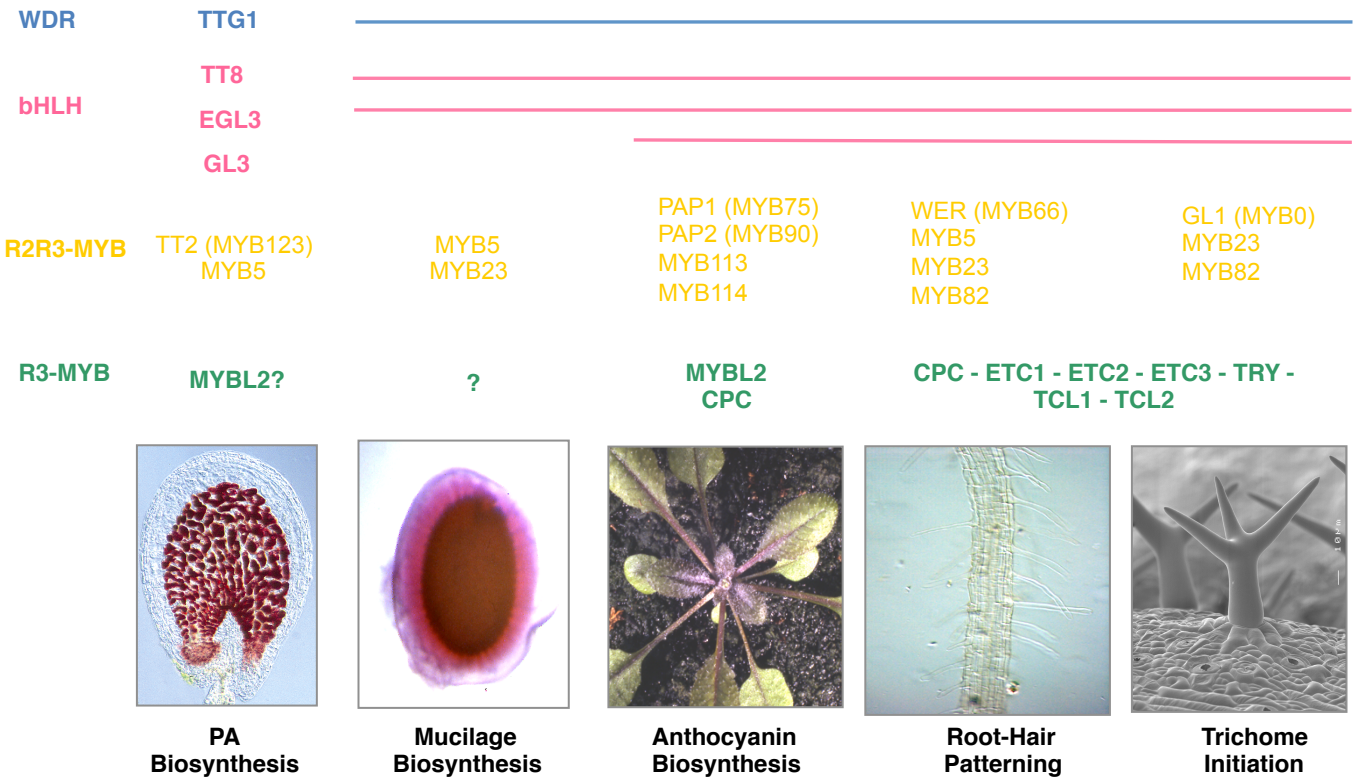


Figure 1-6. MYB-bHLH-WDR complexes and redundant functions in *Arabidopsis thaliana* (Adapted from Lepiniec *et al*., 2006).

TT8. These interactions can disrupt the interaction with the R2R3-MYBs and/or the binding of the complex to DNA (Dubos *et al.*, 2008). MYBL2, has been recently shown to be stabilized by phosphorylation (Ye *et al.*, 2012). Similarly, SPL9, a member of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE family, also represses anthocyanin biosynthesis through a direct interaction with R2R3-MYBs (PAP1 and MYB113) and binds to the target gene promoter (*DFR*) (Gou *et al.*, 2011). Thus, miR156 that targets *SPL9* can indirectly induce the expression of anthocyanin pathway genes. TCP3, a bHLH protein, interacts with both R2R3-MYBs (PAP1, PAP2, and TT2), but also with MYBL2, strengthening the transcriptional activity of the MBW complexes when overexpressed (Li & Zachgo, 2013; Li, 2014). Finally, TT1 protein would interact directly with TT2, enhancing the activity of the MBW protein complex (Appelhagen *et al.*, 2011b).

As described hereafter some MBW complexes and small R3-MYB are not only involved in the regulation of the flavonoid pathway but also in the regulation of different cell fates in Arabidopsis, as well as in other plants (Broun, 2005). Some of these complexes are regulated by nuclear trapping and intercellular movement. For instance, some R3-MYBs are able to move from one cell to another to affect trichome initiation and root-hair patterning by competing with R2R3-MYBs for binding to their bHLH partners (Savage *et al.*, 2008; Wester *et al.*, 2009; Schiefelbein *et al.*, 2014). TTG1 that is also a mobile molecule is trapped in trichomes by GL3, preventing its diffusion in trichome neighbouring cells (Bouyer *et al*., 2008; Pesch & Hulskamp, 2009; Balkunde *et al*., 2011). In addition, physical interaction of TTG1 with GEM (GL2 EXPRESSION MODULATOR) would prevent the formation of MBW complexes modulating GL2 expression and root hair formation (Caro *et al.*, 2007). The dimerization and interaction of the maize bHLH R, with RIF1 (a protein involved in chromatin remodelling) regulates its activity (Hernandez *et al.*, 2007; Kong *et al.*, 2012). Recently it has been shown that BIN2, a GSK3-like kinase, induced by brassinosteroid can inhibit MBW activity by phosphorylating both TTG1 and EGL3/GL3 (Cheng *et al.*, 2014). Other modifications including phosphorylation, disulfide bridge formation or cysteine nitrosylation could affect bHLH, MYB or TTG1 activities (Kwon *et al.*, 2007; Feller *et al.*, 2011), but the biological significance of these modifications for the regulation of MBW activity remains to be demonstrated in most cases.

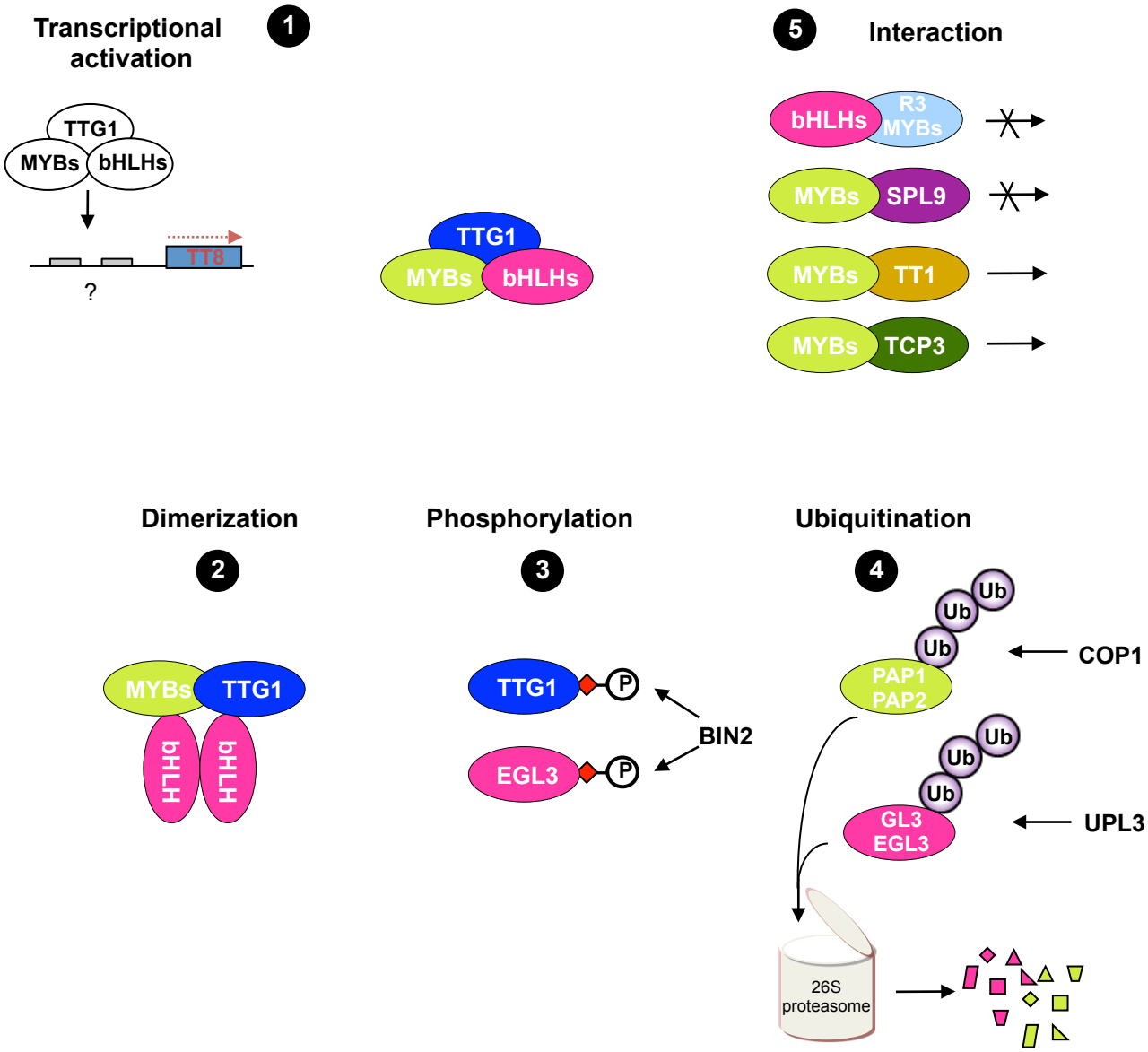


Figure 1-7. Regulation of MBW complexes in *Arabidopsis thaliana*. (1) Transcriptional regulation of *TT8* expression involves TTG1-dependant MBW complexes. It was not known for the binding sites that MBW complexes recognized on the *TT8* promoter before my PhD. (2) bHLHs (TT8, GL3 and EGL3) may occur through dimerization, based on the sequence homology of these bHLHs with maize bHLH R that owns an ACT-like domain participating in the dimerization (Feller *et al*., 2006). (3) BIN2 phosphorylates TTG1 and EGL3 to modulate root epidermal patterning (Cheng *et al*., 2014). (4) Posttranslational degradation of components of the MBW complexes by the 26S ubiquitin proteasome system (Maier *et al*., 2013; Patra *et al*., 2013b, c). (5) The components of MBW complexes interact with different regulators to activate or repress the activities of the complexes (Appelhagen *et al*., 2011; Dubos *et al*., 2008; Gou *et al*., 2011; Li & Zachgo, 2013; Li, 2014; Matsui *et al*., 2008; Zhu *et al*., 2009; Thevenin *et al*., 2012; Wang & Chen, 2014).

1.3 Regulation of MBW genes expression

1.3.1 Developmental regulation

The developmental regulation of MBW genes expression is well characterized in Arabidopsis, especially those controlling flavonoid biosynthesis in the seed coat or in cotyledon (Fig. 1-8). For instance, *TTG1*, the two R2R3-MYB *TT2/MYB5* and the three bHLH *TT8/EGL3/GL3* are all expressed in PA accumulating cells of the seed coat (Fig. 1-8) (Nesi *et al.*, 2000; Nesi *et al.*, 2001; Debeaujon *et al.*, 2003; Baudry *et al.*, 2006; Gonzalez *et al.*, 2009; Li *et al.*, 2009). Similarly, *TTG1*, *TT8/GL3/EGL3*, and *PAP1/PAP2* are expressed in anthocyanin accumulating cells in seedlings (Zhang *et al.*, 2003; Baudry *et al.*, 2006; Gonzalez *et al.*, 2008; Maes *et al.*, 2008). These specific patterns of MBW genes expression are fully consistent with the genetic analyses. Indeed mutations of the three bHLHs or the two R2R3-MYBs (PAP1 and PAP2) are necessary for preventing anthocyanin accumulation in seedlings and, similarly, the mutations of TT2 and MYB5 or of the 3 bHLHs are required to fully prevent the expression of the flavonoid genes and accumulation of anthocyanins in seedlings and PA in the seed coat, respectively (Zhang *et al*., 2003; Gonzalez *et al*., 2008).

The characterization of the positive feedback regulation of *TT8* expression, involving MBW complexes, was a key step in the understanding of the molecular mechanisms regulating the strong and specific accumulation of PA in the endothelium (Fig. 1-9) (Baudry *et al.*, 2006). It has been shown that the ectopic expression of TT2 induced *TT8* expression in all tissues where TTG1 is expressed (Nesi *et al.*, 2001; Baudry *et al.*, 2004). Furthermore, *TT8* transcript accumulation strongly decreases in developing siliques of the *tt8 egl3 gl3* triple mutant, as in the *ttg1* mutant (Fig. 1-9g) (Baudry *et al.*, 2006). Some additional studies confirmed that MBW complexes with MYB5 and GL3, or PAP1, were involved in the accumulation of *TT8* mRNA in developing siliques or vegetative tissues, respectively (Fig. 1-9h) (Baudry *et al.*, 2006; Gonzalez *et al.*, 2008; Gonzalez *et al.*, 2009). Interestingly the level of *TT8* mRNA in these mutants was not null, suggesting that some part of the *TT8* expression was independent of the MBW complexes activities (Baudry *et al.*, 2006).

TT16 is involved in the control of cell differentiation and divisions, as well as *BAN* expression in the endothelium, but not in chalaza/micropyle region (Nesi *et al.*, 2002;

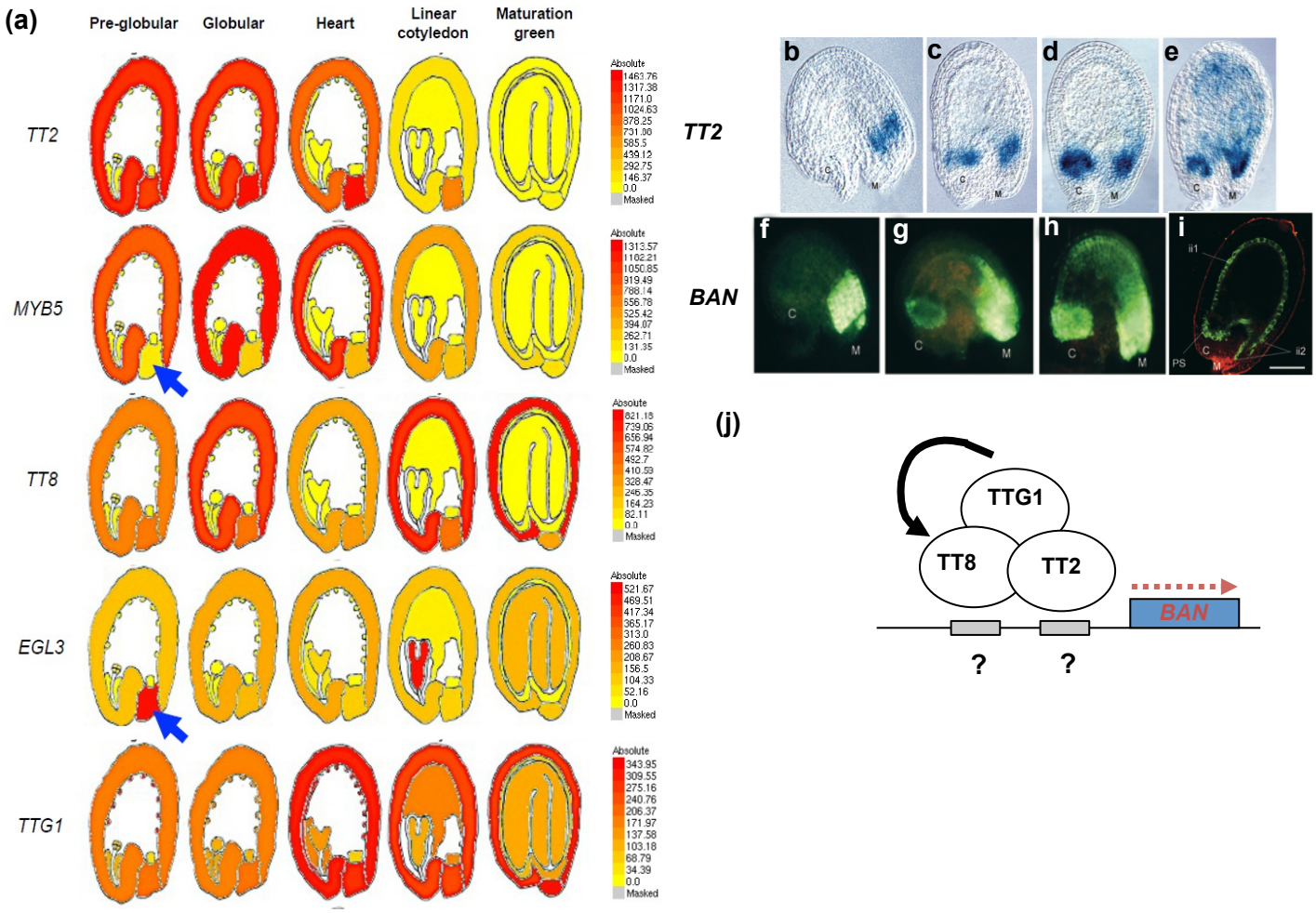


Figure 1-8. Tissue-specific pattern of flavonoid regulator gene and *BAN* promoter activity. (a) Steady state mRNA level of the MBW genes in developing Arabidopsis seeds. Data were generated by combining laser capture micro dissection coupled with ATH1 GeneChip analysis (Le *et al*., 2010). Schematic representations are available at the following URL: http://bar.utoronto.ca/efp/cgi-bin/efpWeb. Blue arrows indicate the absence or presence of mRNA accumulation in the chalaza for MYB5 and EGL3, respectively. (b-e) Pattern of TT2 promoter activity a young fertilized ovule (b), the quadrant stage of embryo development (c), the early globular stage (d) and the globular stage (e) revealed by the detection of GUS in transformants expressing *ProTT2:uidA* (Debeaujon *et al*., 2003). (f-i) Expression of *ProBAN1:mGFP5-ER*. GFP activity is observed on whole mounts, with a standard fluorescence microscope, (f) mature unfertilized ovule, (g) one-cell stage of embryo stage, (h) two-cell stage, (i) Seed at the late globular stage (Debeaujon *et al*., 2003). GFP activity is observed on confocal sections. (j) Model for the structure and regulation of the transcriptional activating complex composed by TT2, TT8, and TTG1 (Adapted from Baudry *et al*., 2004).

Debeaujon *et al.*, 2003), linking cell development and PA deposition. Nevertheless, the ectopic expression of *TT2* is sufficient to induce PA accumulation in the endothelium of *tt16* suggesting that TT16 acts at least indirectly on TT2 and/or MBW activity. Similarly, TT1 mutations have also been shown to alter the morphology of the endothelium and give rise to yellow seeds that display a weak PA accumulation (Sagasser *et al.*, 2002). In addition TT1 can interact directly with TT2, suggesting that it may be necessary to maintain the activity of the MBW complex (Appelhagen *et al.*, 2011b). TTG2 controls PA biosynthesis and outer integument development, but has been suggested to act downstream of TTG1 and TT2 (Johnson *et al.*, 2002; Ishida *et al.*, 2007). Nevertheless, the similarities between the *ttg2* and *ttg1* phenotypes also suggested that TTG2 could act on the activity of the MBW complexes, so indirectly on *TT8* expression. Altogether these results suggest that these transcriptional regulators control the activity of the MBW complex at different levels, but their direct targets remain to be identified. Similarly, ANTHOCYANINLESS2 (ANL2) a homeodomain protein belonging to the GLABRA2 group is required for the accumulation of anthocyanin in leaves but also for root development, suggesting that it could act indirectly on MBW complex activity (Kubo *et al.* 1999). Last, *ABERRANT TESTA SHAPE (ATS)/KANADI4 is* a TF of the GARP family involved in integuments development (McAbee *et al.*, 2006). When ectopically expressed, it can bind the promoters of *TT2*, *TT8* and *TTG1* inhibiting PA accumulation (Gao *et al.*, 2010). It would be interesting to investigate further the physiological significance of these regulation and its interactions with TT1, TT16 or TTG2.

1.3.2 Physiological regulation

Flavonoid biosynthesis is induced by various biotic or abiotic environmental factors (e.g. high light, extreme temperatures, nutriments) and many of them directly target the expression of the *bHLH* and/or *R2R3-MYB* of the MBW complexes, whereas TTG1 seems to be constitutively expressed (Fig. 1-10) (Vandenbussche *et al.*, 2007; Cominelli *et al.*, 2008; Dubos *et al.*, 2008; Brown *et al.*, 2009; Morishita *et al.*, 2009; Olsen *et al.*, 2009; Shi & Xie, 2014). For instance, HY5 (a bZIP protein) appears as a key integrator of light and cold stress for inducing flavonoid biosynthesis via direct binding to G- and ACE-boxes found in the

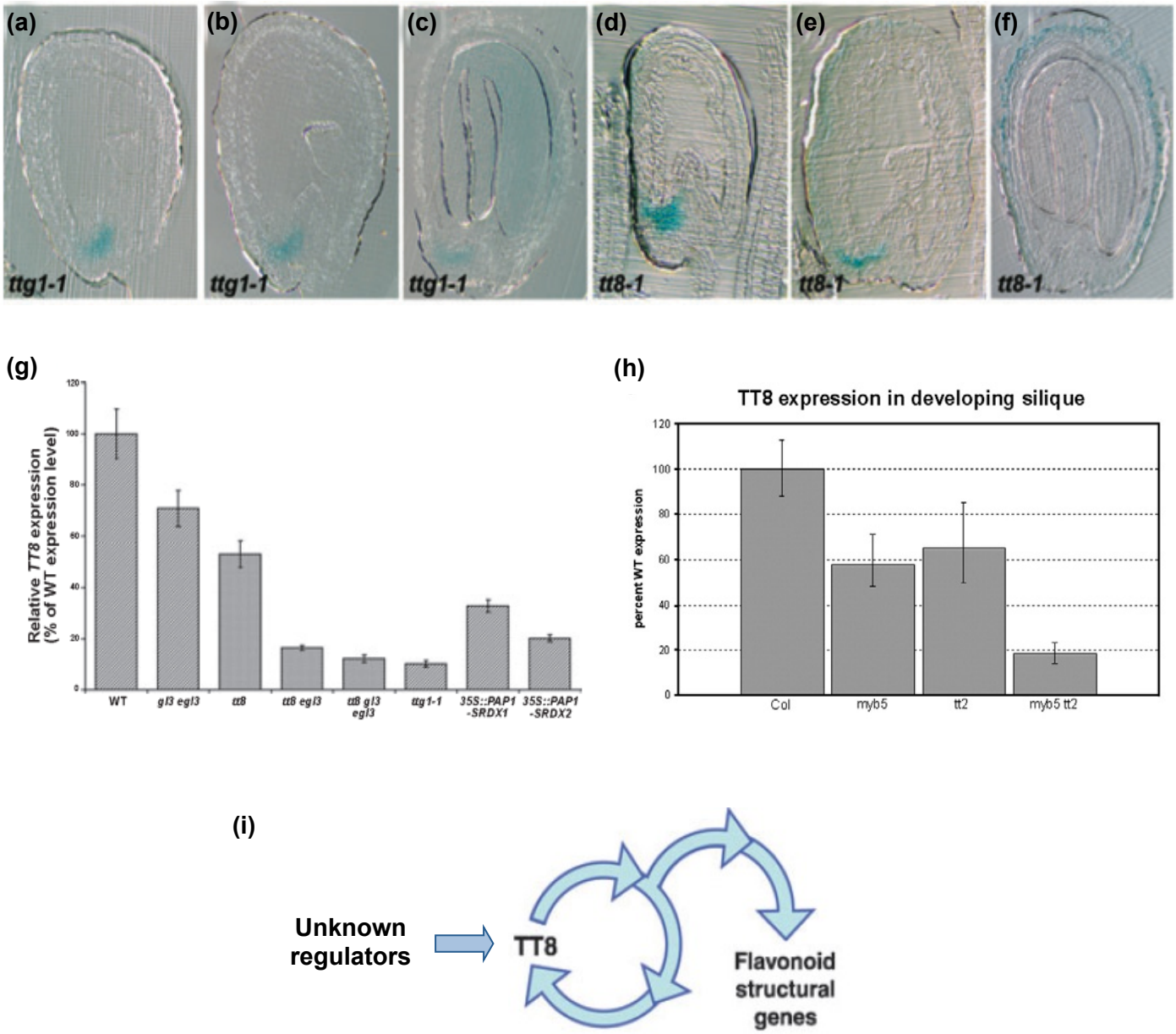


Figure 1-9. *TT8* expression in developing siliques of *Arabidopsis thaliana* (Baudry *et al*., 2006; Gonzalez *et al*., 2009). GUS activity of the *TT8* promoter in seeds of *ttg1-1* (a-c) and *tt8-1* (d-f). TT8 expression levels were measured in the siliques of several mutant backgrounds at the heart stage of embryo development (g). *TT8* expression levels were measured in developing siliques of *tt2*, *myb5* and *myb5tt2* (h). Q-PCR analysis of *TT8* and *BAN* expression in developing siliques reported as percentage of wild-type expression. Error bars indicate range of expression. Model for the regulatory network in the Arabidopsis seed coat. (i) Model for the TT8 regulatory network in the *A. thaliana* seed coat.

promoter of *PAP1* (Shin *et al.*, 2007; Catala *et al.*, 2011; Maier *et al.*, 2013; Shin *et al.*, 2013). STH2, a B-box protein interacts genetically with HY5 to enhance the levels of anthocyanin accumulation (Datta *et al.*, 2007). A NAC protein (ANAC078) is involved in the regulation of flavonoid biosynthesis in response to high light stress, participating to the induction of the *bHLH* (*GL3*, *EGL3* and *TT8*) and *R2R3-MYB* (e.g. *PAP1*) expression (Morishita *et al.*, 2009). LZP1, a zinc finger protein acting downstream HY5, may also induce PAP1 (Chang *et al.*, 2008). Last, high light treatments inhibit the expression of *MYBL2* contributing to the induction of *TT8* and *PAP1*/*PAP2* and anthocyanin accumulation (Dubos et al., 2008). Conversely, high temperature and low light reduce the expression of *TT8*, *EGL3* and *TTG1* and anthocyanin accumulation in Arabidopsis leaves (Rowan *et al.*, 2009).

Sucrose treatment is another well-known inducer of flavonoid genes leading to anthocyanin accumulation (Teng *et al.*, 2005; Solfanelli *et al.*, 2006; Loreti *et al.*, 2008). Surprisingly, sucrose induces a complex regulatory network involving PAP1 and/or closely related PAP2 and MYB113/PAP3, and *Trans-Acting SiRNA gene 4 (TAS4*) allowing fine-tuned and reversible induction of flavonoid biosynthesis (Luo *et al.*, 2012). *TAS4* encodes a small interfering RNA subjected to miR828-directed cleavage, targeting *PAP1* and *PAP2* (Luo *et al.*, 2012; Rock, 2013; Yang *et al.*, 2013). The expression of both *PAP1* and *PAP2* but also *TAS4* was induced specifically by exogenous treatment with sucrose, inducing a negative regulatory feedback controlling *PAP1* and *PAP2* accumulation (Luo *et al.*, 2012). Similar regulation is triggered by phosphate limitation and ABA that induce *PAP1* and *PAP2*, but also the expression of *TAS4* and *miR828* (Hsieh *et al.*, 2009; Luo *et al.*, 2012; Rock, 2013). Interestingly, MYBL2 is another putative target of mir858 adding a level of complexity to this regulatory network.

When inducing anthocyanin production, sucrose interacts synergistically with phosphate and ABA, but also with JA and other hormones involved in the responses to various biotic and abiotic stresses, and antagonistically with GA (Loreti *et al.*, 2008; Qi *et al.*, 2011; De Geyter *et al.*, 2012; Li, Y *et al.*, 2014; Qi *et al.*, 2014; Shi & Xie, 2014). Briefly, DELLAs and JASMONATE-ZIM DOMAINs (JAZs) proteins act as key cross-talking components of GA and JA signalling cascades, respectively, and interact with different partners of the MBW complexes. CTK enhances sucrose-induced anthocyanin biosynthesis through transcriptional

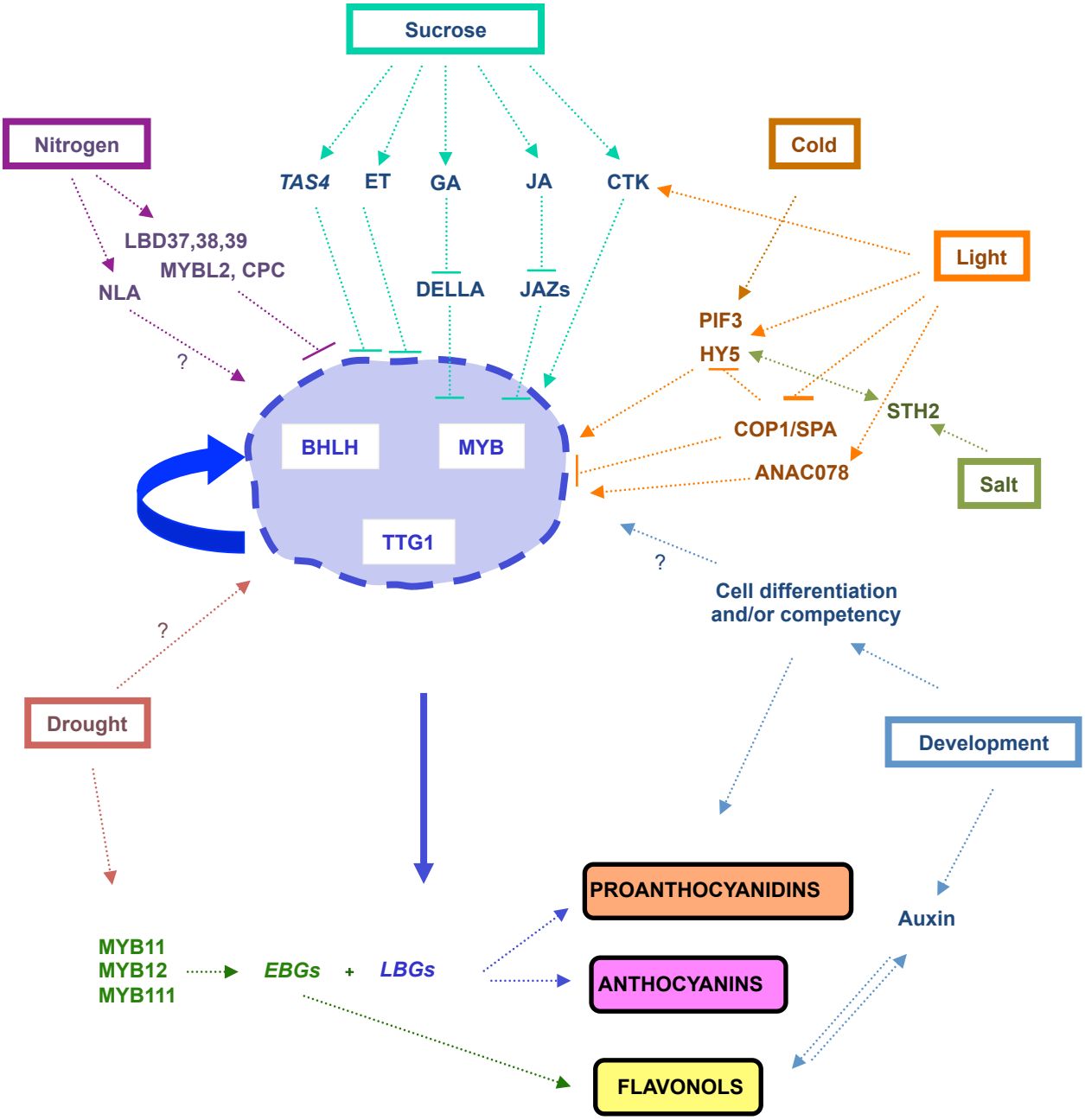


Figure 1-10. Environmental and developmental regulation of flavonoid biosynthesis. References involved are given in the text.

activation of the MBW complexes, while suppressing MYBL2 expression (Das *et al*., 2012). It was also demonstrated that CTK anthocyanin enhancement is partially dependent on phytochrome and HY5, but mainly on photosynthetic electron transport (Das *et al*., 2012). By contrast, Ethylene can inhibit anthocyanin accumulation induced by sucrose and light through suppressing the expression of *TT8, GL3*, and *PAP1*, while stimulating concomitant expression of the negative regulator *MYBL2* (Jeong et al., 2010). In addition MYBL2 may play a role in the Brassinosteroids (BR) signalling pathway and its expression is controlled by a negative regulatory feedback involving BES1 (an atypical bHLH) (Ye *et al.*, 2012). Last, it has been shown in cell culture that different auxins control anthocyanin biosynthesis mainly by inducing the expression of *TT8*, *GL3* and *PAP1* (Liu *et al.*, 2014).

Nitrogen deficiency (low N/C balance) also triggers flavonoid accumulation in seedlings and rosette leaves, especially at low temperature, by inducing *PAP1*, *PAP2* and *GL3* expression (Pourtau *et al.*, 2006; Lea *et al.*, 2007; Lillo *et al.*, 2008; Feyissa *et al.*, 2009; Olsen *et al.*, 2009; Nemie-Feyissa *et al.*, 2014; Shi & Xie, 2014). Interestingly, the expression of some small R3-MYBs was also induced, suggesting there might be a negative feedback loop to avoid overaccumulation of anthocyanins. In this regard, three members of the LATERAL ORGAN BOUNDARY DOMAIN (LBD) protein family, LBD37, 38 and 39, are also repressors of anthocyanin biosynthesis, induced by N/NO_3, and acting upstream of *PAP1* and *PAP2* (Rubin *et al.*, 2009; Zhou *et al.*, 2012). Furthermore, NLA (NITROGEN LIMITATION ADAPTATION), an E3 ligase, is involved in this adaptation to N limitation, but the molecular link with MBW complexes remains to be investigated (Peng *et al.*, 2008). Interestingly, it has been recently shown that autophagy mutants are affected in the expression of *PAP1, PAP2,* and *TT8,* that is fully consistent with the lower level of flavonoid accumulation compared to wild type plants (Masclaux-Daubresse *et al.*, 2014). Some microbe-based fertilizers have been shown to induce the expression of PAP2, MYB113, MYB114, EGL3 and TT8 (Ali & McNear, 2014).

1.4 MBW complexes are involved in other pathways and found in all plant species

1.4.1 Arabidopsis MBW complexes also control epidermal cell fate

In Arabidopsis, some of the MBW complexes are involved in trichome formation, root hair patterning, seed coat differentiation and mucilage production, suggesting that a common regulatory module has been retained for epidermal cell fates (Figure 1-6) (Lee & Schiefelbein, 2001; Zhang *et al.*, 2003; Broun, 2005; Ramsay & Glover, 2005; Lepiniec *et al.*, 2006; Serna & Martin, 2006; Ishida *et al.*, 2007; Zhao *et al.*, 2008; Gonzalez *et al.*, 2009; Schiefelbein *et al.*, 2009; Benítez *et al.*, 2011; Song *et al.*, 2011; Symonds *et al.*, 2011; Tominaga-Wada *et al.*, 2012; Zhao *et al.*, 2012; Cheng *et al.*, 2014; Rishmawi *et al.*, 2014; Schiefelbein *et al.*, 2014). The MBW complexes involved contain at least one of the bHLH from subgroup IIIf (GL3, EGL3, TT8, or MYC1), and an R2R3-MYB from subgroup S15 (WER, GL1 and AtMYB23) and AtMYB5 or AtMYB82. For instance GL1, MYB23 and AtMYB82 control trichome development, WER root-hair patterning and MYB5 mucilage production (Li *et al.*, 2009; Morohashi & Grotewold, 2009; Song *et al.*, 2011; Zhao *et al.*, 2012; Liang *et al.*, 2014). WER and GL1 are functionally similar, but their expression patterns determine their specific functions. Trichome formation on leaf lamina relied on GL3 and EGL3, whereas TT8 is required for marginal trichome development of rosette leaves (Maes *et al.*, 2008).

As for flavonoids, some small R3-MYBs, namely CPC (CAPRICE), TRY (TRIPTYCHON), ETC1 (ENHANCER OF TRY AND CPC 1), ETC2, ETC3/CPL3 (CAPRICE-LIKE 3), TCL1 (TRICHOMELESS 1) and TCL2/CPL4 are also controlling these pathways (Dubos *et al.*, 2010; Benítez *et al.*, 2011; Gan *et al.*, 2011; Wang & Chen, 2014). They all display within their MYB-domain the amino acid signature known to support the interaction with the bHLH (i.e. [D/E]Lx$_2$[R/K]x$_3$Lx$_6$Lx$_3$R) and inhibit MBW complexes by sequestering bHLHs partners. For instance, the WER-GL3/EGL3-TTG1 complexes induce the expression of *GLABRA2* (*GL2*) and *CPC* in non-root hair cells. CPC moves from the non-root hair cells to root hair cell files and bind to GL3 and EGL3 inhibiting the activity of the MBW complexes, and GL3 moves in the opposite direction. By contrast, TCL1 directly inhibits the expression of *GL1* (Wang *et al.*, 2007; Tominaga-Wada & Wada, 2014). Interestingly, it has been recently demonstrated that BR signalling inhibits root hair formation

through BIN2, a GSK3-line kinase that phosphorylates EGL3 and TTG1, preventing EGL3 from moving to non hair cell and inhibiting MBW activity (Cheng *et al*. 2014).

1.4.2 Conserved genes, functions, and regulations in other plants

Similar MBW regulatory genes are found in angiosperms as well as in gymnosperms and mosses (Zimmermann *et al.*, 2004; Koes *et al.*, 2005; Bedon *et al.*, 2007; Bedon *et al.*, 2010; Feller *et al.*, 2011). The bHLH-interacting domain found in R2R3-MYBs involved in MBW is well conserved among higher plant species, suggesting that at least MYB and bHLH interactions and probably MBW complexes arose early in land plant evolution. MBW partners have been found first in maize and then in *Antirrhinum,* petunia, grapevine, apple, strawberry, morning glory, rice or cabbage and all angiosperms analysed so far (Fig. 1-11) (Mol *et al.*, 1998; Xue *et al.*, 2003; Quattrocchio *et al.*, 2006; Pires & Dolan, 2010; Feller *et al.*, 2011; Hichri *et al.*, 2011a; Petroni & Tonelli, 2011; Falcone Ferreyra *et al.*, 2012; Jaakola, 2013; Lai *et al.*, 2013; Patra *et al.*, 2013c; Schaart *et al.*, 2013; Albert *et al.*, 2014). For instance, MBW complexes have been well characterized in maize (C1/Pl-R/B-PAC1), petunia (AN2/AN4-AN1-AN11) or in strawberry (MYB9/MYB11-bHLH3-TTG1) (Spelt *et al.*, 2000; Petroni & Tonelli, 2011; Schaart *et al.*, 2013).

In all plants studied so far MBW complexes determine the spatio-temporal expression of target genes that accounts for tissue-specific accumulation of anthocyanins or PAs. Some TFs from monocots can control the expression of the entire part of the pathway, contrary to the specific control of *LBGs* in Arabidopsis. For instance maize *C1* and *Pl* MYB genes determine the pattern of anthocyanin synthesis in the aleurone layers (Lai *et al.*, 2013). Some studies indicate that *EBG* expression can be closely correlated with R2R3-MYB regulator genes and R2R3-MYBs can induce *EBGs* in transient or transgenic plants, but there is no evidence demonstrating that R2R3-MYBs of dicot species directly activate *EBGs* (Lai *et al*., 2013). Therefore the specific control of *LBG* in Arabidopsis may be a recent trait (eg. TT2 specificallly inducing *LBG* and PA in the testa). It is to be noticed that specific flavonoid pathways, such as phlobaphenes in maize or flavonols in Arabidopsis, are regulated by different R2R3-MYBs that are not acting through bHLH interactions, namely P in maize and

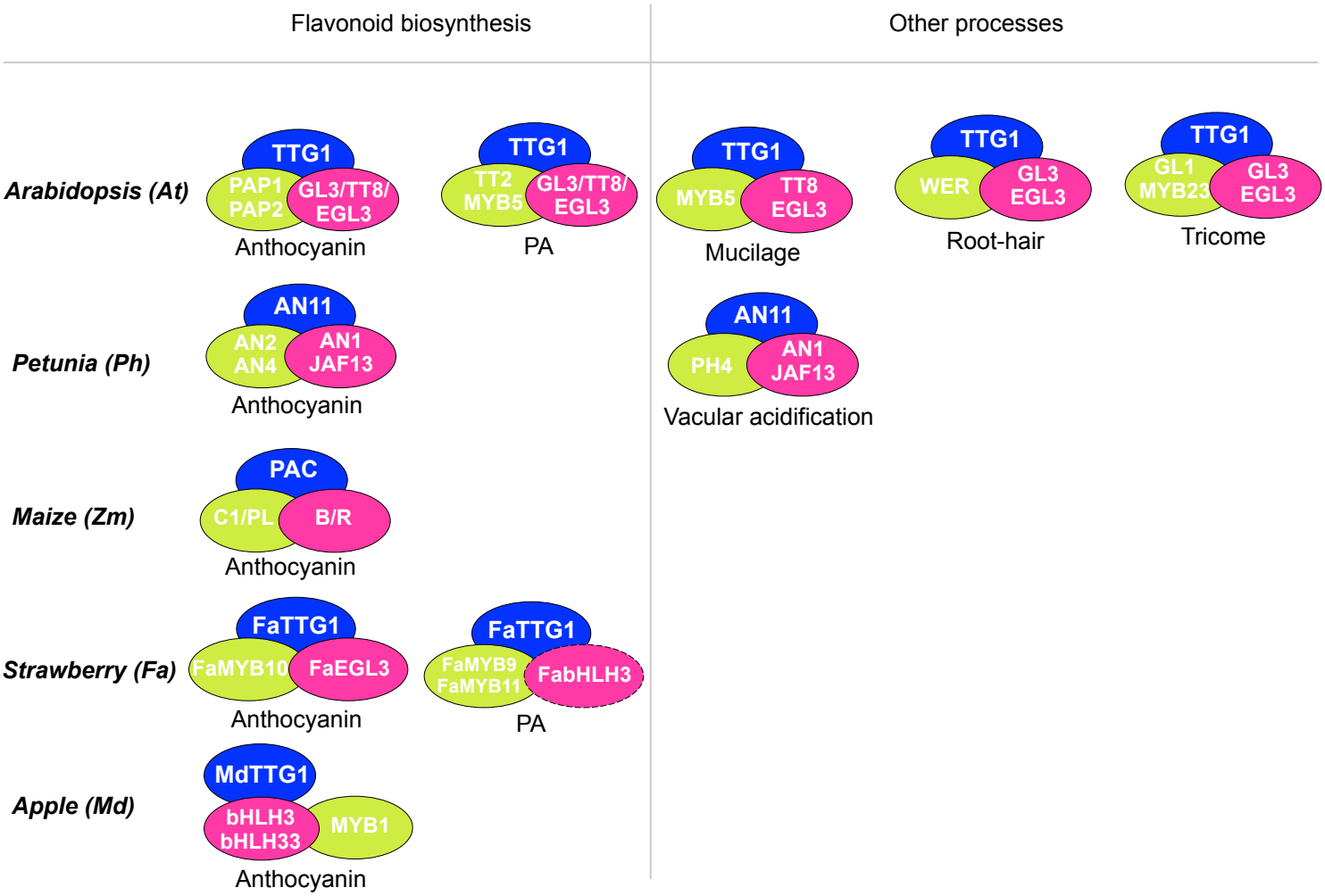


Figure 1-11. Schematic representation of protein MBW complexes in plants. Different plant species include: Arabidopsis (*Arabidopsis thaliana*), petunia (*Petunia hybrida*), maize (*Zea maize*), strawberry (*Fragaria ananassa*) and Apple (*Malus domestica*). bHLH (pink), R2R3-MYB (green) and WDR proteins (blue) in dotted lines indicate that an interaction has been proposed but not experimentally proven. References involved are given in the text.

MYB11, 12 and 111 in Arabidopsis (Mehrtens *et al.*, 2005; Stracke *et al.*, 2007; Feller *et al.*, 2011; Lai *et al.*, 2013).

Small R3-MYB repressors have also been identified in other species. For instance, in petunia the R3-MYB "MYBx", modulates the transactivation of the MBW complex (Albert *et al.*, 2014). Interestingly, activation of AN1, MYB27 (an R2R3-MYB with EAR inhibitory domain) and MYBx by the MBW complex and the inhibition of AN1 activity by the small MYBx suggest that a negative regulatory feedback, similar to those found in Arabidopsis, may be widespread in dicotyledonous species (Spelt *et al.*, 2000; Albert *et al.*, 2014). Intercellular movement is also found in petunia for AN11 and MYBx, which may facilitate anthocyanin pigment pattern formation (Albert *et al.*, 2014). The positive feedback regulation of *TT8* expression, involving MBW complexes, (Baudry *et al.*, 2006) might be also evolutionary conserved, as exemplified by the self activation of *AN1* and *VvMYC1* expression in petunia and grapevine, respectively (Spelt *et al.*, 2000; Hichri *et al.*, 2010). Last, miR828 has been found in various angiosperms but also in gymnosperms, suggesting that it plays an important and conserved role in plant fitness, consistent with flavonoid functions (Rock, 2013).

Although the flavonoid R2R3-MYBs are well conserved, their DNA-binding specificities may vary depending on plant species and interactions with the bHLH factors (Lai *et al.*, 2013). Nevertheless, as in Arabidopsis, at least MYB-core binding sites and E-box were identified in the promoters of most of the target genes (Roth *et al.*, 1991; Koshino-Kimura *et al.*, 2005; Ryu *et al.*, 2005; Quattrocchio *et al.*, 2006; Hichri *et al.*, 2011a; Song *et al.*, 2011; Kong *et al.*, 2012; Lai *et al.*, 2013). In addition, maize R/C1 complex can have different conformations activating targets through different sequences (*5'-CAGTTA-3'* and G-box motif found in *BZ1,* or C1 binding site in *A1*) (Kong *et al.*, 2012; Patra *et al.*, 2013c). Similarly, in grapevine, the target gene selectivity of VvMYB5b depends on the interaction with bHLH co-factors (Hichri *et al.*, 2011b). Some plants also display more specific and/or different regulatory mechanisms. For instance, the maize bHLH gene *INTENSIFIER 1* identified by recessive mutations that increase pigmentation of the kernel (Burr *et al.*, 1996). Although the gene is homologous to *TT8*, most of the mRNAs are incorrectly spliced and encode truncated proteins that could act as dominant negative factors.

Last, it is interesting to note that the other pathways involving MBW in Arabidopsis (i.e. trichome, root hair, or mucilage formation) or petunia (vacuolar pH) appear to be limited to some species (Serna and Martin, 2006; Quattrocchio *et al.*, 2006). In petunia, AN1/AN11 also control vacuole pH in petals, but not trichome formation (Spelt *et al.*, 2000; Spelt *et al.*, 2002). Similarly, the maize PAC1 (WDR) is specifically involved in the anthocyanin pathway but can complement all *ttg1* phenotypes in Arabidopsis (Selinger & Chandler, 1999; Carey *et al.*, 2004; Feller *et al.*, 2011). These results suggest that the control of the flavonoid pathway would be the ancestral function of the MBW complexes in plants. Nevertheless, because the MBW partners are all well conserved among the eukaryotes (Feller *et al.*, 2011), they probably have been recruited from another ancestral function present in all these organisms such as the regulation the cell cycle and/or cell differentiation.

1.5 Objectives of the thesis

At the start of my PhD, our understanding of the transcriptional control of flavonoid biosynthesis in Arabidopsis was becoming clearer, however some questions still needed to be answered.

Within the framework of this PhD thesis we wished to develop the study of the nature and spatio-temporal activity of the MYB-bHLH-WDR (MBW) complexes, and identify their direct targets. For this purpose we have tried to answer a set of linked but different basic questions:

i. How *TT8* expression is regulated in seeds?

- Is *TT8* expression regulated by the other known TFs (*i.e.* TT1, TT16 & TTG2)?

- How the MBW complexes feedback regulation occurs?

- Which *cis*-regulatory elements present on *TT8* promoter are involved in this regulation?

ii. How the entire flavonoid biosynthetic pathway is regulated by the MBW complexes in seeds and seedlings?

- Which biosynthetic genes are directly regulated, at the transcriptional level, by the MBW complexes?

- What is the nature of the MBW complexes involved in regulating the transcription of these biosynthetic genes?

- What are the *cis*-regulatory elements to which the MBW complexes bind *in vivo*?

In order to get dissection for the regulation of the PA biosynthetic pathway by the TT2-TT8-TTG1 MBW complex, a combination of genetic, biochemical and molecular methods was adopted during my PhD. First, gene expression analyses by quantitative RT-PCR and gene promoter activity by histochemical detection with GUS analysis in wild type and different loss-of-function mutant backgrounds were carried out. Then, chimeric inducible TFs were used to validate putative targets for the TTG1-dependant MBW complexes. In parallel, a new developed transient expression system that relies on the moss *Physcomitrella patens* protoplasts was coupled to the study of functional redundancy among the homologous proteins. Furthermore, the determination of the link existing between the MBW complex and the target promoters was studied using transient and stable expression in combination with yeast one-hybrid technique to proceed molecular dissection and functional analysis of the target gene promoters.

1.6 Presentation of the thesis

For this thesis, the presentation of the results in deciphering the MBW regulatory network controlling flavonoid biosynthesis in Arabidopsis seed is introduced into three chapters. Firstly, the results for the regulation of *TT8* expression are presented in the manner of an article published in the New Phytologist (Chapter II, Xu and Grain *et al*., 2013). The second chapter gives a detailed description for the nature and spatio-temporal activities of the MYB-bHLH-WDR complexes regulating flavonoid biosynthesis and their targets, in the form of an article published in the New Phytologist (Chapter III, Xu *et al*., 2014a). In the fourth chapter, we further analysed the transcriptional capacity of the TT2-TT8-TTG1 MBW complex through transactivation assays in moss protoplasts and overexpression in Arabidopsis siliques (Chapter IV). These results presented in Chapter IV provide new information for biotechnological engineering of PA biosynthesis in *planta*, and have been published in the Plant Signaling & Behavior (Xu *et al*., 2014b). During my PhD, I have been

also involved in the development of a new method to study the interactions between TFs and their target gene promoters by transient expression in *P. patens* protoplasts, working together with Johanne Thévenin in my laboratory under the direction of Dr Bertrand Dubreucq. In this case, I am the co-author of the publication about this new technique (Annexes: Article 4 - Thévenin and Dubos *et al*., 2012 in New Phytologist). The implementation of this new technique has been performed on another project, namely 'Characterisation of TT2, TT8 & TTG1 homolgues in strawberry'. Therefore, I am the co-authors of the additional publication (Annexes: Article 5 - Schaart *et al*., 2013 in New Phytologist). Last, the introduction of my thesis will be used for writing a review that is currently in preparation.

CHAPTER II

TRANSCRIPTIONAL REGULATION OF *TT8* EXPRESSION

Article 1: Regulation of flavonoid biosynthesis involves an unexpected complex transcriptional regulation of TT8 expression, in Arabidopsis

Wenjia Xu[1, 2, *], Damaris Grain[1, 2, *], José Le Gourrierec[1, 2], Erwana Harscoët[1, 2], Adeline Berger[1, 2], Vincent Jauvion[1, 2], Aurélie Scagnelli[1, 2], Nathalie Berger[1, 2], Przemyslaw Bidzinski[1, 2], Zsolt Kelemen[1, 2], Fabien Salsac[1, 2], Antoine Baudry[1, 2], Jean-Marc Routaboul[1, 2], Loïc Lepiniec[1, 2], Christian Dubos[1, 2, 3]

Accepted for the publication in New Phytologist (2013, vol.198)

[1] INRA, Institut Jean-Pierre Bourgin, Saclay Plant Sciences, RD10, F-78026 Versailles, France
[2] AgroParisTech, Institut Jean-Pierre Bourgin, Saclay Plant Sciences, RD10, F-78026 Versailles, France
[3] Current address: Université d'Angers, Institut de Recherche en Horticulture et Semences (INRA, Agrocampus-Ouest, Université d'Angers), SFR 149 QUASAV, F-49045 Angers, France

[*] These authors contributed equally to this work.

Authors for correspondence:
Christian Dubos
Tel: +33 (0) 4 9961 2613
Email: christian.dubos@versailles.inra.fr

Loïc Lepiniec
Tel: +33 (0) 1 3083 3686
Email: loic.lepiniec@versailles.inra.fr

2 Article 1

2.1 SUMMARY

• *TT8*/*bHLH042* is a key regulator of anthocyanins and proanthocyanidins (PAs) biosynthesis in *Arabidopsis thaliana*. TT8 transcriptional activity has been studied extensively, and relies on its ability to form, with several R2R3-MYB and TTG1 (WD-Repeat protein), different MYB-bHLH-WDR (MBW) protein complexes. By contrast, little is known on how *TT8* expression is itself regulated.

• Transcriptional regulation of *TT8* expression was studied using molecular, genetic and biochemical approaches.

• Functional dissection of the *TT8* promoter revealed its modular structure. Two modules were found to specifically drive *TT8* promoter activity in PA- and anthocyanin-accumulating cells, by differentially integrating the signals issued from different regulators, in a spatio-temporal manner. Interestingly, this regulation involves at least six different MBW complexes, and an unpredicted positive feedback regulatory loop between *TT8* and *TTG2*. Moreover, the results suggest that some putative new regulators remain to be discovered. Finally, specific *cis*-regulatory elements through which *TT8* expression is regulated were identified and characterized.

• Taken together, these results provide a molecular model consistent with the specific and highly regulated expression of *TT8*. They shed new light into the transcriptional regulation of flavonoid biosynthesis and provide new clues and tools for further investigation in Arabidopsis and other plant species.

2.2 INTRODUCTION

Flavonoids are secondary metabolites that play important roles throughout the plant life cycle. For example, flavonoids were shown to be involved in the attraction of pollinators or seed dispersers, in the defence against biotic or abiotic stresses and in the regulation of plant growth and development (Lepiniec *et al.*, 2006; Petroni & Tonelli, 2011; Grunewald *et al.*, 2012). In *Arabidopsis thaliana* three main classes of flavonoid compounds are accumulated:

anthocyanins in the vegetative tissues, proanthocyanidins (PAs) in seed integuments, and flavonols in both, seeds and vegetative tissues.

The expression of the genes that encode enzymes involved in the specific steps that lead to the biosynthesis of anthocyanins and PAs (i.e. the late biosynthetic genes) are controlled by the combined action of specific R2R3-MYB (subgroup 5 and 6) and R/B-like bHLH (subgroup IIIf) transcription factors (TF) together with TTG1 (WD-repeat protein), in a MYB-bHLH-WDR (MBW) ternary protein complex (Heim *et al.*, 2003; Baudry *et al.*, 2004; Lepiniec *et al.*, 2006; Dubos *et al.*, 2010; Thevenin *et al.*, 2012). *TT2/MYB123*, *TT8/bHLH042* and *TTG1* have been demonstrated as being the three main regulators of PA biosynthesis (Baudry *et al.*, 2004; Lepiniec *et al.*, 2006). By contrast, the regulation of anthocyanin biosynthesis was shown to involve a larger set of TFs, which includes *PAP1*/*MYB75*, *PAP2*/*MYB90, GL3/bHLH001*, *EGL3/bHLH002*, *TT8/bHLH042* and *TTG1* (Zhang *et al.*, 2003; Baudry *et al.*, 2006; Feyissa *et al.*, 2009; Appelhagen *et al.*, 2011a). Similar MBW complexes have been also characterized in other plant species such as maize or petunia (Hernandez *et al.*, 2004; Koes *et al.*, 2005; Albert *et al.*, 2011; Petroni & Tonelli, 2011). Transcriptional TTG1-dependent activities are not restricted to the regulation of flavonoid biosynthesis as, for example, additional MBW complexes involving R2R3-MYB belonging to subgroup 15 (in combination with GL3 and EGL3) were shown to regulate trichome formation and root hair patterning (Zhao *et al.*, 2008; Dubos *et al.*, 2010; Song *et al.*, 2011). The activity of MBW complexes is negatively regulated by a group of single MYB repeat proteins, the R3-MYBs. This group includes *MYBL2* a negative regulator of anthocyanin biosynthesis in vegetative tissues that can repress PA biosynthesis when overexpressed in seeds (Dubos *et al.*, 2008). Similarly to *MYBL2*, *CPC* (*CAPRICE*) has been proposed to regulate anthocyanin biosynthesis in response to nitrogen deprivation (Zhu *et al.*, 2009). Moreover *CPC*, together with six additional R3-MYB, has been shown to be involved in the regulation of trichomes and root hair patterning (Wang *et al.*, 2008; Gan *et al.*, 2011).

The regulation of PA biosynthesis involves some additional transcriptional regulators that belong to other families of TF, namely *TT1*/*WIP1* (Zn finger), *TT16*/*ABS*/*AGL32* (MADS box) and *TTG2*/*DSL1*/*WRKY44*, for which the precise mode of action remain to be elucidated (Johnson *et al.*, 2002; Nesi *et al.*, 2002; Sagasser *et al.*, 2002; Lepiniec *et al.*, 2006). It is

hypothesized that *TT1* and *TT16* regulate the competency and the differentiation, respectively, of cells that are destined to accumulate PAs, whereas *TTG2* is proposed to control the overall seed coat development, including the integument cell elongation process (Johnson *et al.*, 2002; Garcia *et al.*, 2005).

If the regulation of PA and anthocyanin biosynthesis by the MBW complexes is currently relatively well described and understood, the means by which the expression of the transcriptional regulators is regulated have been less investigated. It has been demonstrated that some environmental cues such as light intensity or temperature influence anthocyanin accumulation in the vegetative tissues by modulating the expression of the R2R3-MYB (e.g. *PAP1* and *PAP2*) and R/B-like bHLH (e.g. *TT8*, *GL3* and *EGL3*) TF (Dubos *et al.*, 2008a; Olsen *et al.*, 2009; Rowan *et al.*, 2009). Similarly, the carbon to nitrogen balance as well as the concentration of certain phytohormones, such as jasmonic acid, gibberelins or cytokinins, are also important factors that have an impact on the expression of these TF (Lea *et al.*, 2007; Maes *et al.*, 2008; Feyissa *et al.*, 2009). By contrast, the regulation of *TT2* and *TT8* expression in seeds appeared to be developmentally regulated (Lepiniec *et al.*, 2006). However, the precise nature of the molecular mechanisms that govern the expression of these anthocyanin and PA biosynthesis regulators is still elusive.

The regulation of *TT8* expression was chosen as a first step in the elucidation of this fundamental question because of the key role of TT8 in the regulation of both anthocyanin and PA biosynthesis in different tissues. In a previous study, it has been found that *TT8* was regulating its own expression through a positive feedback regulatory loop, which implied a TTG1-dependent regulatory mechanism (Nesi *et al.*, 2001; Baudry *et al.*, 2006). It has been also proposed that different MBW complexes involving closely related bHLH (i.e. EGL3 and GL3) could regulate *TT8* expression in the mucilage cell layer of the seed coat and in the margins of seedling cotyledons (Baudry *et al.*, 2006; Gonzalez *et al.*, 2008). Additional studies revealed that *TT8* mRNA accumulation in developing siliques was controlled by TT2 and MYB5 (Gonzalez *et al.*, 2008). Together, these data indicate that more than one MBW complex is involved in the regulation of *TT8* expression in both seeds and vegetative tissues.

With the aim of providing a comprehensive model for the complex pattern of *TT8* expression, we first carried out a functional dissection of the *TT8* promoter, revealing its

modular structure. Two modules were sufficient to drive *TT8* expression in PA- and anthocyanin-accumulating cells, whereas the third one was regulating the strength of the promoter activity. The activity of these two-first modules was assayed in different regulatory mutants affected in the formation of the MBW complexes, as well as in *tt1*, *tt16* or *ttg2*. Each of the regulatory modules was integrated the regulatory signals from these different regulators in a tissue-specific manner. We also found that *TT1*, *TT16* and *TTG2* were involved in the regulation of *TT8* expression, and that some new regulatory factors are most probably yet to be discovered. Furthermore, this approach revealed that six different MBW complexes are involved in the regulation of *TT8* promoter activity, with three being specifically dedicated to the PA-accumulating cells in seeds and the other three to the cotyledon margins where anthocyanins accumulate. Finally, specific *cis*-regulatory elements through which *TT8* expression is regulated have been identified and characterized.

2.3 MATERIALS AND METHODS

Histochemical detection of *β-Glucuronidase* (GUS) activity, fluorimetric (quantitative) GUS analyses, microscopy, *Physcomitrella patens* transfection assays and yeast one-hybrid experiments were carried out as described elsewhere (Jefferson *et al.*, 1987; Baud *et al.*, 2003; Debeaujon *et al.*, 2003; Baudry *et al.*, 2006; Thevenin *et al.*, 2012). All methods and conditions used for plant growth, plant transformation and selection for transgenic lines were as reported previously (Nesi *et al.*, 2000).

2.3.1 Plant material

The following *Arabidopsis thaliana* (L.) Heynh. mutants were used in this study: *tt2-1* (CS83), *myb5-1* (salk_030942), *tt2 myb5-1* (salk_005260, salk_030942), *pap1* (PST_16228), *pap2* (salk_093731), *tt8-3* (DEB122), *egl3-1 tt8-1*, *egl3-1 gl3-1 tt8-1*, *ttg1-1* (CS89), *tt1-1* (CS82), *tt16-1* (DXT32), *ttg2-4* (CTA18).

2.3.2 Arabidopsis gene IDs studied

MYB5, At3g13540; *PAP1*/*MYB75*, At1g56650; *PAP2*/*MYB90*, At1g66390; *TT2/MYB123*, At5g35550; *GL3*/*bHLH001*, At5g41315; *EGL3*/*bHLH002*, At1g63650; *TT8/bHLH042*, At4g09820; *TTG1*, At5g24520; *TTG2*/*WRKY44*/*DSL1*, At2g37260; *TT1*/*WIP1*, At1g34790; *TT16*/*AGL32*/*ABS*, At5g23260.

2.3.3 Constructs

All the PCR products were obtained using high-fidelity Phusion DNA polymerase with the HF buffer (Thermo Scientific-Finnzymes, Waltham, MA, USA). The primers (from Sigma-Aldrich) used in this study are described in the Supporting Information Table 2-1. Mutagenesis of *proTT8-1* and *proTT8-6* were carried out using the QuikChange® Site-Directed Mutagenesis Kit according to manufacturer instructions (Stratagene-Agilent, Massy, France). The integrity of each construct was ensured by DNA sequencing. Following Arabidopsis transformation, between 12 and 36 independent lines were assayed for GUS activity, for each construct.

proTT8:GUS fusions The *TT8* promoter fragments used for the deletion studies were amplified from the Wassilewskija Arabidopsis accession. The position of each fragment refers to the 3'-end of the previously described 'full length' *TT8* promoter (Baudry *et al.*, 2006). These DNA fragments were then introduced into the pDONR207 vector by a BP recombination (Invitrogen), and finally transferred into the binary vector pBI-101GUS containing the Gateway cassette (Figs 2-1, 2-8) or in pGWB3 (Figs 2-2b, 2-3b,c), by a LR recombination (Baudry *et al.*, 2006; Nakagawa *et al.*, 2007). The 'full length' *TT8* promoter fused to GUS was introduced into the different mutants background by crossing (Figs 2-2a, 2-3a). The 3'-end deletion fragments were fused by PCR (Table 2-1) to a modified minimal 35S promoter from cauliflower mosaic virus (from the pRTL2 vector) carrying a transcription start site, prior to the BP recombination.

proTT8:green flurescent protein (GFP) fusions The two *TT8* promoter fragments that were

Name	Sequence	Purpose
pTT8-1-B1	Attb1-GCAGAAGCAAAAGCCAAGC	5'- & 3'-end promoter deletion series
pTT8-2-B1	Attb1-TTAACCATACCCAAAATCG	5'- & 3'-end promoter deletion series
pTT8-3-B1	Attb1-GCTAGTGGCTGCTACTTGGC	5'- & 3'-end promoter deletion series
pTT8-4-B1	Attb1-ATTCTTTACTTACCGGTCAGG	5'- & 3'-end promoter deletion series
pTT8-5-B1	Attb1-TTCATCCAATAATTCCAACTTC	5'- & 3'-end promoter deletion series
pTT8-6-B1	Attb1-TTCTTTGCTCCTTGTAGCAG	5'- & 3'-end promoter deletion series
pTT8-7-B1	Attb1-CAATTAGTATATGACACGTC	5'- & 3'-end promoter deletion series
pTT8-B2	Attb2-GCTCTCTCTCTAAAAATC	5'- & 3'-end promoter deletion series
pTT8-6-35Smini-B2	Attb2-gtcctctccaaatgaaatgaacttccttatatagaggaagggtcttgcCTGCTACAAGGAGCAAAGAA	3'-end promoter deletion series
pTT8-5-35Smini-B2	Attb2-gtcctctccaaatgaaatgaacttccttatatagaggaagggtcttgcGAAGTTGGAATTATTGGATGAA	3'-end promoter deletion series
pTT8-4-35Smini-B2	Attb2-gtcctctccaaatgaaatgaacttccttatatagaggaagggtcttgcCCTGACCGGTAAGTAAAGAAT	3'-end promoter deletion series
pTT8-3-35Smini-B2	Attb2-gtcctctccaaatgaaatgaacttccttatatagaggaagggtcttgcGCCAAGTAGCAGCCACTAGC	3'-end promoter deletion series
pTT8-2-35Smini-B2	Attb2-gtcctctccaaatgaaatgaacttccttatatagaggaagggtcttgcCGATTTTGGGTATGGTTAA	3'-end promoter deletion series
pTT8-6-Protoplasts-B2	Attb2-gtcctctccaaatgaaatgaacttccttatatagaggaagggtcttgcGAATGATCCATTCACCAGACG	3'-end promoter deletion series
pTT8-6-m1-F	TCACTTGTTTGTCATACAGCTATTTTATCTAC	Mutagenesis
pTT8-6-m1-R	GTAGATAAAATAGCTGTATGACAAACAAGTGA	Mutagenesis
pTT8-6-m2-F	GTCCAATTAGTATATGACACGCCTACAAGATACATAGATC	Mutagenesis
pTT8-6-m2-R	GATCTATGTATCTTGTAGGCGTGTCATATACTAATTGGAC	Mutagenesis
pTT8-1-m3-F	CCAACGTCTGGTGAATGGATCATTCAAAAATCAAAAGTC	Mutagenesis
pTT8-1-m3-R	GACTTTTGATTTTTGAATGATCCATTCACCAGACGTTGG	Mutagenesis
pTT8-1-m4-F	CTCCTTTATTACATGGATCCATTTTCTCATTCTTTACTTACCGG	Mutagenesis
pTT8-1-m4-R	CTCCTTTATTACATGGATCCATTTTCTCATTCTTTACTTACCGG	Mutagenesis
cMYB5-B1	Attb1-TCATGATGTCATGTGGTGGG	cDNA cloning
cMYB5-B2	Attb2-CTAGTCATGTCCTAAGCTAG	cDNA cloning
cPAP1-B1	Attb1-TCATGGAGGGTTCGTCCAAAGGGCTGCGA	cDNA cloning
cPAP1-B2	Attb2-CTAATCAAATTTCACAGTCTC	cDNA cloning
cGL3-B1	Attb1-TCATGGCTACCGGACAAAACAG	cDNA cloning
cGL3-B2	Attb2-TCAACAGATCCATGCAACCC	cDNA cloning
cEGL3-B1	Attb1-TCATGGCAACCGGAGAAAACAGAACGG	cDNA cloning
cEGL3-B2	Attb2-TTAACATATCCATGCAACCCTTTGAAGTGCCTCTTG	cDNA cloning

AttB1 and AttB2 refer to the GatewayTM recombination sequences (Invitrogen).

Table 2-1. Primers used in this study

used in *P. patens* transient expression assays were recombined into the pBS TPp-B vector (Thevenin *et al.*, 2012). In order to be able to compare the results obtained for both promoters, *proTT8-6* TATA box was removed and replaced by the 35S minimal promoter, as in *proTT8-17* (Fig. 2-1d).

Coding sequence (cds) cloning and expression *MYB5*, *PAP1*, *GL3* and *EGL3* cds were amplified from cDNA that were reverse transcribed from mRNA extracted from 4-d-old seedlings grown *in vitro* on MS media supplemented with 3% sucrose. The fragments obtained were then recombined first into the pDONR207 vector and subsequently into pBS TPp-A expression vector (Thevenin *et al.*, 2012). Expression of *TT2*, *TT8* and *TTG1* in *P. patens* protoplast and yeast was carried out as described elsewhere (Baudry *et al.*, 2006; Thevenin *et al.*, 2012).

2.4 RESULTS

2.4.1 Functional dissection of TT8 promoter

In order to get better insight into how *TT8* expression is regulated, a 5'-end promoter deletion series, every *c*. 100 bp on a 1005 bp DNA fragment, was carried out (Fig. 2-1a-c). These different *TT8* promoter fragments were fused to the *uidA* (GUS, β-glucuronidase) reporter gene, and transferred into wild type (WT) plants. The first six distal promoter constructs (until *proTT8-6*, -511 bp) displayed the same GUS pattern in seeds and seedlings as the studied previously and so-called 'full length' promoter fragment (1518 bp, Fig. 1a,c) (Baudry *et al.*, 2006). By contrast, no GUS activity was detected when *proTT8-7* (-356 bp) fragment was tested (Fig. 2-1a-c). This experiment demonstrated that *proTT8-6* (-511 bp) was sufficient to specifically drive GUS activity in PA- and anthocyanin-accumulating cells in seeds and seedlings, respectively. In order to get further insights into the observed decrease in GUS staining (Fig. 2-1a), the activity of *TT8* promoter constructs was quantified in 4-d-old siliques, at the time when the *BAN* gene (a direct target of *TT8*) reaches its peak of expression (Debeaujon *et al.*, 2003; Baudry *et al.*, 2004). The activity was clearly affected in a promoter

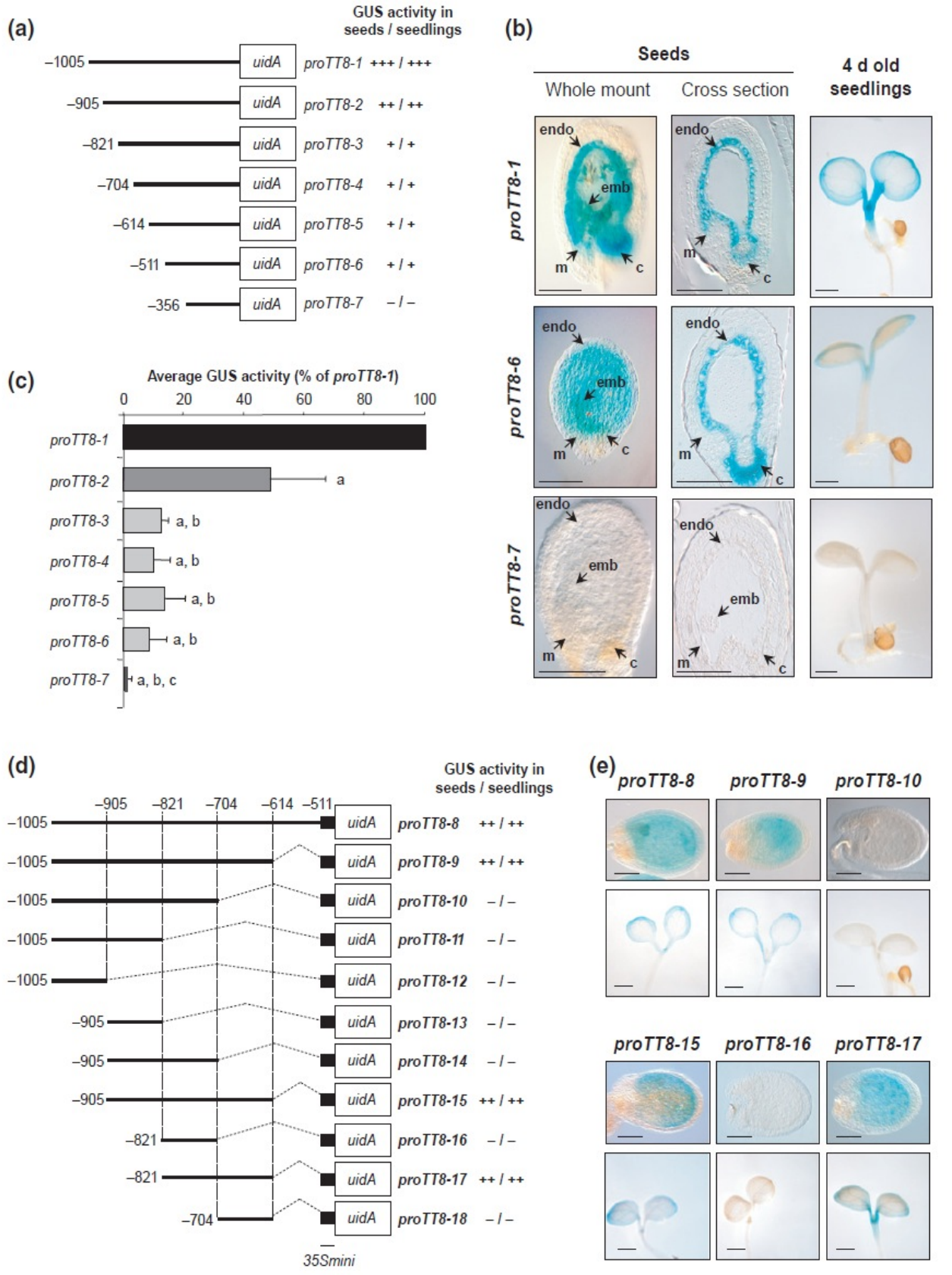


Figure 2-1. Molecular and functional dissection of the *TT8* promoter. (a) Scheme of 5'-end deletion constructs. (b) Quantitative *β-Glucuronidase* (GUS) analysis of the 5'-end deletion constructs. *t*-test significance at $P < 0.01$ when compared to (a) *proTT8-1*, (b) *proTT8-2* and (c) *proTT8-6*. Error bars +SE. (c) The GUS activity driven by the 5'-end deletion constructs in *Arabidopsis thaliana* seeds (globular stage) and seedlings. M, micropyle; c, chalaza; endo, endothelium; emb, embryo. (d) Scheme of 3'-end deletion constructs. (e) The GUS activity driven by the 3'-end deletion constructs in seeds (globular stage) and seedlings. Bars: 0,1 mm (seeds), 2 mm (seedlings). The 'full length' *TT8* promoter was used as reference for numbering. + and – refer to the visual observation of the GUS strength.

length-dependent manner (Fig. 2-1b). This experiment revealed that 50% of *proTT8-1* (-1005 bp) activity was lost with *proTT8-2* (-905 bp) and that an additional 40% of *proTT8-1* activity was lost with *proTT8-3* (-821 bp). Finally, this experiment also revealed that only 10% of *proTT8-1* (-1005 bp) activity was driven by *proTT8-6* (-511 bp). These results showed that some sequences located between -1005 and -821 bp were involved in the strength of *TT8* promoter. As a second step, a 3'-end deletion series, between -1005 and -511 bp, was carried out. This approach led to the analysis, in WT plants, of 11 additional *TT8* promoter fragments (*proTT8-8* to *proTT8-18*) placed upstream of a minimal 35S promoter (Fig. 2-1d,e). The results demonstrated that a second promoter fragment located between -821 and -614 bp (corresponding to *proTT8-17*) was sufficient to reproduce the activity of the studied previously 'full length' *TT8* promoter fragment (Baudry *et al.*, 2006a). These analyses showed that a first module (located between -1005 and -821 bp) is responsible for the strength of the *TT8* promoter, whereas two others modules (i.e. *proTT8-6* and *proTT8-17*) control its tissue specificity (Fig. 2-1).

2.4.2 Determination of the genetic relationships existing between TT8 promoter activity and the TT1/WIP1, TT16/ABS/AGL32 and TTG2/WRKY44 transcription factors (TFs)

When the characterized previously 'full length' *TT8* promoter fragment fused to GUS was introduced into *tt1*, no difference with the WT GUS pattern was observed in micropyle, chalaza or endothelium, suggesting that *TT1* was not playing a major role in the control of *TT8* promoter activity (Fig. 2-2a). Nevertheless, in a *tt1* mutant background, the activities of both *proTT8-6* (-511 bp) and *proTT8-17* (-821 to -614 bp) were restricted to the chalazal area and the distal part of the endothelium, respectively, whereas *proTT8-1* (-1005 bp) was found to be active in both tissues, but not in micropyle (Fig. 2-2b). This later result revealed a *TT1*-independent activity in the micropyle of the region located upstream the -1005 bp DNA fragment. Thus, although *TT1* is not necessary for a basal activity of the *TT8* promoter in PA-accumulating cells (Fig. 2-2c), it modulates this activity by acting on the different modules in a tissue-specific manner.

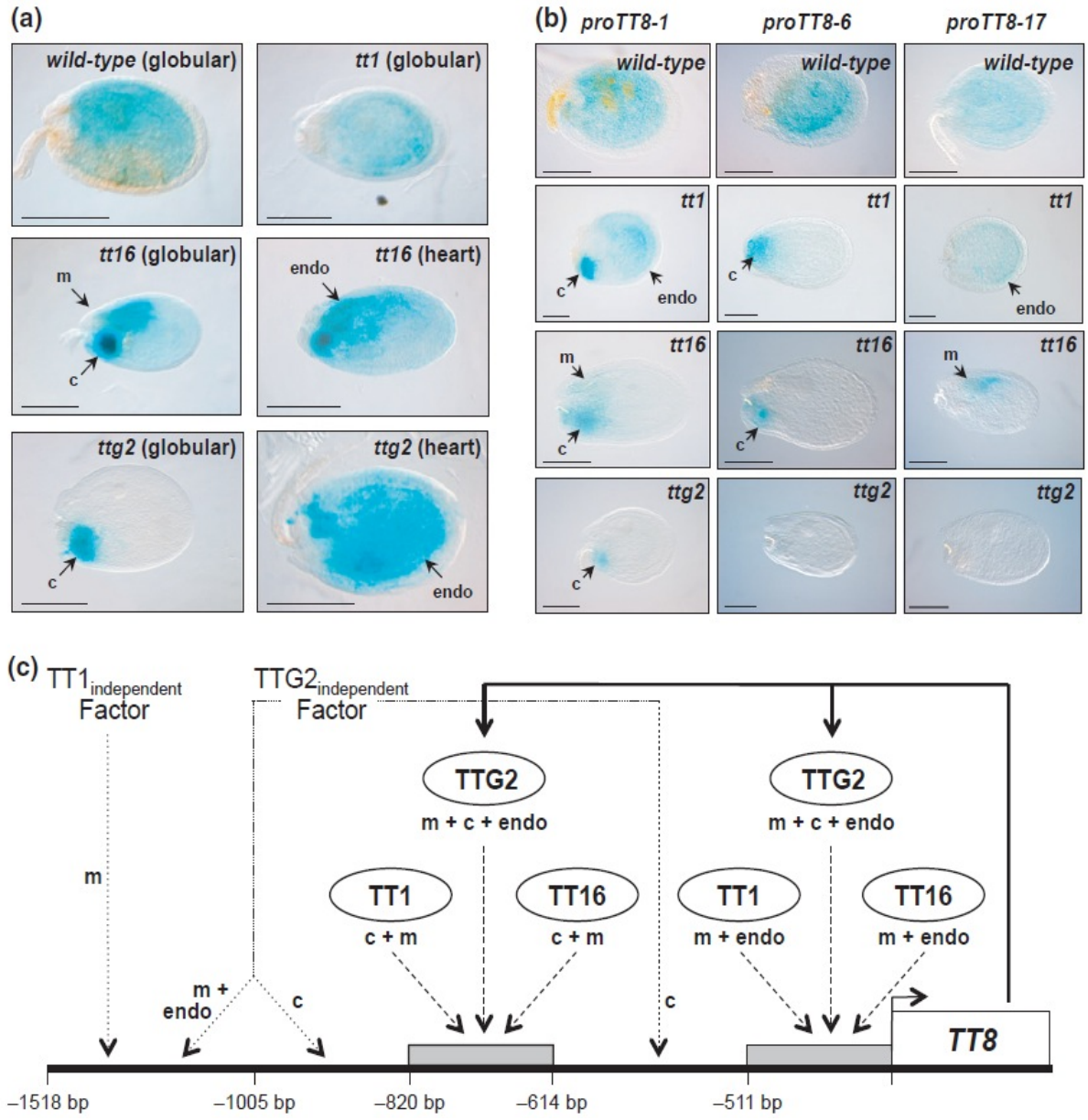


Figure 2-2. *TT8* promoter activity in *tt1*, *tt16* and *ttg2 Arabidopsis thaliana* mutant seeds. Analysis of the *β-Glucuronidase* (GUS) activity driven by the (a) 'full length' promoter fragment in seeds or (b) *proTT8-1* in comparison with, *proTT8-6* and *proTT8-17* (globular stage), in the *tt1*, *tt16* and *ttg2* backgrounds. Bars: 0.1 mm. M, micropyle; c, chalaza; endo, endothelium. (c) Schematic regulation of *TT8* promoter activity by *TT1*, *TT16* and *TTG2*. Solid arrows, link previously demonstrated; dashed arrows, regulations identified in this study; dotted arrows, putative regulations. The 'full length' *TT8* promoter was used as reference for numbering.

In a *tt16* mutant background, the activity of the 'full length' promoter was essentially restricted to the micropyle and the chalazal areas, although some weak activity was also observed in the endothelium in later developmental stages (from heart stage onward, Fig. 2-2a). This GUS pattern perfectly matches the pattern of PA accumulation observed in this mutant (Nesi *et al.*, 2002). Similarly, *proTT8-1* (-1005 bp) activity was detected in the same two tissues, whereas the activity of *proTT8-6* (-511 bp) and *proTT8-17* (-821 to -614 bp) were restricted to the chalaza and the micropyle, respectively (Fig. 2-2b). These data indicated that *TT16* was necessary for the activity of *TT8* promoter in the endothelium, and that *TT16* signalling was integrated through the *proTT8-6* and *proTT8-17* modules (Fig. 2-2c).

In the *ttg2* mutant the activity of the 'full length' *TT8* promoter was restricted to the chalaza area at early stage of seed development (globular) and was detected in all the PA-accumulating cells in the later stages (from heart stage onward; Fig. 2-2a). Consistent with this result, *proTT8-1* (-1005 bp) activity is observed in the chalaza strand (Fig. 2-2b). When *proTT8-6* (-511 bp) and *proTT8-17* (-821 to -614 bp) were assayed in *ttg2*, no blue staining was detected (Fig. 2-2b). Together, these data showed that the activity of the *TT8* promoter in the endothelium was strictly dependent on *TTG2* early during seed development, but this was not the case later during seed maturation or in the chalaza. Moreover, the detection of GUS activity in chalaza with *proTT8-1* (-1005 bp) suggested that either *TTG2* regulation in this tissue was independent of *proTT8-6* and *proTT8-17* or that a threshold effect was at play (Fig. 2-2c).

2.4.3 Regulation of TT8 promoter activity by the MYB-bHLH-WDR (MBW) protein complexes

It has been shown previously that TT2 controls the specific expression of *TT8* in PA-accumulating cells through its interaction with TTG1 and TT8 itself (Baudry *et al.*, 2006; Lepiniec *et al.*, 2006). Therefore, the activity of *TT8* promoter was assayed in *tt2* (as well as in *tt8* and *ttg1* as controls). This experiment confirmed that the activity of *TT8* promoter was restricted to the chalaza in both *tt8* and *ttg1* mutants (Fig. 2-3a). In *tt2* seeds, although no

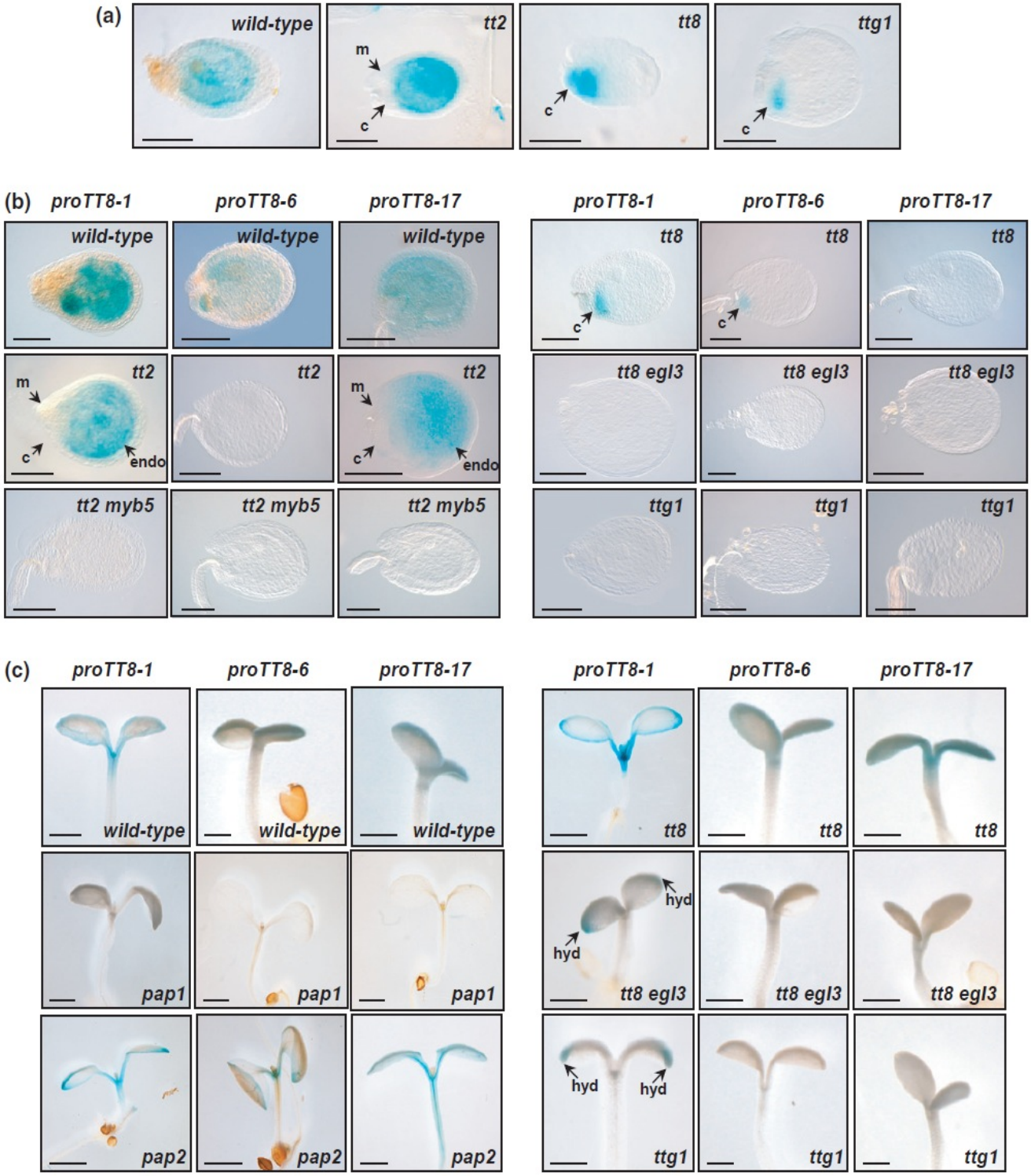


Figure 2-3. *TT8* promoter activity in R2R3-MYB (*tt2*, *myb5*, *pap1* and *pap2*), R/B-like bHLH (*tt8*, *gl3* and *egl3*) and *ttg1* *Arabidopsis thaliana* mutants. The activity of the (a) 'full length' promoter fragment was studied in seeds (globular stage), whereas *proTT8-1*, *proTT8-6* and *proTT8-17* activities were studied in (b) seeds (globular stage) and (c) seedlings. m, micropyle; c, chalaza; endo, endothelium; hyd, hydathodes. Bars: 0.1 mm (seeds), 2 mm (seedlings).

activity was detected in micropyle and chalazal areas, it was apparently unaffected in the endothelium. These results suggest that putative redundant MBW complexes regulate *TT8* promoter in a partially overlapping manner in seeds (Fig. 2-3a). In order to determine which MBW complexes could be involved, the promoter activity of *proTT8-1* (-1005 bp), *proTT8-6* (-511 bp) and *proTT8-17* (-821 to -614 bp) were assayed in different mutants backgrounds (Fig. 2-3b). *proTT8-1* (-511 bp) and *proTT8-17* (-821 to -614 bp) conferred GUS activity in the endothelium when expressed in *tt2*, whereas no more activity was detected in *myb5 tt2* double mutant. The mutation of *tt2* was sufficient to block the activity of *proTT8-6* (-511 bp). In *tt8*, some blue staining was observed in the chalaza when *proTT8-1* (-1005 bp) and *proTT8*-6 (-511 bp) were assayed, which disappeared in the *tt8 egl3* double mutant. Conversely, *proTT8-17* (-821 to -614 bp) was inactive in *tt8*. Finally, none of the three promoters was active in *ttg1*.

In 4-d-old seedlings the activities of these promoters were found to be dependent on *PAP1* (and independent of *PAP2*; Fig. 2-3c). In addition, *proTT8-6* (-511 bp) and *proTT8-17* (-821 to -614 bp) were regulated by *TT8*, *EGL3* and *TTG1*. Interestingly, *proTT8-1* was still active in the hydathodes of the *tt8 egl3* double mutant, but not in the *tt8 egl3 gl3* triple mutant, suggesting that a threshold effect involving GL3 was at play in this tissue (Fig. 2-3c, 2-4). In *ttg1*, *proTT8-1* (-1005 bp) was also active in the hydathodes, suggesting that either some *TTG1*-independent regulators may be involved in the regulation of this promoter in these specific cells, or that some activities could be driven by the R2R3-MYB and R/B-like bHLH factors independently of TTG1. This second hypothesis implies a threshold effect as no GUS activity was detected in *ttg1* seedling carrying either the *proTT8-6* (-511 bp) or the *proTT8-17* (-821 to -614 bp) construct (Fig. 2-3c). Together, these data exemplified how different MBW complexes regulate *TT8* activity in both seeds and cotyledons.

In order to provide molecular support to the genetic analyses the different putative MBW complexes were assayed in transient activation assays in *P. patens* protoplasts using *proTT8-6* (-511 bp) and *proTT8-17* (-821 to -614 bp) fused to the *GFP* reporter gene (Fig. 2-5a). *P. patens* protoplasts were used as this system was previously described as a fast, sensitive and robust method to carry out transient expression assays (Thevenin *et al*., 2012). This approach allows quantitative analysis of gene expression using up to four different

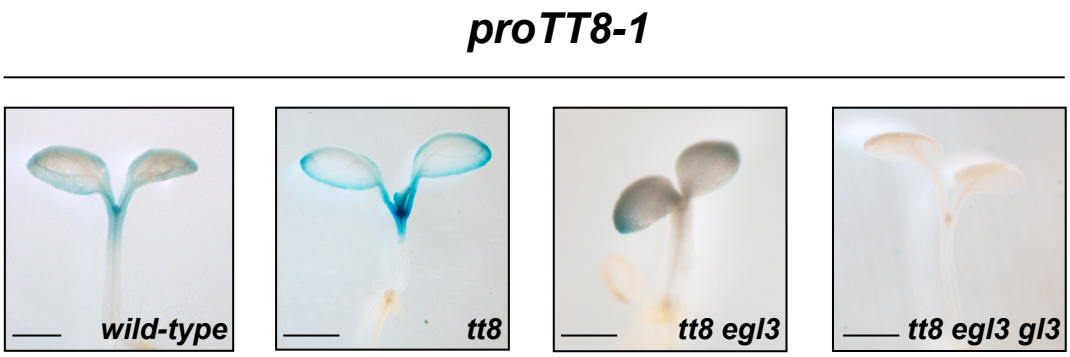


Figure 2-4. Redundant R/B-like bHLH activities involved in the regulation of *proTT8-1* activity in cotyledons.

regulatory proteins in one experiment. In our assays, three R2R3-MYBs (i.e. TT2, MYB5 and PAP1) were tested in combination with three R/B-like bHLHs (i.e. TT8, EGL3 and GL3) and TTG1. For both modules the strongest activation was obtained with TT2-TT8-TTG1 and TT2-EGL3-TTG1 complexes. The other seven possible complexes were able to activate significantly both promoters, but to a much lower extent. Interestingly, R2R3-MYBs together with the R/B-like bHLH TFs were able to slightly activate both modules, independently of TTG1 (Fig. 2-6). This observation is consistent with previous results obtained using *BAN* promoter in a yeast one-hybrid experiment and in transient activation assays in *P. patens* protoplasts (Baudry *et al.*, 2004; Thevenin *et al.*, 2012).

The data demonstrate that diverse MBW complexes have the ability to regulate *TT8* promoter activity: three MBW complexes are specific to PA-accumulating cells in seeds, namely TT2-TT8-TTG1, TT2-EGL3-TTG1 and MYB5-TT8-TTG1, whereas PAP1-TT8-TTG1, PAP1-EGL3-TTG1 and PAP1-GL3-TTG1 are specific to the cotyledon margins (Fig. 2-5b,c). Interestingly, these data also indicate that *TT8* expression in the chalaza and cotyledon hydathodes might also involve some *TTG1*-independent regulatory mechanisms (Fig. 2-5b,c).

2.4.4 Identification of the cis-regulatory targets of the MBW complexes

We selected proTT8-6 (-511 bp) as a model to be studied in depth in order to determine through which cis-regulatory elements the MBW complexes could regulate TT8 expression. As a starting point, *proTT8-6* sequences were amplified in five additional accessions (Bay-0, Shahdara, L*er*, Cvi-0 and Col-0), aligned and analysed using the PLACE web tool (Fig. 2-7; http://www.dna.affrc.go.jp/PLACE/, Higo *et al.*, 1999) with the aim of identifying conserved putative MYB and bHLH binding sites. This analysis led to the identification of two putative binding sites (MYB-core binding site, *CAGTTA*, and G-box like *CACGTC*) that display strong similarities with the *cis*-regulatory elements of the *BAN* promoter shown to be the targets of the TT2-TT8-TTG1 complex (Thevenin *et al.*, 2012).

The MYB-core and G-box sites were mutagenized and replaced by the *CATACA* (*proTT8-6-m1*) and *CACGCC* (*proTT8-6-m2*) sequences, respectively. The mutated

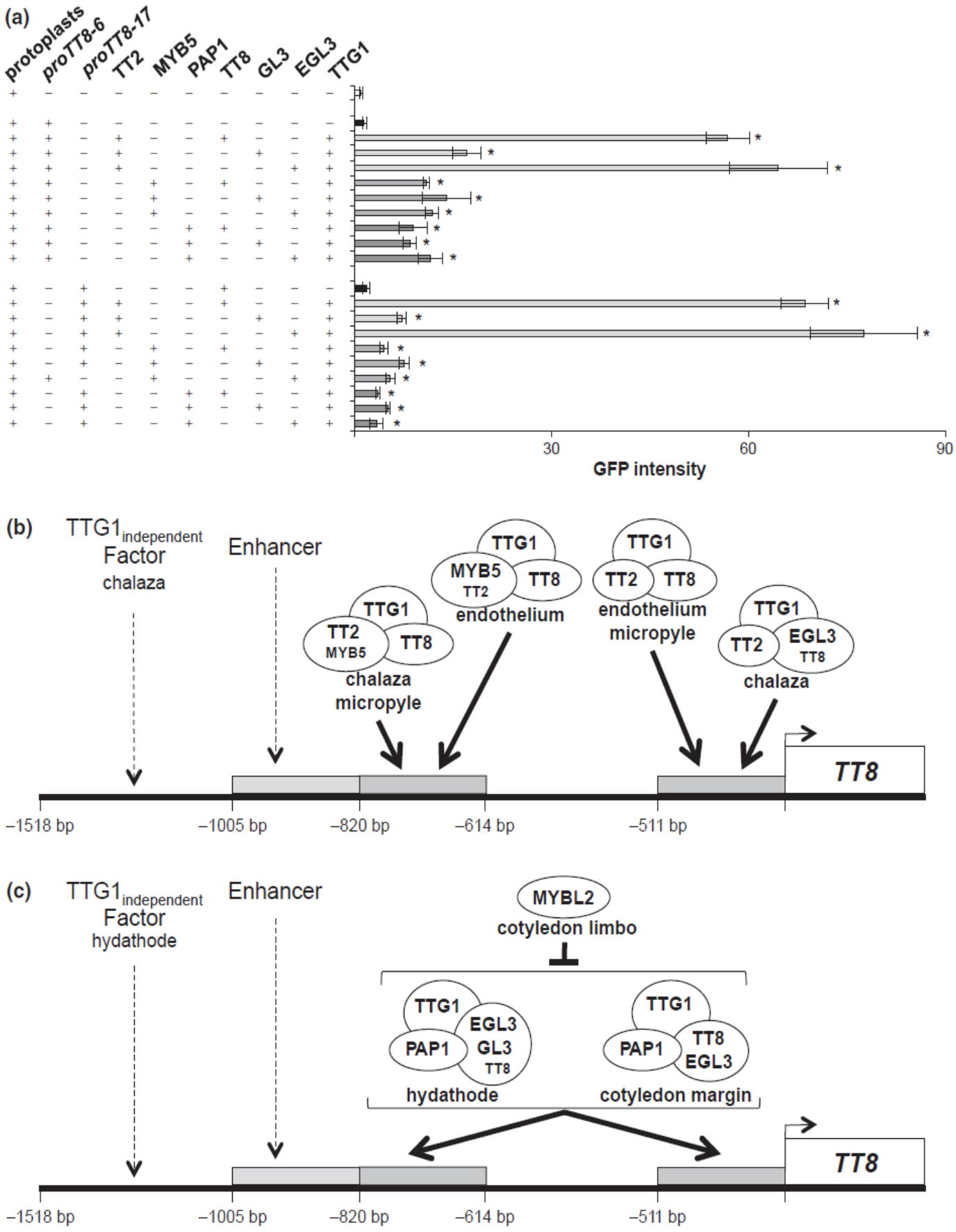


Figure 2-5. Regulation of *TT8* expression by the MBW complexes. (a) Green fluorescent protein (GFP) intensity (arbitrary units) measured in *Physcomitrella patens* protoplasts co-transfected with R2R3-MYB (*TT2*, *MYB5* or *PAP1*), R/B-like bHLH (*TT8*, *GL3* and *EGL3*), *TTG1* and either *proTT8-6* or *proTT8-17*. Error bars XXSE. *t*-test significance: *, $P < 0.01$. Schematic representation of *TT8* promoter regulation by the MBW complexes in (b) seeds and (c) cotyledons. Solid arrows, interaction demonstrated in this study; dashed arrows, putative regulatory mechanism. The 'full length' *TT8* promoter was used as reference for numbering.

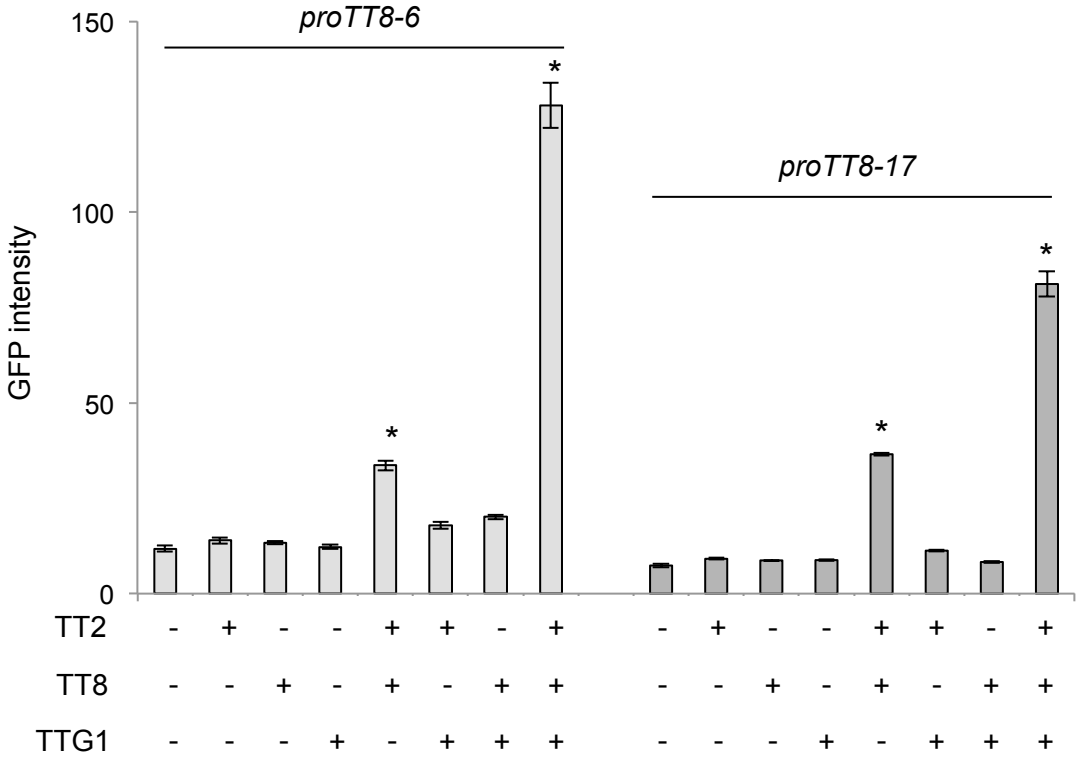


Figure 2-6. R2R3-MYB together with R/B-like bHLH, in the absence of TTG1, can trigger *proTT8-6:GFP* and *proTT8-17:GFP* activity in *P. patens* protoplasts. Error bars show SE. *t*-test significance: *, $P < 0.01$.

promoters were assayed in yeast one-hybrid experiments, as described in (Baudry *et al.*, 2006) and in transient expression assays in *P. patens* protoplasts (Fig. 2-8a,b). These experiments confirmed that TT2 and TT8 regulate *proTT8-6* activity through these two putative *cis*-regulatory elements. The mutated promoters were subsequently tested *in planta* (Fig. 2-8c). Both promoters lack activity in seed endothelium and cotyledon margins, their activity being restricted to chalaza and cotyledon hydathodes. These data demonstrated that these two *cis*-regulatory elements played an important role in the regulation of *TT8* promoter activity, at least in the endothelium and cotyledons margins. However, no similar binding sites were identified in *proTT8-17*, suggesting that MBW complexes may also recognise different type of regulatory sequences (Fig. 2-9).

2.4.5 Identification of AC-rich regulatory sequences involved in TT8 promoter activity

The implication of *AC*-rich sequences (*AC*-element, AC) in the regulation of structural genes involved in secondary metabolism has been extensively reported (Patzlaff *et al.*, 2003; Prouse & Campbell, 2012). Two palindromic motifs, which differed by only one nucleotide and are similar to the well described ACII MYB binding site (*ACCAACC*; Patzlaff *et al.*, 2003), were identified in *proTT8-1* (-1005 bp) - the first motif in *proTT8-6* (*ACCAACCA*; -511 bp, Fig. 2-7) and the second motif in *proTT8-17* (*ACCAAACCA*; -821 bp to -614 bp, Fig. 2-9).

In order to determine if these two putative *cis*-regulatory sequences are involved in the regulation of *TT8* promoter activity, they were mutated alone or in combination in the *proTT8-1* context (Fig. 2-8d). The impact of these mutations was then assessed through the analysis of GUS activity in seeds and cotyledons of WT plants (Fig. 2-8e). The mutation of the *proTT8-6* (-511 bp) element into *ATGGATCA* (*proTT8-1-m3*) did not affect activity in seeds, whereas in cotyledons the blue staining was only observed in the hydathodes. Conversely, mutation of the *proTT8-17* (-821 bp to -614 bp) element into *AATGGATCA* (*proTT8-1-m4*) did not modify GUS activity. However, when both mutations were combined (*proTT8-1-m3m4*), no more blue staining was detected in either seeds or cotyledons (Fig. 2-8d,e). This later result suggests that these two *AC*-rich sequences have partly overlapping

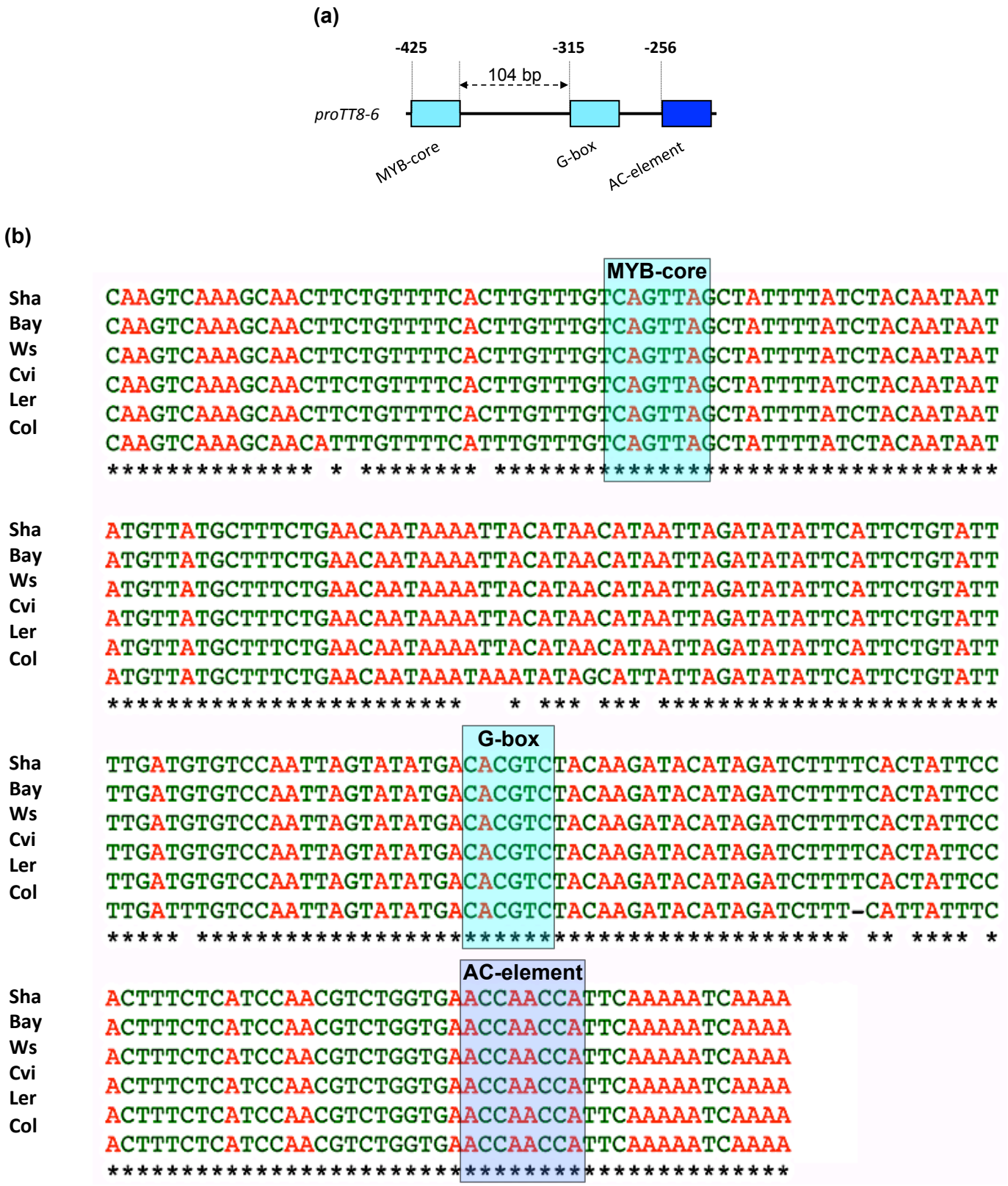


Figure 2-7. *proTT8-6* key *cis*-regulatory elements in six different Arabidopsis accessions. (a) Structure, (b) sequence (from -459 bp to -235 bp).

function in the regulation of *proTT8-1* activity. To test this hypothesis, the *m3* and *m4* mutations were introduced into *proTT8-6* and *proTT8-17*, respectively, and assayed *in planta*. This experiment showed that GUS activity was totally lost in seeds and seedlings of WT plants expressing these constructs, supporting the important role played by these *AC*-elements in the regulation of *TT8* promoter activity (Fig. 2-8f).

2.5 DISCUSSION

2.5.1 Modular structure of the TT8 promoter

Previous studies aimed at understanding the role and the mode of action of *TT8* in PA accumulation during *Arabidopsis* seed development reveal that *TT8* regulates PA biosynthesis through its involvement in a MYB-bHLH-WDR (MBW) ternary protein complex. In this MBW complex TT8 has been shown to control the expression of both the so-called late biosynthetic genes (which encode the enzyme specifically committed to the biosynthesis of PAs and anthocyanins) and its own expression (Nesi *et al.*, 2000; Debeaujon *et al.*, 2003; Baudry *et al.*, 2004; Baudry *et al.*, 2006). A 'full length' promoter sufficient to drive reporter gene expression in the PA-accumulating cells (chalaza, micropyle and endothelium) in seeds and in the cotyledon margins (including the hydathodes) in which anthocyanins accumulate in young seedlings, has been isolated (Baudry *et al.*, 2006). Analysis of the *TT8* promoter revealed that a 1 kb fragment (*proTT8-1*) was sufficient to drive gene expression in a similar way to the previously isolated 'full length' promoter. *proTT8-1* is composed of at least three regulatory modules: two specifically driving gene expression in PA- and anthocyanin-accumulating cells (*proTT8-6*, -511 bp and *proTT8-17*, between -821 and -614 bp) and one influencing the overall promoter strength (Fig. 2-1).

2.5.2 The activity of the TT8 promoter is strongly connected to the seed coat differentiation

In order to explore how *TT8* promoter activity is influenced by the developmental stage of the Arabidopsis seed coat, the role of three key genes, namely *TT1* (Zn finger), *TT16* (MADS

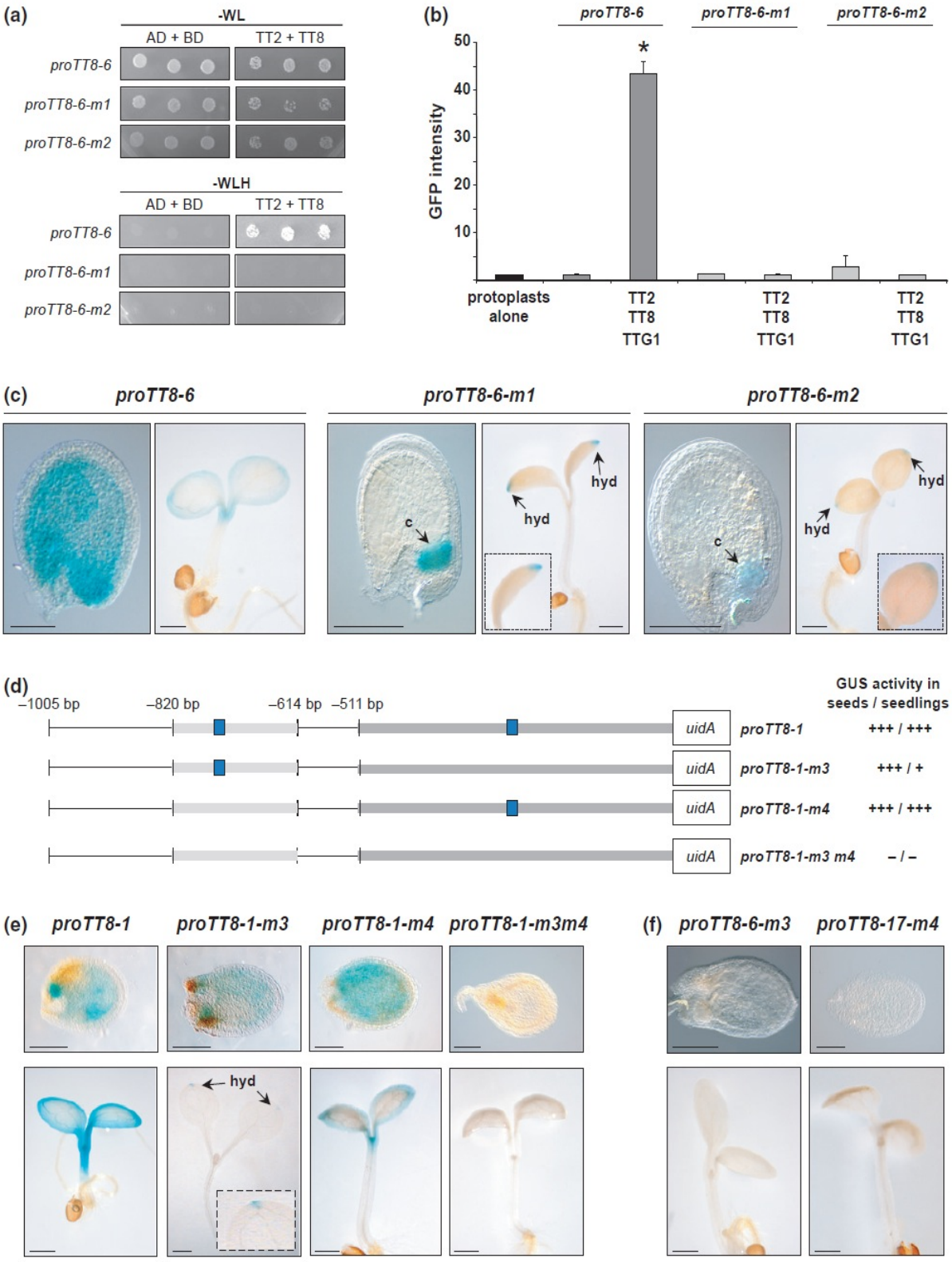


Figure 2-8. Identification and characterization of *cis*-regulatory elements involved in the regulation of *TT8* promoter activity. (a) Yeast one-hybrid experiment. Yeast were stably transformed with wild-type (WT) or mutated *proTT8-6:HIS3* constructs. These different yeast strains were then co-transfected with either the empty vectors (AD + BD) or with the same vectors containing TT2 and TT8 (TT2 + TT8). Upper panel, growth on control media deprived of W and L amino acids. Lower panel, growth on selective media deprived of W, L and H amino acids. AD, GAL4 activation domain; BD, GAL4 binding domain. (b) Green fluorescent protein (GFP) intensity measured in *Physcomitrella patens* protoplasts co-transfected with TT2, TT8, TTG1 and the WT or the mutated versions of *proTT8-6:GFP*. Error bars XXSE. *t*-test significance: *, $P < 0.01$. (c) WT or mutated *proTT8-6:GUS* activity in WT seeds (globular stage) and seedlings. (d) Schematic representation of the position of the analysed *AC*-element on the WT and the mutated *proTT8-1*. The 'full length' *TT8* promoter was used as reference for numbering. + and – refer to the visual observation of the β-Glucuronidase (GUS) strength. (e) WT or mutated *proTT8-1:GUS* activity in WT seeds (globular stage) and seedlings. (f) Impact of the *AC*-element mutations on *proTT8-6:GUS* and *proTT8-17:GUS* activities in WT seeds and seedlings. c, chalaza; hyd, hydathode. Bars: 0.1 mm (seeds), 2 mm (seedlings).

box) and *TTG2* (WRKY), involved in this process was studied (Johnson *et al.*, 2002; Nesi *et al.*, 2002; Sagasser *et al.*, 2002). The activity of the *TT8* promoter was found to be regulated by these three TFs, and the specific pattern driven by *proTT8-1* (-1005 bp) and the 'full length' promoter, corresponded essentially to the sum of the activities driven by the *proTT8-6* (-511 bp) and *proTT8-17* (-821 to -614 bp) regulatory modules (Fig. 2-2). The fact that *proTT8-1* was active in the chalaza of all three mutants, indicated that another regulation was at play in this tissue. It could be hypothesized that any of these three TF is sufficient to activate *proTT8-1* in this tissue, or the involvement of another regulatory mechanism. Interestingly, this work also revealed an unpredicted positive feedback regulatory loop between *TT8* and *TTG2*, and an unexpected regulatory link between *TT1* and *TT8* expression (Johnson *et al.*, 2002; Ishida *et al.*, 2007; Gonzalez *et al.*, 2009; Albert *et al.*, 2011; Appelhagen *et al.*, 2011). Together, these data indicate that *proTT8-6* and *proTT8*-17 differentially integrate the developmental signals from TT1, TT16 and TTG2, and that perhaps some unknown regulators are likely to be discovered.

2.5.3 Several MBW protein complexes control the spatio-temporal expression of TT8

Various studies have indicated that different sets of MBW complexes regulate the biosynthesis of PAs and anthocyanins, by modulating the expression of the structural genes specifically dedicated to these biosynthetic processes. From these studies, redundant TTG1-dependent activities between proteins belonging to the same TF subgroups were identified. For example, the Arabidopsis R2R3-MYB *PAP1* and *PAP2* have been shown to regulate anthocyanin biosynthesis in the vegetative tissues, whereas *TT2* and *MYB5* have been proposed to regulate PA biosynthesis in seeds (Baudry *et al.*, 2006; Gonzalez *et al.*, 2008; Gonzalez *et al.*, 2009; Appelhagen *et al.*, 2011). Similarly, two additional Arabidopsis R/B-like bHLH factors, namely *EGL3* and *GL3*, have been also shown to regulate these two branches of the flavonoid pathway (Zhang *et al.*, 2003; Baudry *et al.*, 2006; Feyissa *et al.*, 2009). Interestingly, *TT8* expressiom was found to be directly regulated by TT8 itself through a positive feedback regulatory loop involving redundant MBW complexes (Nesi *et al.*, 2001;

Baudry *et al.*, 2006). We took advantage of the characterization of the different modules involved in the regulation of *TT8* promoter activity to further explore these mechanisms. This approach revealed that the control of *TT8* expression much more complex than hypothesized previously, as it involves the combined action of at least six functionally redundant MBW complexes interacting with different modules in a tissue-dependent manner (Fig. 2-3). This level of complexity is particularly well exemplified for the PA-accumulating cells where the sole TT2-TT8-TTG1 complex was thought to regulate *TT8* expression, and for which two additional complexes have been identified (Baudry *et al.*, 2006). It would be interesting to investigate if this level of complexity also occurs in the other aspects of plant growth and development that are controlled by MBW complexes.

2.5.4 TT8 promoter activity is orchestrated by a highly diverse set of cis-regulatory DNA sequences

In-depth analysis of *proTT8-6* (-511 bp) resulted in the identification of a functional MYB binding site similar to the one characterized in *BAN* (which encodes the first enzyme of the flavonoid pathway specifically dedicated to PA biosynthesis) promoter, and closely related to the canonical MYB binding site ($C[^{C}/_{A}]GTT[^{G}/_{A}]$; Weston, 1992; (Weston, 1992; Thevenin *et al.*, 2012). Equivalent MYB binding sites were identified on the promoter of some MBW complexes target genes in maize or Arabidopsis (Roth *et al.*, 1991 ; Koshino-Kimura *et al.*, 2005; Ryu *et al.*, 2005; Song *et al.*, 2011). By contrast, *proTT8-17* did not display any canonical MYB binding site, suggesting that divergent MYB target DNA sequences recognized by the MBW complexes remain to be identified, as has been described for other promoters (Figs 2-9, 2-10; Koshino-Kimura *et al.*, 2005; Ryu *et al.*, 2005; Song *et al.*, 2011). In addition, two AC-rich MYB binding sites (also called AC-elements) were also found to be functionally redundant and necessary for the tissue-specific activity of the *TT8* promoter. Similar DNA motifs have been shown in various plant species to play a central role in the regulation of the promoter activity of some structural genes involved in secondary metabolism (Patzlaff *et al.*, 2003; Prouse & Campbell, 2012). These results suggest that another R2R3-MYB protein could participate in the regulation of *TT8* promoter activity

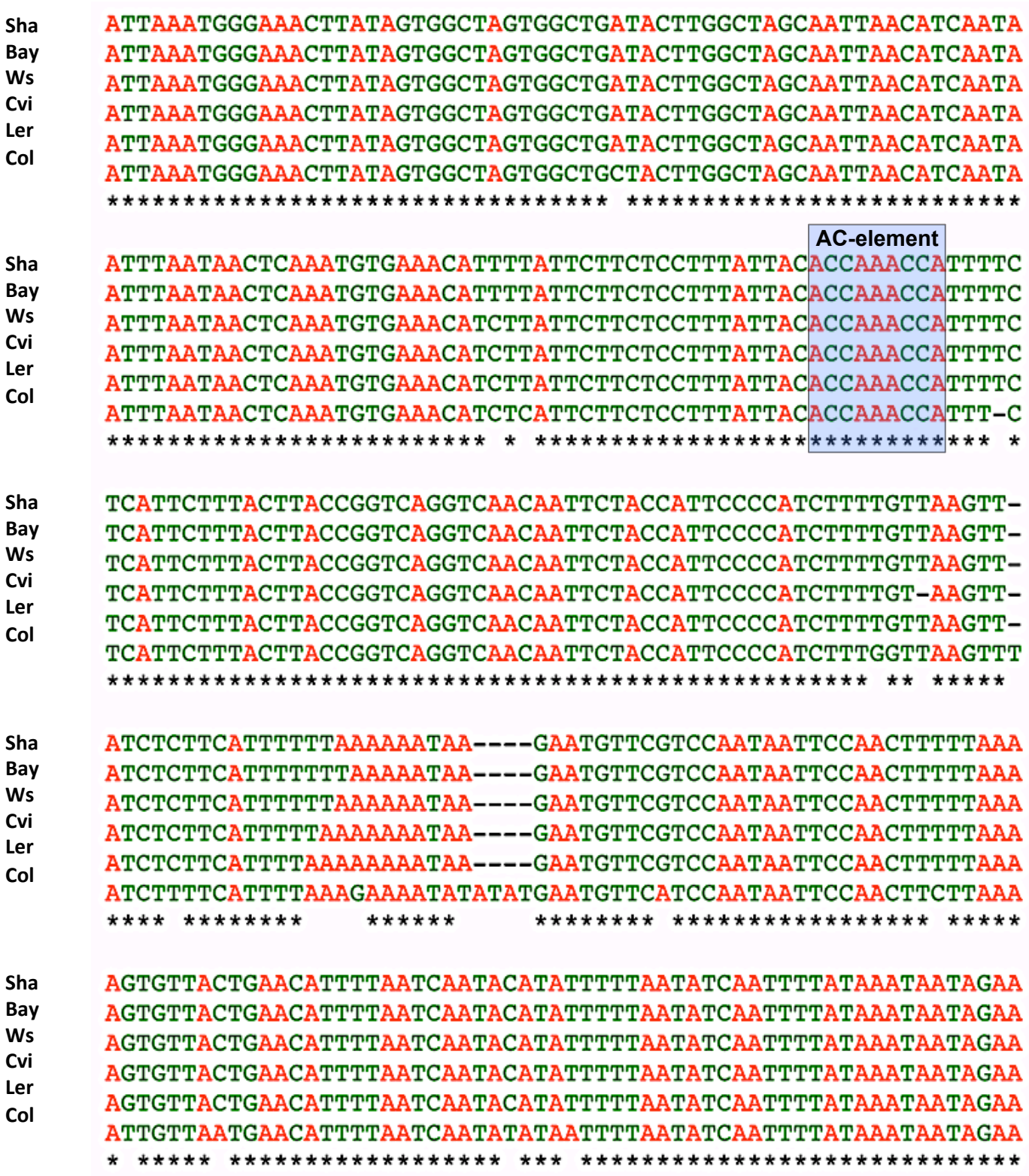


Figure 2-9. *proTT8-17* sequence in six different Arabidopsis accessions.

independently of any R/B-like bHLH TF, as has been shown for the maize P protein (Grotewold *et al.*, 1994). However, it cannot be excluded that the MBW complexes, or another type of TF, may regulate *TT8* promoter activity through these *AC*-rich sequences. On the other hand, TT8, GL3 and EGL3 were predicted, like other plant bHLH proteins to bind to the palindromic G-box DNA sequences (Heim *et al.*, 2003; Toledo-Ortiz *et al.*, 2003; Thevenin *et al.*, 2012). Interestingly, a G-box like sequence has been found to be necessary for the regulation of *proTT8-6* (-511 bp) activity, suggesting that the consensus sequence to which TT8, EGL3 and GL3 bind could be *CACGT[$^{G}/_{C}$]*, in which the *ACGT* core sequence is conserved (Figs 2-7, 2-8). This observation is consistent with some results obtained with *Caenorhabditis elegans*, suggesting that bHLH proteins may recognise divergent DNA motifs, depending on their interacting protein partners (Grove *et al.*, 2009). This level of complexity in the bHLH binding capacity may explain why no canonical bHLH binding site (i.e. E-box or *ACGT* core sequence) was found in *proTT8-17* (-821 bp to -614 bp). These results point out that some important efforts remain to be done in order to determine the entire diversity of *cis*-regulatory sequences recognised by the bHLH proteins in plants, which would be of great interest in order to better understand how MBW complexes regulate their target genes. High throughput genomic approaches (e.g. ChIP-chip or ChIP-seq) would be helpful to build a comprehensive view of this regulation, although it seems that the relevance of each interaction found has to be confirmed *in vivo* (Morohashi & Grotewold, 2009).

2.5.5 Modularity and complexity of spatio-temporal regulation of promoter activity

In this study we found that the regulation of TT8 promoter activity, and consequently TT8 expression, requires different regulatory modules in the promoter, several partially redundant proteins complexes and various cis-regulatory DNA sequences to which TF bind. The partially redundant DNA regulatory modules and proteins complexes identified, probably reflect the underlying evolutionary processes. It would be interesting to investigate similar regulatory networks in distantly related plant species in order to understand the structure of the ancestral network and its evolution. Overall, the surprising complexity of this regulation

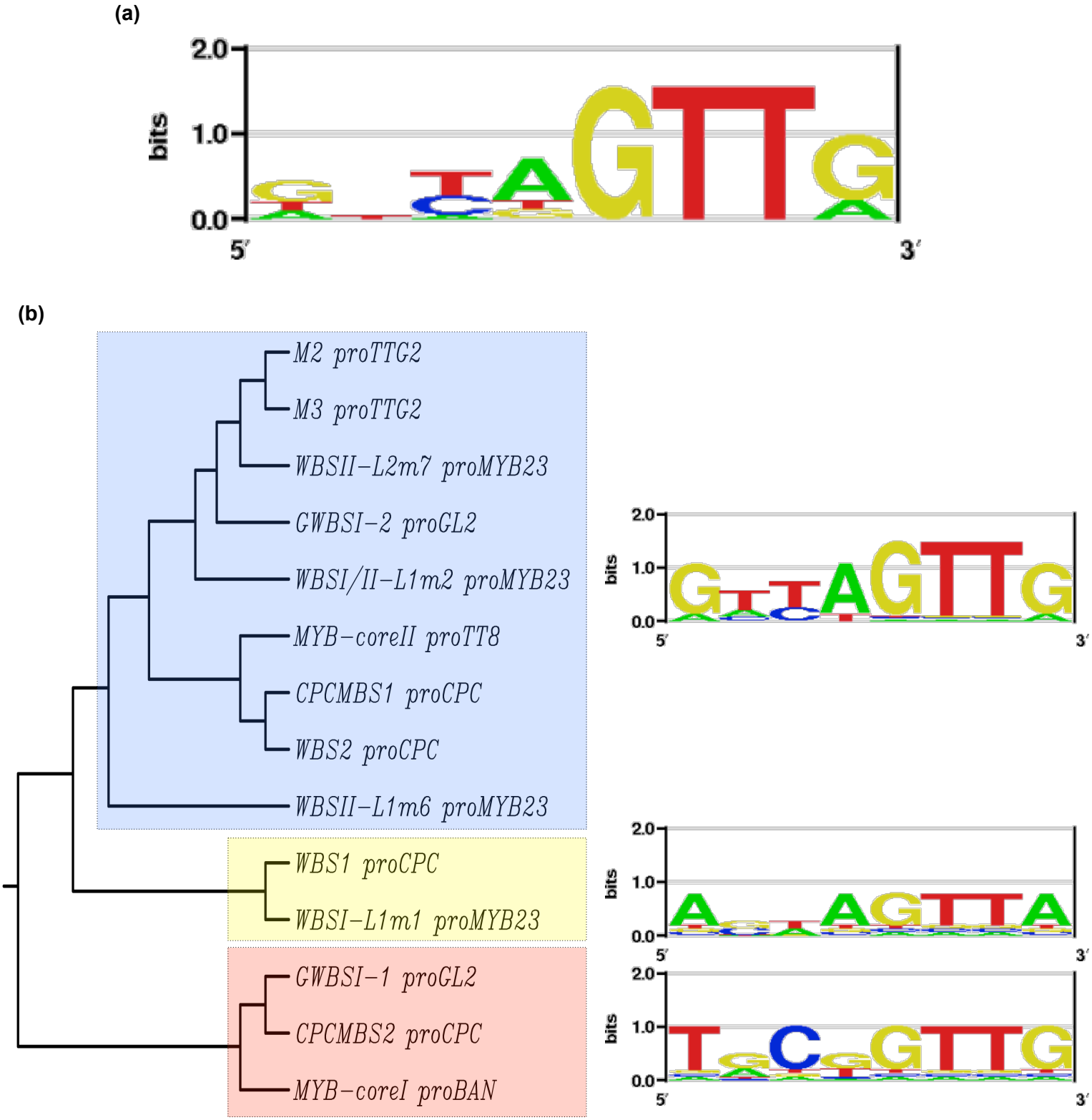


Figure 2-10. *cis*-regulatory element recognised by the Arabidopsis R2R3-MYB involved in MBW complexes. (a) Consensus sequence obtained using the enoLOGOS web tool. (b) These regulatory element can be grouped, accordingly to their DNA sequence, in three classes. Left panel: Rooted distance tree, right panel: consensus sequences (Ishida *et al*., 2007, Kang *et al*., 2009, Koshino-Kimura *et al*., 2005, Ryu *et al*., 2005, Song *et al*., 2011).

indicates that the study of genes expression through the regulation of their promoter activity cannot be reduced to the search for, and the analysis of either a minimal or a full-length promoter if the aim is to fully understand the intricate relations that exist, at the spatio-temporal levels, between the various TFs involved and their target DNA sequences.

2.6 ACKNOWLEDGEMENTS

We thank Dr Bertrand Dubreucq for useful discussions, and Johanne Thévenin, Louise Vaisman, Cecile Brousse and Sophie Bobet for their excellent technical support, as well as the 'Plateforme de cytologie et imagerie végétale (PCIV)'. The *egl3 tt8*, *egl3 gl3 tt8* and *tt2 myb5* mutant seeds were kindly provided by Dr Alan Lloyd, and the *pap1* and *pap2* seeds by Prof Bernd Weisshaar and Dr Ralph Stracke. This research was carried out as part of the EU-funded project FLAVO (Food CT-2004-513960). W.X.'s work was generously supported by the China Scholarship Council (CSC).

CHAPTER III

COMPREHENSIVE ANALYSES OF MYB–BHLH–WDR COMPLEXES AND THEIR TARGETS IN ARABIDOPSIS SEED

Article 2: Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB–bHLH–WDR complexes and their targets in Arabidopsis seed

Wenjia Xu[1, 2], Damaris Grain[1, 2], Sophie Bobet[1, 2], José Le Gourrierec[1, 2], Johanne Thévenin[1, 2], Zsolt Kelemen[1, 2], Loïc Lepiniec[1, 2], Christian Dubos[1, 2]

Accepted for the publication in New Phytologist (2014, vol. 202)

[1] INRA, Institut Jean-Pierre Bourgin, Saclay Plant Sciences, RD10, F-78026 Versailles, France

[2] AgroParisTech, Institut Jean-Pierre Bourgin, Saclay Plant Sciences, RD10, F-78026 Versailles, France

Author for correspondence:

Christian Dubos

Tel: +33 499612613

Email: christian.dubos@supagro.inra.fr

3 Article 2

3.1 SUMMARY

- In *Arabidopsis thaliana*, proanthocyanidins (PAs) accumulate in the innermost cell layer of the seed coat (i.e. endothelium, chalaza and micropyle). The expression of the biosynthetic genes involved relies on the transcriptional activity of R2R3-MYB and basic helix-loop-helix (bHLH) proteins, which form ternary complexes ('MBW') with TRANSPARENT TESTA GLABRA1 (WD repeat protein). The identification of the direct targets and the determination of the nature and the spatio-temporal activity of these MBW complexes are essential steps towards a comprehensive understanding of the transcriptional mechanisms that control flavonoid biosynthesis.
- In this study, various molecular, genetic and biochemical approaches were used.
- Here, we have demonstrated that, of the 12 studied genes of the pathway, only *dihydroflavonol-4-reductase* (*DFR*), *leucoanthocyanidin dioxygenase* (*LDOX*), *BANYULS* (*BAN*), *TRANSPARENT TESTA 19* (*TT19*), *TT12* and H^+*-ATPase isoform 10* (*AHA10*) are direct targets of the MBW complexes. Interestingly, although the TT2-TT8-TTG1 complex plays the major role in developing seeds, three additional MBW complexes (i.e. MYB5-TT8-TTG1, TT2-EGL3-TTG1 and TT2-GL3-TTG1) were also shown to be involved, in a tissue-specific manner. Finally, a minimal promoter was identified for each of the target genes of the MBW complexes.
- Altogether, by answering fundamental questions and by demonstrating or invalidating previously made hypotheses, this study provides a new and comprehensive view of the transcriptional regulatory mechanisms controlling PA and anthocyanin biosynthesis in Arabidopsis.

3.2 INTRODUCTION

Flavonoids are secondary metabolites that fulfill a multitude of functions during plant growth and development (Winkel-Shirley, 2001b; Lepiniec *et al.*, 2006). Three main classes of flavonoids are found in *Arabidopsis thaliana*, namely flavonols, anthocyanins and

proanthocyanidins (PAs). PAs, also called condensed tannins, are flavonoid polymers resulting from the condensation of flavan-3-ol units that specifically accumulate in the innermost cell layer of the testa (chalaza, micropyle and endothelium), conferring to the mature seeds its characteristic brown colour (Debeaujon *et al.*, 2003b; Pourcel *et al.*, 2005; Lepiniec *et al.*, 2006). In Arabidopsis, most of the mutants impaired in flavonoid accumulation have been isolated through screening for altered seed pigmentation which results in the *transparent testa* (*tt*) phenotype (Koornneef, 1990). The molecular characterization of the *tt loci* has allowed the flavonoid biosynthetic pathway to be deciphered.

The structural genes leading to PA biosynthesis in Arabidopsis seeds are usually divided into two groups, the so-called early (*EBGs*) and late (*LBGs*) biosynthetic genes (Pelletier *et al.*, 1999; Lepiniec *et al.*, 2006). The EBGs include *chalcone synthase* (*CHS*), *chalcone isomerase* (*CHI*), *flavonol 3-hydroxylase* (*F3H*) and *flavonol 3'-hydroxylase* (*F3'H*), which are involved in precursor biosynthesis for the three classes of Arabidopsis flavonoids. The *LBGs* comprise *dihydroflavonol-4-reductase* (*DFR*), *leucoanthocyanidin dioxygenase* (*LDOX*) and *BANYULS/anthocyanidin reductase* (*BAN*/*ANR*). A third group of structural genes, which includes *laccase 15* (*TT10*), *MATE transporter* (*TT12*), *glutathione-S-transferase* (*TT19*) and H^+*-APTase* (*AHA10*), has been shown to be involved in PA modification, transport and oxidation (Kitamura *et al.*, 2004; Baxter *et al.*, 2005; Pourcel *et al.*, 2005; Marinova *et al.*, 2007; DeBolt *et al.*, 2009; Routaboul *et al.*, 2012). Finally, a gene whose function has not yet been clearly elucidated, namely *TT15/UGT80B1* (*UDP-glucose:sterol-glucosyltransferase*), has been shown to affect the PA accumulation process in seeds. TT15 catalyses the synthesis of steryl glycosides, which, in turn, affect the trafficking of lipid polyester precursors and, most probably indirectly, PA accumulation (Kitamura *et al.*, 2004; Baxter *et al.*, 2005; Pourcel *et al.*, 2005; Marinova *et al.*, 2007; DeBolt *et al.*, 2009; Routaboul *et al.*, 2012).

Numerous studies have indicated that the expression of the genes involved in the flavonoid biosynthetic pathway is controlled by distinct mechanisms in a tissue- or species-specific manner (Winkel-Shirley, 2001b; Koes *et al.*, 2005; Lepiniec *et al.*, 2006; Petroni & Tonelli, 2011; Schaart *et al.*, 2013). This regulation involves different sets of

transcriptional regulators that are specific to each class of flavonoids. In Arabidopsis, *EBGs* and *FLS1* (*FLAVONOL SYNTHASE 1*, which leads to the accumulation of flavonols) expression is controlled by at least three R2R3-MYBs, namely PFG1/MYB12, PFG2/MYB11 and PFG3/MYB111 (Stracke *et al.*, 2007). The regulation of PA and anthocyanin biosynthesis involves the combined action of specific R2R3-MYB (subgroup 5 and 6) and R/B-like basic helix-loop-helix (bHLH) (subgroup IIIf) transcription factors (TFs), together with TRANSPARENT TESTA GLABRA1 (TTG1) (WD repeat protein), in a MYB-bHLH-WDR (MBW) ternary protein complex (Heim *et al.*, 2003; Baudry *et al.*, 2004; Zimmermann *et al.*, 2004; Lepiniec *et al.*, 2006; Dubos *et al.*, 2010; Thevenin *et al.*, 2012). In vegetative tissues, anthocyanin biosynthesis is regulated by different MBW complexes involving PRODUCTION OF ANTHOCYANIN PIGMENT 1 (PAP1)/MYB75 and PAP2/MYB90, in combination with GLABRA3 (GL3)/bHLH00, ENHANCER OF GLABRA 3 (EGL3)/bHLH002, TT8/bHLH042 and TTG1 (Zhang *et al.*, 2003; Baudry *et al.*, 2006; Feyissa *et al.*, 2009; Appelhagen *et al.*, 2011). In seeds, TT2/MYB123, TT8 and TTG1 are the main PA biosynthesis regulators (Lepiniec *et al.*, 2006). The TT2-TT8-TTG1 protein complex directly regulates the expression of *BAN*, which encodes the first enzyme committed to the PA biosynthetic pathway (Baudry *et al.*, 2004). In addition, *DFR* and *LDOX* are direct targets of the TTG1:GR chimeric protein (a translational fusion between TTG1 and the glucocorticoid receptor; (Gonzalez *et al.*, 2008). Moreover, two recent studies on *TT1*/*WIP1* (WIP zinc finger), a TF that has been hypothesized to regulate the competency of endothelium cells to synthesize and accumulate PAs, have suggested that the transcriptional regulatory network that controls PA biosynthesis and accumulation may be more complicated than previously thought (Appelhagen *et al.*, 2010; Appelhagen *et al.*, 2011). TT1 may act synergistically with the TT2-TT8-TTG1 complex to regulate the expression of the *EBGs*, such as *CHS*, as already suggested for other MBW complexes during vegetative development (Borevitz *et al.*, 2000; Tohge *et al.*, 2005; Appelhagen *et al.*, 2011; Appelhagen *et al.*, 2011). In addition, the overexpression of MYBL2 (R3-MYB), a negative regulator of MBW complex transcriptional activity in PA accumulating cells, leads to a lower accumulation of PAs and *LBGs*, *CHS*, *F3H* and *F3'H* mRNA levels (Dubos *et al.*, 2008b). This later finding

may be the result of either a negative metabolic feedback or a negative regulation of a transcriptional activator of *EBGs*. In this respect, it has also been shown that *TT8* promoter activity is itself partially regulated by TT1 (Xu *et al.*, 2013b). From these studies it is tempting to speculate that MYBL2 may also inhibit the activity of both the TT2-TT8-TTG1 complex and the TT1 TF (Zimmermann *et al.*, 2004; Dubos *et al.*, 2008; Appelhagen *et al.*, 2011). Finally, functional redundancy involving MYB5 and either EGL3 or GL3 could be at play in this regulatory process, but the set of structural genes that may be directly regulated by these additional TFs remains to be identified (Baudry *et al.*, 2004; Lepiniec *et al.*, 2006; Gonzalez *et al.*, 2009).

TFs regulate the expression of their target genes through the interaction with specific DNA *cis*-regulatory elements usually localized upstream the transcribed region (within the promoters). Therefore, the identification of these *cis*-elements is an essential step towards a comprehensive understanding of the transcriptional regulation. The means by which the TT2-TT8-TTG1 complex regulates its target genes have been previously explored through the study of the *BAN* promoter activity, allowing the identification of a 236-bp-long minimal promoter upstream from the translational start site (Debeaujon *et al.*, 2003; Baudry *et al.*, 2004; Thevenin *et al.*, 2012). In these studies specific R2R3-MYB and bHLH binding sites through which the TT2-TT8-TTG1 complex regulates *BAN* expression were identified and characterized. Similar *cis*-regulatory motifs were also identified in the promoter of key Arabidopsis genes involved in trichome and root hair differentiation (i.e. *GL2*, *CAPRICE (CPC)*, *TTG2* and *MYB23*), from which additional MBW complexes (involving GL1/MYB0, WEREWOLF (WER)/MYB66 or MYB23 together with GL3, EGL3 and TTG1) regulate their expression (Koshino-Kimura *et al.*, 2005; Ryu *et al.*, 2005; Ishida *et al.*, 2007; Kang *et al.*, 2009; Song *et al.*, 2011). Altogether, these studies suggest that the MBW transcriptional regulatory complexes may regulate the expression of the entire set of their target genes through specific interaction with MYB and bHLH *cis*-regulatory motifs, although this remains to be proven.

In this study a combination of genetic, biochemical and molecular methods was undertaken in order to dissect the regulation of PA biosynthetic genes orchestrated by the

MBW complexes. The regulation of gene expression was investigated through the characterisation of mRNA accumulation level (using quantitative reverse transcription-polymerase chain reaction, qRT-PCR) and promoter activity (using the *β-Glucuronidase* (*GUS*) reporter gene) for the so-called *EBGs* and *LBGs*, as well as for *TT10*, *TT12*, *TT15*, *TT19* and *AHA10*, in single and multiple loss-of-function mutant backgrounds. This strategy revealed that only *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* are regulated by the TT2-TT8-TTG1 complex. The use of transgenic plants that express the inducible TTG1:GR chimeric protein showed that these genes are direct targets of the MBW complex. Genetic analyses demonstrated that partially overlapping and redundant transcriptional regulatory activities were at play for all the direct target genes, except *TT19*, in both seed and seedlings. We found that MYB5 was acts redundantly with TT2 only in the endothelium, whereas TT8, EGL3 and GL3 act together in the chalaza area. The means by which these MBW complexes interact with the promoters of their target genes were investigated through a promoter deletion strategy using both transient and stable transformation assays in *Physcomitrella patens* protoplasts and Arabidopsis seeds, respectively. Taken together, these results provide a comprehensive view of the transcriptional regulation of the flavonoid biosynthetic pathway by the MBW regulatory complexes in Arabidopsis seed.

3.3 MATERIALS AND METHODS

Unless otherwise stated below, all the primers used in this study are described in Supporting Information Table 3-1.

3.3.1 Plant materials

Arabidopsis thaliana (L.) Heynh accessions Wassilewskija (WS) and Landsberg (Ler) were used as wild-type (WT) controls where appropriate. The *Arabidopsis thaliana tt2-1*, *tt8-3*, *ttg1-1*, *tt8 egl3, tt8 gl3 egl3*, and *myb5-1* mutant lines, as well as the lines overexpressing TTG1 fused to the GR domain, have been described elsewhere (Debeaujon *et al.*, 2003; Zhang *et al.*, 2003; Baudry *et al.*, 2004; Gonzalez *et al.*, 2009). The *tt2 myb5* double mutant

Name	Sequence	Application
attB1-proCHS	attB1-TGGGCCTGAAGACTTTAAGC	*CHS* promoter cloning
attB2-proCHS	attB2-TATAGTATACACCAACTTGG	*CHS* promoter cloning
attB1-proCHI	attB1-AATTAAGAAAGTGTGGAGC	*CHI* promoter cloning
attB1-proCHI	attB2-TGTTGAGTCGGTTGGAATTTCG	*CHI* promoter cloning
attB1-proDFR	attB1-TAGAATCCACGTGGACGAGG	*DFR* promoter cloning
attB1-proDFR	attB2-TTTTGTGGTTATATGATAG	*DFR* promoter cloning
attB1-proLDOX	attB1-TCCCTTAGAGATAGAAAAGGG	*LDOX* promoter cloning
attB1-proLDOX	attB2-CTGTTTAAAGCTAAAACAGAGCAG	*LDOX* promoter cloning
attB1-proTT10	attB1-ACGTCACATCAACCACTCCAGC	*TT10* promoter cloning
attB2-proTT10	attB2-TTTGGAAGAGTTTTAGTAAAT	*TT10* promoter cloning
attB1-proTT12	attB1-GAATTCACAATCGGAAAGTCACG	*TT12* promoter cloning
attB2-proTT12	attB2-GGTCCGTTTATTAGTTCCTCTG	*TT12* promoter cloning
attB1-proTT15	attB1-TTTGGTCCCGTAATGTAGCACGG	*TT15* promoter cloning
attB1-proTT15	attB2-CACCTTTAAAAACTGAATTCAACTAAACAGC	*TT15* promoter cloning
attB1-proTT19	attB1-AAGCTTGCTTGTAAAGCACG	*TT19* promoter cloning
attB1-proTT19	attB2-TCTATAAGTTTACAATAAC	*TT19* promoter cloning
attB1-proAHA10	attB1-ACCATTGGTCCATCGTTAATCC	*AHA10* promoter cloning
attB1-proAHA10	attB2-ACTTAAAAGCTTCAAAATGTATTACC	*AHA10* promoter cloning
attB1-proDFR-1	attB1-TAGAATCCACGTGGACGAGG	promoter deletion cloning
attB1-proDFR-2	attB1-CAAGATTTATAATTATTTTAGG	promoter deletion cloning
attB1-proDFR-3	attB1-GTCAACGTATTTCACCCACCGG	promoter deletion cloning
attB1-proDFR-4	attB1-ACACATCTCTTTAGTCCTTCG	promoter deletion cloning
attB1-proDFR-5	attB1-GTTCCCCACGTGCTTCTCCGGTTG	promoter deletion cloning
attB1-proDFR-6	attB1-TTCTCCGGTTGGTACTCACGTG	promoter deletion cloning
attB1-proDFR-7	attB1-GTACTCACGTGACCGGCAGCTTC	promoter deletion cloning
attB2-proDFR$_{35Smini}$	attB2-35Smini-TTTTGTGGTTATATGATAG	promoter deletion cloning
attB1-proLDOX-1	attB1-TCCCTTAGAGATAGAAAAGGG	promoter deletion cloning
attB1-proLDOX-2	attB1-GTCAAATTGTTTTTTTAGTAATTG	promoter deletion cloning
attB1-proLDOX-3	attB1-TTGACCAATCAGTCAACCCAACC	promoter deletion cloning
attB1-proLDOX-4	attB1-TGAAGTGAGTCACGCACTTACC	promoter deletion cloning
attB2-proLDOX$_{35Smini}$	attB1-35Smini-CTGTTTAAAGCTAAAACAGAGCAG	promoter deletion cloning
attB1-proTT12-1	attB1-GAATTCACAATCGGAAAGTCACG	promoter deletion cloning
attB1-proTT12-2	attB2-CAGGTACAGTTCATCAAATGC	promoter deletion cloning
attB1-proTT12-3	attB1-ACATTCGTTGAAGAACATTAGG	promoter deletion cloning
attB1-proTT12-4	attB2-ACAAGAAAAATATTAATCGGC	promoter deletion cloning
attB1-proTT12-5	attB1-AGTCACAAGAGATCTTAACC	promoter deletion cloning
attB1-proTT12-6	attB2-TCAGTCGTAATAAAAAGACGCC	promoter deletion cloning
attB2-proTT12$_{35Smini}$	attB2-35Smini-GGTCCGTTTATTAGTTCCTCTG	promoter deletion cloning
attB1-proTT19-1	attB1-AAGCTTGCTTGTAAAGCACG	promoter deletion cloning
attB1-proTT19-2	attB2-GAACTTTGTTATTGATAGTTGG	promoter deletion cloning
attB1-proTT19-3	attB1-TACCCAACCCTCACAACAACC	promoter deletion cloning
attB1-proTT19-4	attB2-ATTTCTTCTCATTTATAAACC	promoter deletion cloning
attB1-proTT19-5	attB1-TTGATAGACAAGATATTAATAAGTGC	promoter deletion cloning
attB2-proTT19$_{35Smini}$	attB2—35Smini-TCTATAAGTTTACAATAAC	promoter deletion cloning
attB1-proAHA10-1	attB1-ACCATTGGTCCATCGTTAATCC	promoter deletion cloning
attB1-proAHA10-2	attB2-AAAGAGGAAGCTAGTGAGAGTG	promoter deletion cloning
attB1-proAHA10-3	attB1-ATGCTTCCAATAATGGTGATAC	promoter deletion cloning
attB1-proAHA10-4	attB2-GAATCTCCTCTTCTCCACACTC	promoter deletion cloning
attB1-proAHA10-5	attB1-GTCAGCCATGGAATGGTCGTTTC	promoter deletion cloning
attB2-proAHA10$_{35smini}$	attB2-35Smini-ACTTAAAAGCTTCAAAATGTATTACC	promoter deletion cloning
QTT19-F	CATCTTCTTCGTCAGCCATTTGGT	qRT-PCR
QTT19-R	CGTCACATTTCTCGCCTAACCT	qRT-PCR
QAHA10-F	CCGAGGATTTGGACAAGCCA	qRT-PCR
QAHA10-R	GACCTAGACTTTGACTATTAGC	qRT-PCR
	attB1: GGGGACAAGTTTGTACAAAAAAGCAGGCT	
	attB2: GGGGACCACTTTGTACAAGAAAGCTGGGTC	
	35smini: GTCCTCTCCAAATGAAATGAACTTCCTTATATAGAGGAAGGGTCTTGC	

Table 3-1. Primers used in this study

line used in this study was obtained by crossing *tt2-1* with *myb5-1*. Plant growth, transformation, and selection for transgenic lines were performed as reported previously (Nesi *et al.*, 2000).

3.3.2 Studied Arabidopsis gene IDs

TT2/MYB123, At5g35550; *MYB5*, At3g13540; *TT8/bHLH042*, At4g09820; *GL3*/*bHLH001*, At5g41315; *EGL3*/*bHLH002*, At1g63650; *TTG1*, At5g24520; *CHS*, At5g13930; *CHI*, At2g43570; *F3H*, At3g51240; *F3'H*, At5g07990; *DFR*, At5g42800; *LDOX*, At4g22880; *BANYULS*/*ANR*, At1g61720; *TT10*, At5g48100; *TT12*, At3g59030; *TT15*, At1g43620; *TT19*, At5g17220; *AHA10*, At1g17260.

3.3.3 Constructs

The *proOsActine:TT2*, *proOsActine:MYB5*, *proOsActine:TT8*, *proOsActine:GL3*, *proOsActine:EGL3*, and *proOsActine:TTG1* vectors used to express the corresponding genes in the moss *P. patens* protoplasts have been described elsewhere (Thevenin *et al.*, 2012; Xu *et al.*, 2013).

All promoters (ranging from 0.5 to 2kb, which correspond to the whole intergenic regions) in this study were amplified by PCR using high-fidelity Phusion DNA polymerase (New England Biolabs) from WS genomic DNA and subsequently recombined into *pDONR207* vector (BP Gateway® reaction) according to manufacturer's instructions, and sequenced to ensure their integrity. Promoters were then inserted into destination vector by LR reactions (Gateway®). The destination vectors used for the promoters were the pBS TPp-B (green flurescent protein (GFP) reporter gene; (Thevenin *et al.*, 2012) and the *pGWB3* (GUS reporter gene; (Nakagawa *et al.*, 2007) for *P. patens* protoplast transfection assays and Arabidopsis stable transformation, respectively.

The 5' deletion fragments were amplified from the promoters inserted in the *pGWB3* vector as templates and cloned as described above. Primers used for amplification were modified by the addition of the *35S* cauliflower mosaic virus (CaMV) minimal promoter

sequence before the 5'-end of the reverse primers.

The *proTT8:MYB5* vector was generated by an LR recombination as described above from the two following vectors: *proTT8:GTW* (TT8 promoter) and *pDONR207-MYB5* (*MYB5* coding sequence). Both vectors are described elsewhere (Dubos *et al.*, 2008; Xu *et al.*, 2013).

3.3.4 Transient expression assays in Physcomitrella patens protoplasts

Moss culture, protoplast preparation, protoplast transformation and Flow cytometry measurement analysis were carried out as previously described (Thevenin *et al.*, 2012). Each experiment was repeated at least three times, with three technical repetitions per experiment.

3.3.5 Expression analysis

Total RNA was extracted from 4-day-old siliques, treated with DNase, reverse transcribed into cDNA, and assayed using qRT-PCR as previously described (Baudry *et al.*, 2004; Baudry *et al.*, 2006). The primer pairs used in this study for *CHS*, *CHI*, *F3H*, *F3'H*, *DFR*, *LDOX*, *BAN*/*ANR* and *TT10* have been described elsewhere (Baudry *et al.*, 2004; Baudry *et al.*, 2006; Dubos *et al.*, 2008). Primers used for *TT12* (At_TT12_1_SG QuantiTect primer assay) and *TT15* (At_AT1G43620_1_SG QuantiTect primer) were purchased from Qiagen. For each experiment (i.e. mutant analysis or dexamethasone (DEX) induction), cDNA samples were generated from three different plants per genotype, each being assayed three times by qRT-PCR (technical repetitions). All the experiments were repeated at least three times.

3.3.6 DEX induction experiments and RNA analysis

For each treatment, four siliques (4 d after pollination) were taken from three independent plants for each line, opened and incubated in 24-well plates in the presence of Mock buffer alone or with 100 μM cycloheximide (CHX; Sigma-Aldrich) before a 30-min vacuum treatment to ensure effective infiltration of CHX. DEX was then added to a final

concentration of 10 μM, and another 30-min vacuum treatment was applied. After 3 h, the reaction media were renewed, and incubation was continued for 3 h. After 6 h, samples were collected into 1.5-mL Eppendorf Safe-Lock Tubes™ (Hamburg, Germany) and immediately frozen into liquid nitrogen (Baudry *et al.*, 2004).

3.3.7 Histochemical detection of GUS activity

GUS staining for seeds expressing *promoter:uidA* gene fusion (and deleted versions) constructs was performed in the presence of 2 mM potassium ferri-/ferrocyanide, when necessary, as described elsewhere (Debeaujon *et al.*, 2003; Berger *et al.*, 2011). For each construct, 8 to 24 independent transgenic plants were analysed, and representative observations are presented.

3.3.8 Complementation of the transparent testa phenotype of tt2

The *tt2-1* mutant plants were stably transformed with the *proTT8:MYB5* vector, from which 12 independent transgenic lines displayed a WT seed colour (brown).

3.3.9 Statistical analysis

Each experiment (i.e. qRT-PCR analysis and transient assays in *P. patens* protoplasts) was repeated at least three times (biological repetitions). For each sample, three measures were carried out (technical repetitions). Student's *t*-tests were used to compare samples against their corresponding controls. The null hypothesis was rejected when the *P* value was below 5% ($P<0.05$).

3.4 RESULTS

3.4.1 Contribution of TT2, TT8 and TTG1 to the expression of PA biosynthetic genes

The involvement of TT2, TT8 and TTG1 in the expression of 12 characterized flavonoid

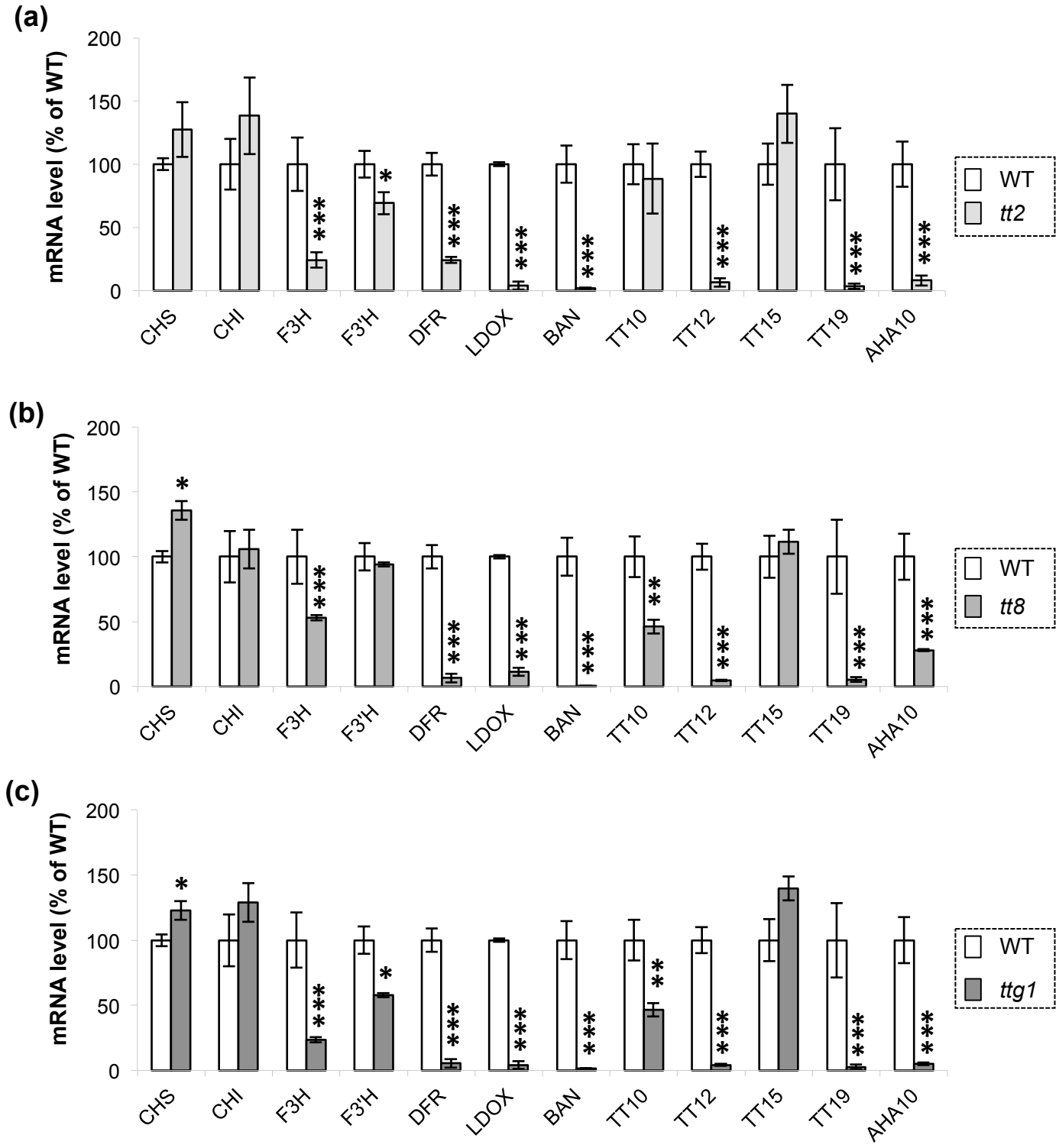


Figure 3-1. mRNA level of proanthocyanidin (PA) biosynthetic genes in wild-type (WT) and *tt2*, *tt8* and *ttg1* mutant siliques. Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis of the mRNA levels for the PA biosynthesis genes was measured in (a) *tt2*, (b) *tt8* and (c) *ttg1* loss-of-function *Arabidopsis thaliana* mutants and their corresponding WT in siliques 4 d post-fertilization. Measurements are expressed as the percentage of the WT level. *t*-test significant difference: * = $P<0.05$; ** = $P<0.01$; *** = $P<0.001$. Error bars, ±SE from three biological repetitions. In (a, c) WT is Landsberg erecta; (b) WT is Wassilewskija.

biosynthetic genes leading to PA accumulation was investigated by comparing mRNA accumulation levels between wild type (WT) and the *tt2*, *tt8* and *ttg1* mutants (Fig. 3-1). The mRNA steady-state levels were quantified in 4-d-old siliques by using qRT-PCR as, at this stage of silique development (globular stage), most of these mRNAs accumulate to their maximum (Nesi *et al.*, 2000; Nesi *et al.*, 2001; Kleindt *et al.*, 2010). A significant decrease in *F3H, DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* mRNA accumulation was observed in all three mutants when compared with the WT (Fig. 3-1). *F3'H* also displayed a lower transcript accumulation in the *tt2 and ttg1* mutants when compared with the WT, whereas no variation was observed in the *tt8* mutant. A weak decrease in *TT10* mRNA accumulation was also observed in the *tt8* and *ttg1* mutants when compared with the WT, but not in the *tt2* mutant. A slight but insignificant increase in *CHS* mRNA accumulation was observed in the *tt2* mutant and for *CHI* in the *tt2* and *ttg1* mutants. Finally, a significant increase in *CHS* mRNA accumulation was observed in *tt8* and *ttg1*.

3.4.2 Activity of 12 PA pathway promoter sequences in Arabidopsis

The promoters (i.e. intergenic region of c. 0.5-2 kb upstream from the translation initiation start (TIS) site) of the 12 studied structural genes were fused to the *uidA* (GUS) reporter gene and stably introduced into the WT and various regulatory mutant backgrounds.

As shown previously for the *F3H*, *BAN*, *TT10*, *TT12*, *TT15* and *AHA10* promoters, all the promoters analysed in this study, except for *F3'H*, were active in PA-accumulating cells (i.e. chalaza, micropyle and endothelium) of WT seeds (Fig. 3-2, 3-3; Harper *et al.*, 1994; Debeaujon *et al.*, 2003; Pourcel *et al.*, 2005; Marinova *et al.*, 2007; DeBolt *et al.*, 2009; Berger *et al.*, 2011). Although c. 24 transgenic plants were analysed, no GUS activity was found for the *F3'H* promoter, preventing further analyses.

The activity remained unchanged for *CHS*, *CHI*, *F3H*, *TT10*, and *TT15* promoters in the seed coat of *tt2*, *tt8* and *ttg1* mutants. This result is fully consistent with mRNA accumulations found for *CHS*, *CHI* and *TT15*. It also suggests that the lower mRNA accumulation observed for *F3H* and *TT10* does not result from a direct transcriptional regulation by the TT2-TT8-TTG1 complex in the seed coat.

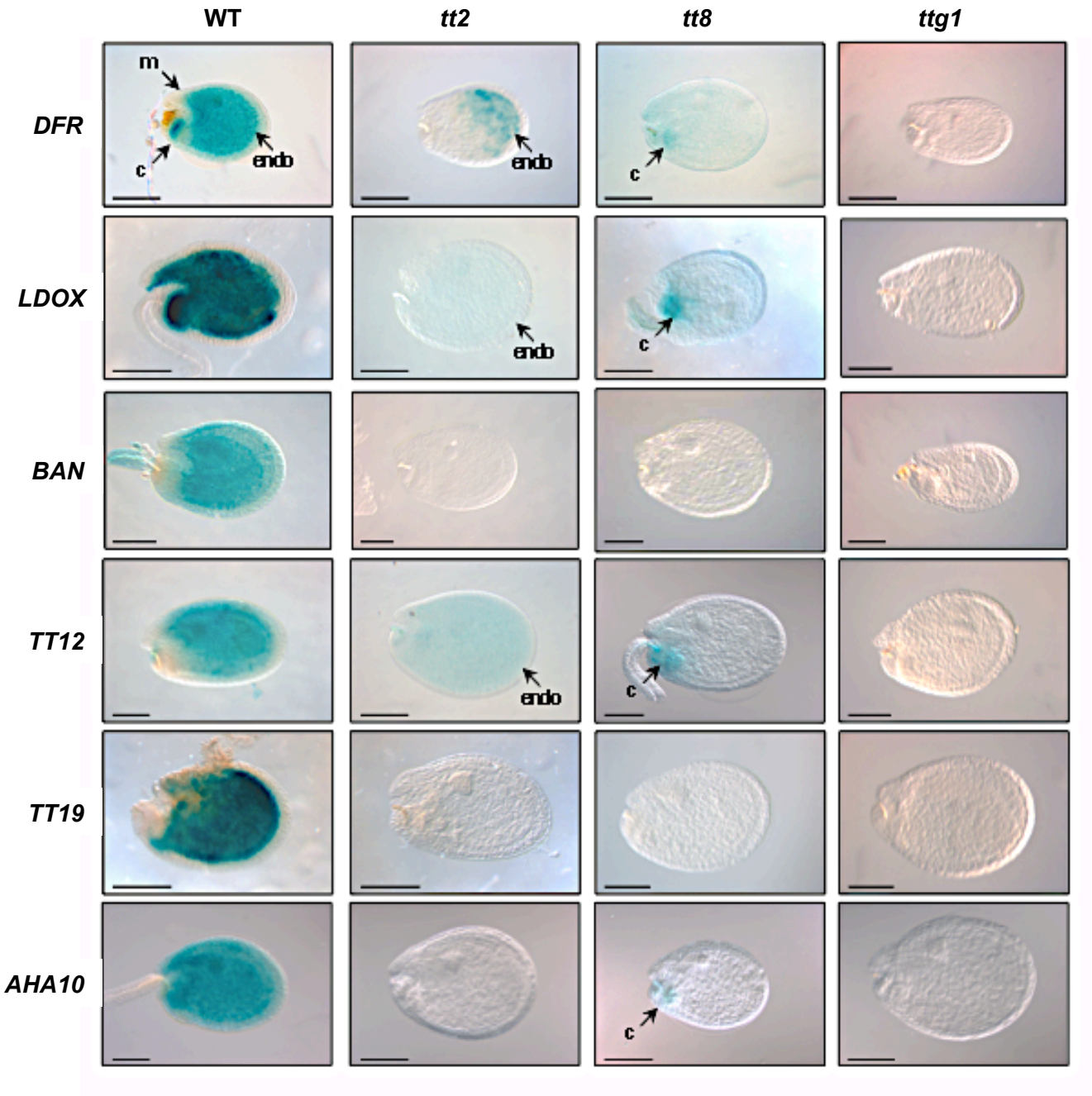


Figure 3-2. Promoter activity of proanthocyanidin (PA) biosynthetic genes in seed. Histochemical analysis of *β-Glucuronidase* (GUS) activity generated by the promoters fused to the *uidA* reporter gene that were differentially affected in *tt2*, *tt8* and *ttg1* loss-of-function *Arabidopsis thaliana* mutants when compared with the wild-type (WT). Photographs correspond to whole-mount seeds at the globular stage. c, chalaza; endo, endothelium; m, micropyle. Bars, 0,1 mm. 2 mM potassium ferricyanide/potassium ferrocyanide was used for the promoter of *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* in WT, and *DFR* in *tt2* and *tt8*.

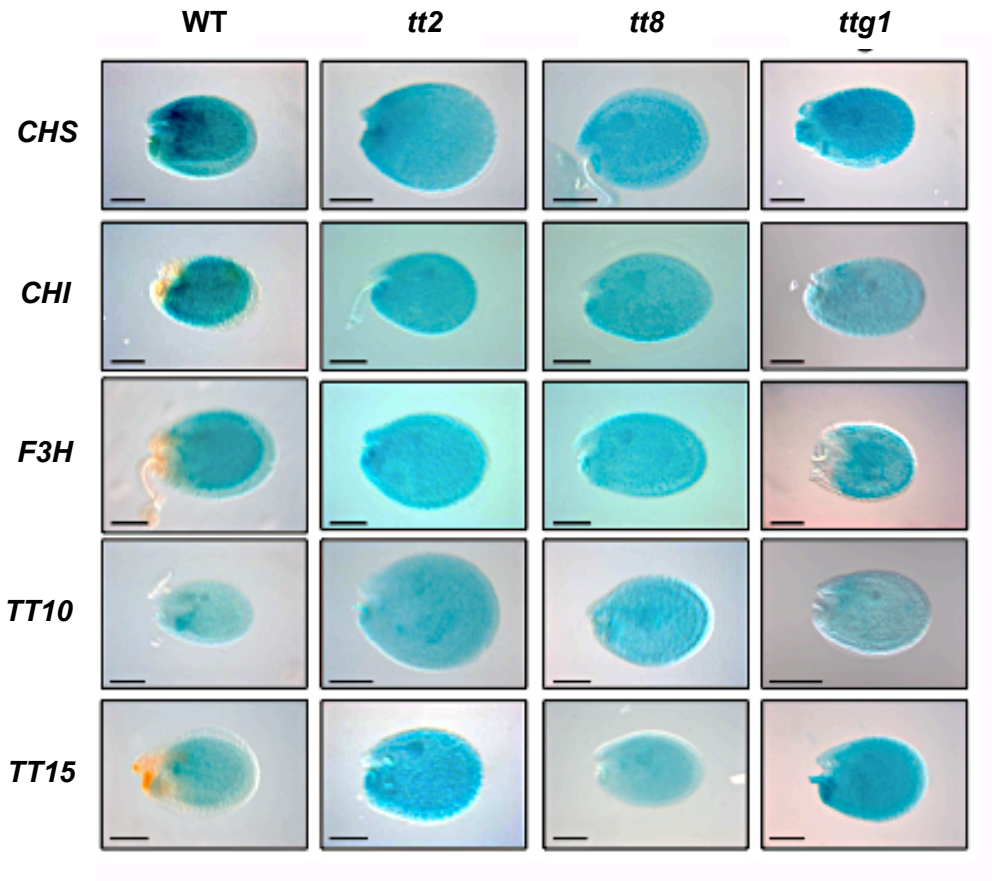


Figure 3-3. Promoter activity of PA biosynthetic genes that are not directly regulated by the TTG1-dependent complexes in seed. Pictures correspond to whole-mount seeds at the globular stage. Scale bars: 0.1 mm. 2 mM potassium ferricyanide/potassium ferrocyanide was used.

By contrast, *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* promoter activities were strongly altered in the three studied mutant backgrounds (Fig. 3-2). No activity could be observed for the six genes in the *ttg1* mutant, even though a faint and diffuse GUS activity was observed in a few lines for the *DFR* promoter. Interestingly, *DFR*, *LDOX* and *TT12* showed similar GUS patterns in the seed coat of *tt8* (in chalaza) and *tt2* (in endothelium), whereas no activity was found in *tt2* and *tt8* for both *TT19* and *BAN* promoters. It is noteworthy that, in the present study, the previously characterized *BAN* minimal promoter was used (236-bp fragment upstream from the translational initiation site; Debeaujon *et al.*, 2003). Finally, when *AHA10* promoter activity was tested in these three mutants, blue staining was only found in the chalaza of *tt8* (Fig. 3-2).

Because anthocyanins accumulate in vegetative tissues, the activity of the different promoters was assayed in WT and mutant 10-d-old seedlings. No activity was found for the promoters of *BAN*, *TT10, TT12 and AHA10* (which code for proteins specifically involved in PA metabolism). By contrast, *DFR*, *LDOX*, and *TT19* promoter activity was detected in the cotyledon margin, independent of *TT2* and *TT8*. Nevertheless, these activities were totally lost in *ttg1* (Fig. 3-4a). Finally, the activities of the *CHS*, *CHI*, *F3H* and *TT15* promoters were found throughout the whole seedling, with the exception of the roots for the promoter of *TT15*, and were unaffected in the three mutant backgrounds tested (Fig. 3-4b).

Finally, promoter activities were assayed in different organs of 5-wk-old plants (Fig. 3-5). *BAN*, *TT10* and *TT12* promoters were found to specifically drive GUS activity in seeds, and not in any other of the organs tested. In rosette leaves, depending on the assayed promoter, blue staining was found in the whole organ (*CHS* and *TT15*), in the leaf margin (*CHI* and *F3H*) or mainly in the mid-vein and lamina (*DFR*, *LDOX*, *TT19* and *AHA10*). In cauline leaves, GUS activity was detected in the whole tissues or in the margin, for the *CHS* and *TT15*, and the *CHI*, *F3H*, *DFR* and *LDOX*, promoters, respectively. The *CHS*, *CHI*, *F3H* and *TT15* promoters also displayed some GUS activity in petals, whereas the *TT10* promoter was active in anthers and stigmas. In the apical part of the stems, blue staining was found for the *CHS*, *F3H* and *TT15* promoters. In the lower part of the stem, GUS activity was found after cutting induction in the cortex and the cambium for the *CHS*, *CHI*, *DFR LDOX* and

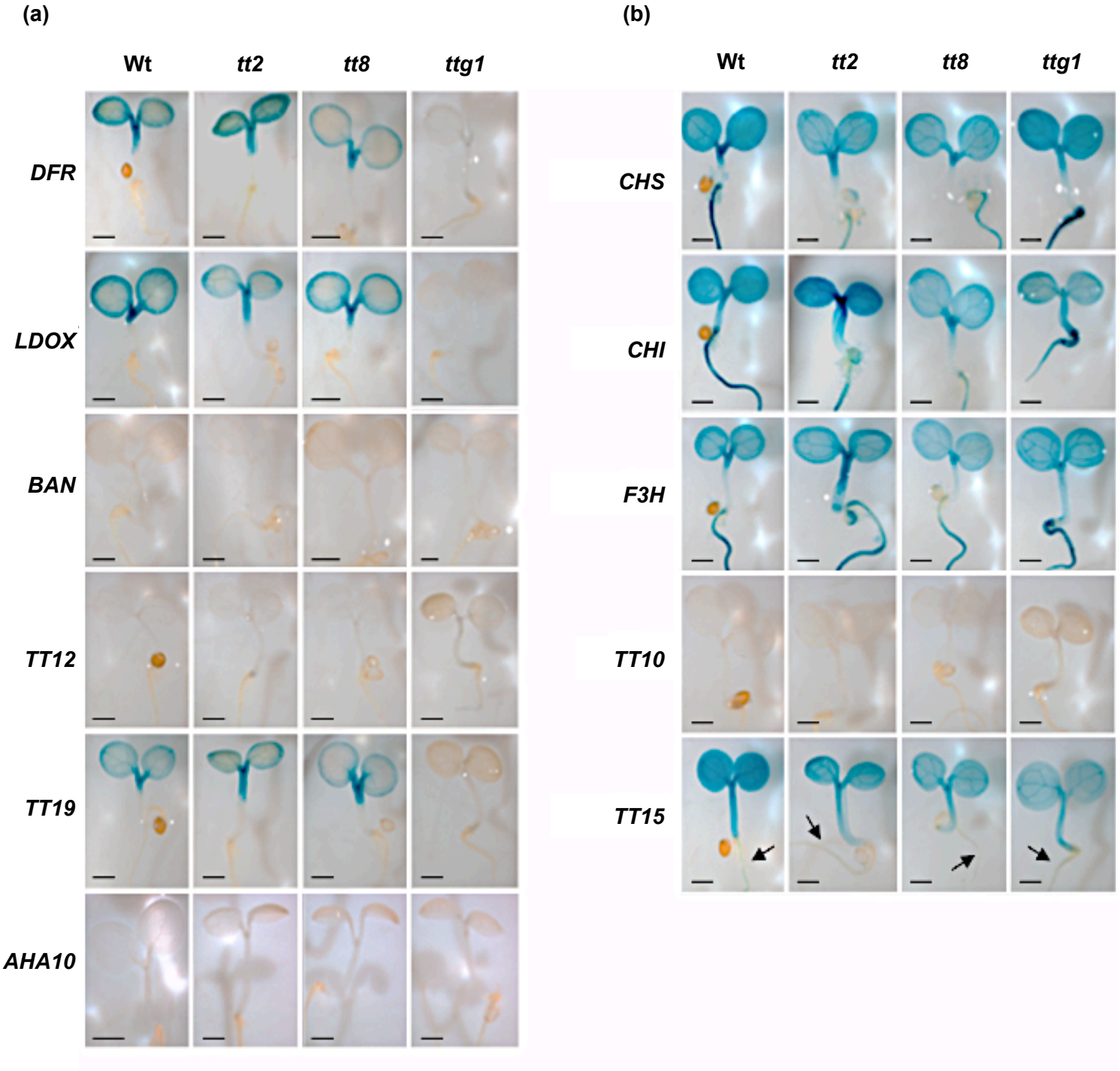


Figure 3-4. Promoter activity of PA biosynthetic genes, in 10-day-old seedlings in WT, *tt2*, *tt8* and *ttg1* mutant backgrounds. (a) TTG1-dependent and (b) -independent transcriptional regulation. Arrows indicate the absence of GUS staining in roots. Scale bars: 0.2 mm.

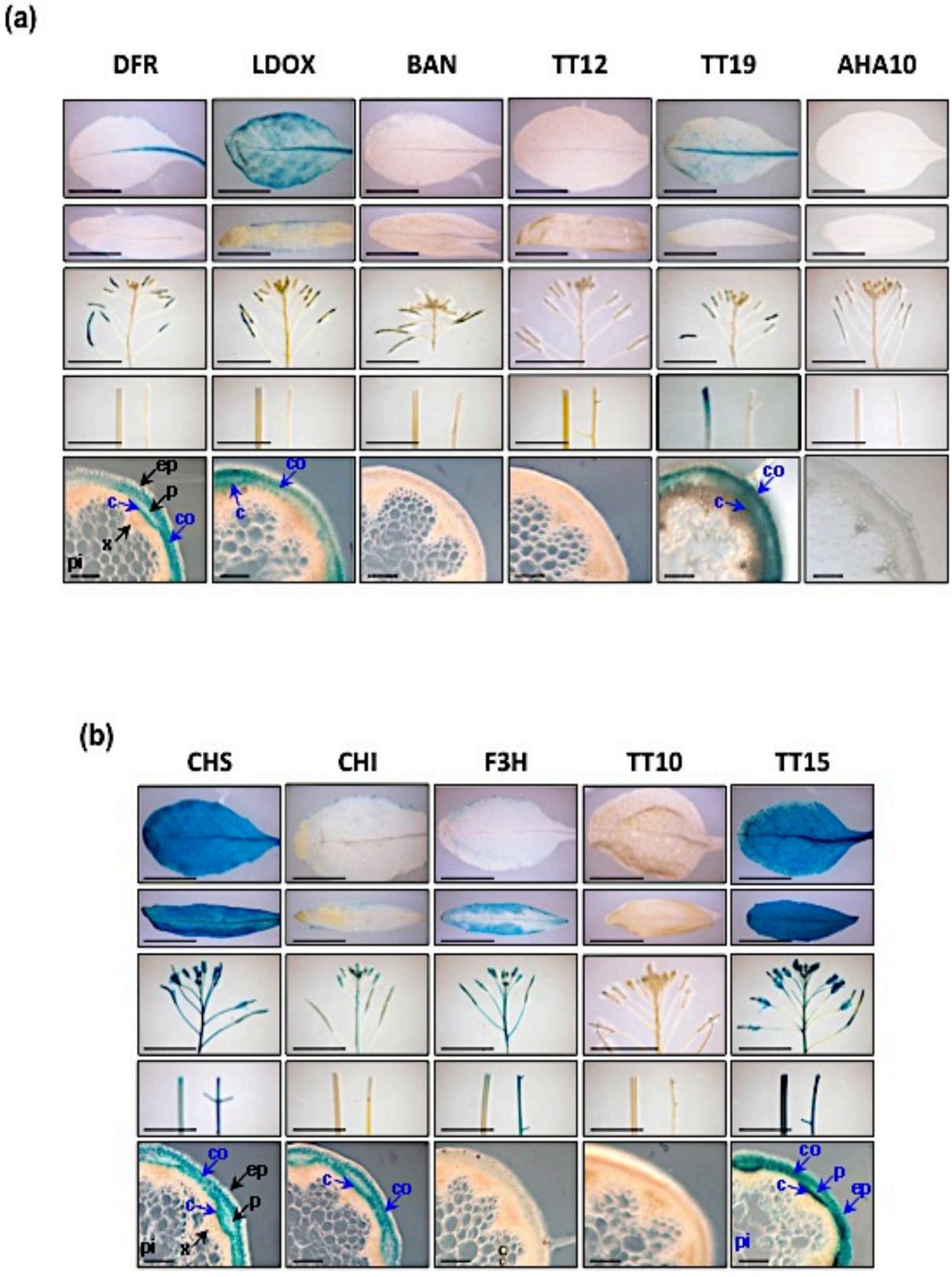


Figure 3-5. Promoter activity of PA biosynthetic genes in 5-week-old WT plants. (a) TTG1-dependent and (b) TTG1-independent transcriptional regulations. From top to bottom: Rosette leave, cauline leave, inflorescence, stem (left: lower part, right: upper part), stem cross section (ep: epidermis, c: cambium, co: cortex, p: phloem, pi: pith, x: xylem).

TT19 promoters. By contrast, *TT15* promoter activity was found throughout the stem (except in the xylem).

3.4.3 Identification of the primary targets of TTG1-dependent regulatory complexes

The characterization of promoter activities and mRNA accumulations suggested that *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* are primary targets of TT2, TT8 and TTG1. This is consistent for *BAN*, which has been shown previously to be directly regulated by TT2, TT8, and TTG1 (Baudry *et al.*, 2004; Baudry *et al.*, 2006).

This later finding was based on the analysis of plants that constitutively overexpressed the chimeric TTG1:GR protein, which is unable to translocate into the nucleus without the addition of DEX. Primary target genes were identified as being induced at the transcriptional level after the simultaneous treatment of DEX and CHX, an inhibitor of protein translation. Thus, the same transgenic lines were used to determine whether a TTG1-dependent regulatory complex directly controls the expression of *DFR*, *LDOX*, *TT12*, *TT19* and *AHA10*. The corresponding mRNA steady-state levels were determined by qRT-PCR on 4-d-old siliques, after 6-h treatments (Fig. 3-6). Significant and reproducible inductions of *BAN* mRNA accumulation were obtained after DEX, and DEX plus CHX, treatments, compared with the Mock and CHX controls. By contrast, no effect on *TT10* mRNA accumulation was observed with any of these treatments. These positive (*BAN*) and negative (*TT10*) controls demonstrated that the experimental conditions were effective. Interestingly, DEX treatment resulted in a significant induction of *DFR*, *LDOX*, *TT12*, *TT19* and *AHA10* mRNA levels, which was also observed when DEX was used in combination with CHX. Altogether, these data demonstrate that *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* are primary targets of TTG1-dependent regulatory complexes.

3.4.4 Characterization of functionally redundant MYB-bHLH-TTG1 complexes in seeds

MYB5 and the two bHLH GL3 and EGL3 are known to be involved in TTG1-dependent

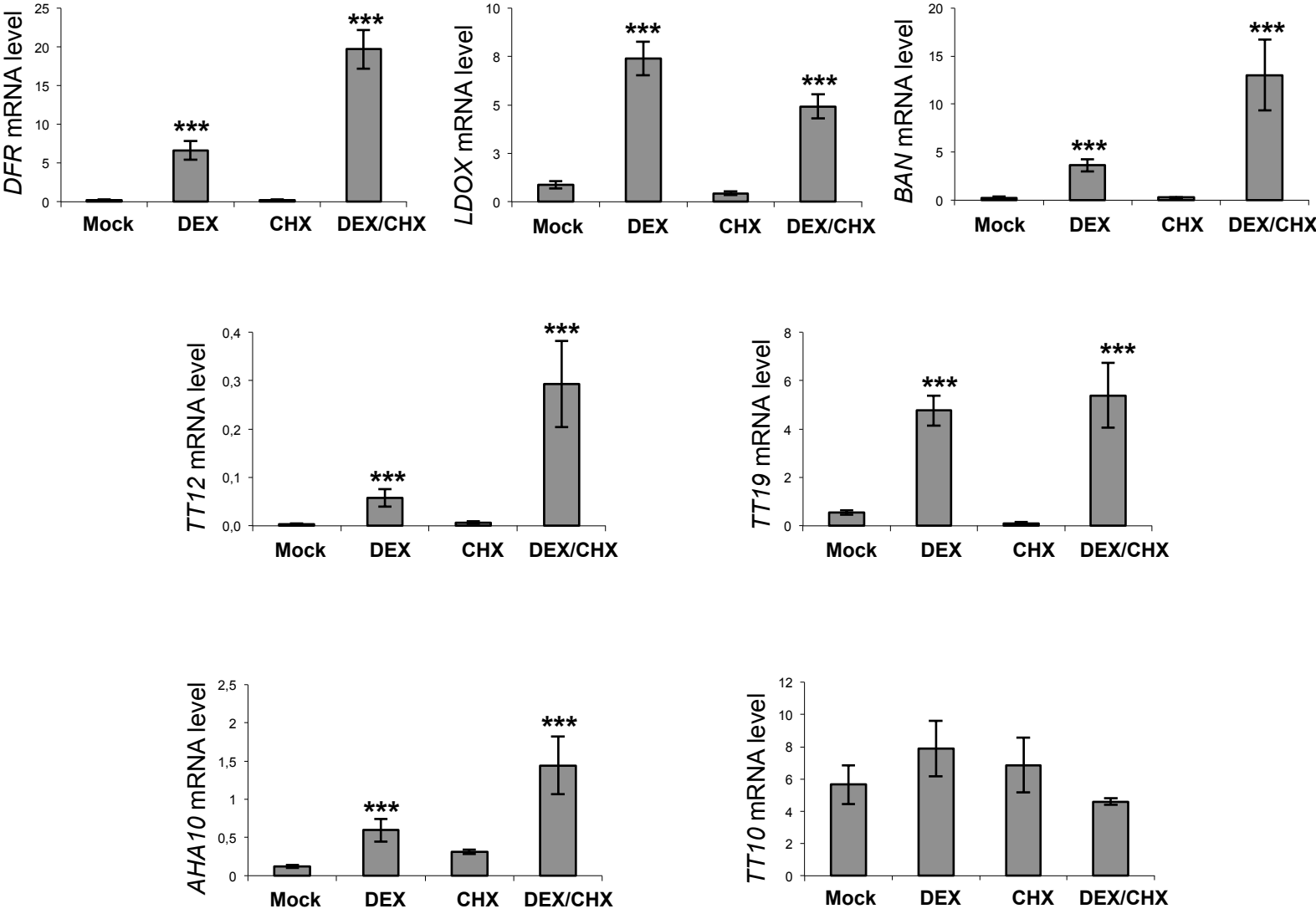


Figure 3-6. Direct activation of *DFR*, *LDOX*, *BAN*, *TT19*, *TT12* and *AHA10* expression by the TTG1:GR fusion protein in *Arabidopsis thaliana* siliques. The level of *DFR*, *LDOX*, *BAN*, *TT19*, *TT12*, *AHA10* and *TT10* mRNA was measured by quantitative reverse transcription-polymerase chain reaction (qRT-PCR). Values are expressed as a percentage of the constitutively expressed *EF1αA4* gene. *t*-test significant difference: ***, *P*<0.001. Error bars, ±SE from three biological repetitions. Mock, buffer; DEX, dexamethasone; CHX, cycloheximide.

regulatory complexes (Heim *et al.*, 2003; Zhang *et al.*, 2003; Gonzalez *et al.*, 2009; Li *et al.*, 2009). In order to test if these proteins could also regulate the activity of the *LBG* promoters in seeds, two complementary strategies were used.

First, the different MBW combinations were assayed against the *DFR*, *LDOX*, *TT12* and *AHA10* promoters using the *P. patens* protoplast transient expression system (Thevenin *et al.*, 2012). These four promoters have been chosen because they are direct targets of TTG1-dependent complexes that still display some activity in the *tt8* and/or *tt2* mutant seeds (Fig. 3-2). This approach showed that EGL3 could functionally replace TT8 in combination with TT2 and TTG1 to induce the activity of the four promoters, and that MYB5 was also able to functionally replace TT2 with either TT8 or EGL3, but to a lesser extent (Fig. 3-7). Weak but significant activations of the *DFR* and *AHA10* promoters were also observed when GL3 was assayed in combination with MYB5.

Secondly, genetic validation was carried out *in planta* by analysing mRNA accumulation or promoter activity in double or triple loss-of-function mutants (i.e. *tt2 myb5, tt8 egl3 or tt8 egl3 gl3*). The accumulation of *DFR*, *LDOX* and *TT12* mRNAs in the *tt2 myb5* double mutant revealed that the expression of these three genes is regulated in the endothelium by both TT2 and MYB5 (Fig. 3-8a). *AHA10* was not investigated in the *tt2 myb5* double mutant because no promoter activity was detected in *tt2* (Fig. 3-2). In this respect, no GUS activity was detected in the seeds of the *tt2 myb5* double mutant when the promoters of *DFR*, *LDOX* and *TT12* were analysed (Fig. 3-8b). These findings are supported by the ability of MYB5 to partially complement the *transparent testa* phenotype of *tt2* when overexpressed in seeds (i.e. light brown colour of the complemented seeds) under the control of the *TT8* promoter (Fig. 3-8c) (Baudry *et al.*, 2006; Dubos *et al.*, 2008). In addition, GUS analysis showed that, in the chalaza, TT8 and EGL3, but not GL3, regulate the expression of *LDOX*, *TT12* and *AHA10*, whereas the three bHLHs are involved in the regulation of *DFR* expression (Fig. 3-8d, e).

3.4.5 Functional dissection of the MBW target gene promoters

Transient expression in *P. patens* protoplasts was used to characterize, within the *DFR*,

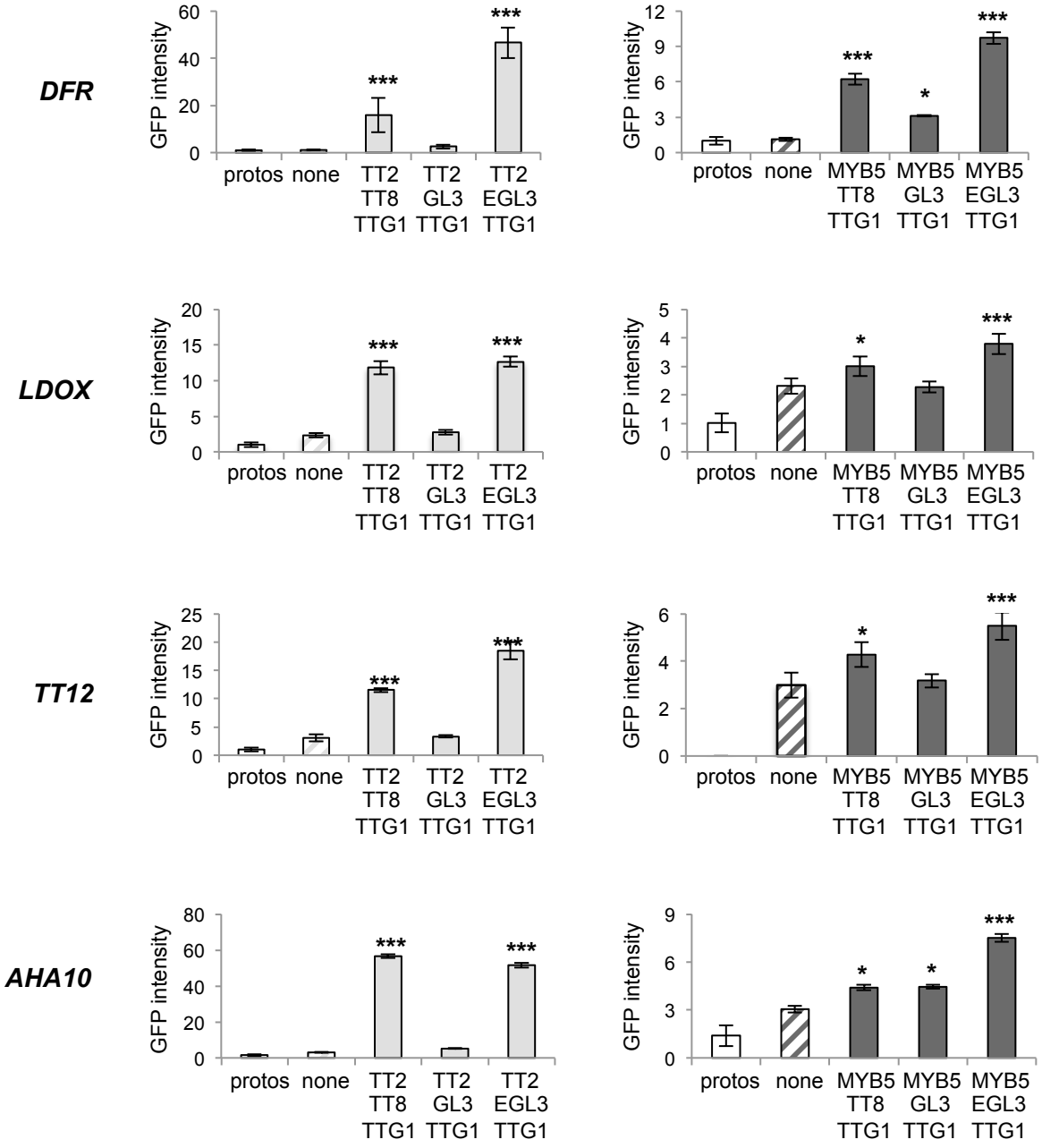


Figure 3-7. Effect of different R2R3-MYB (TT2 and MYB5) and R/B-like bHLH (TT8, EGL3 and GL3), with TTG1, on *DFR*, *LDOX*, *TT12* and *AHA10* promoter activity. Promoters were fused to the green fluorescent protein (GFP) reporter gene, and tested by transient expression assays in *Physcomitrella patens* protoplasts. Transactivation activities were monitored by GFP fluorescence. Co-transfection assays were carried out with either TT2 (left panels) or MYB5 (right panels). *t*-test significant difference: *, $P<0.05$; ***, $P<0.001$. Error bars, ±SE from three biological repetitions. protos, protoplasts alone; none, protoplasts transformed with the assayed promoters alone.

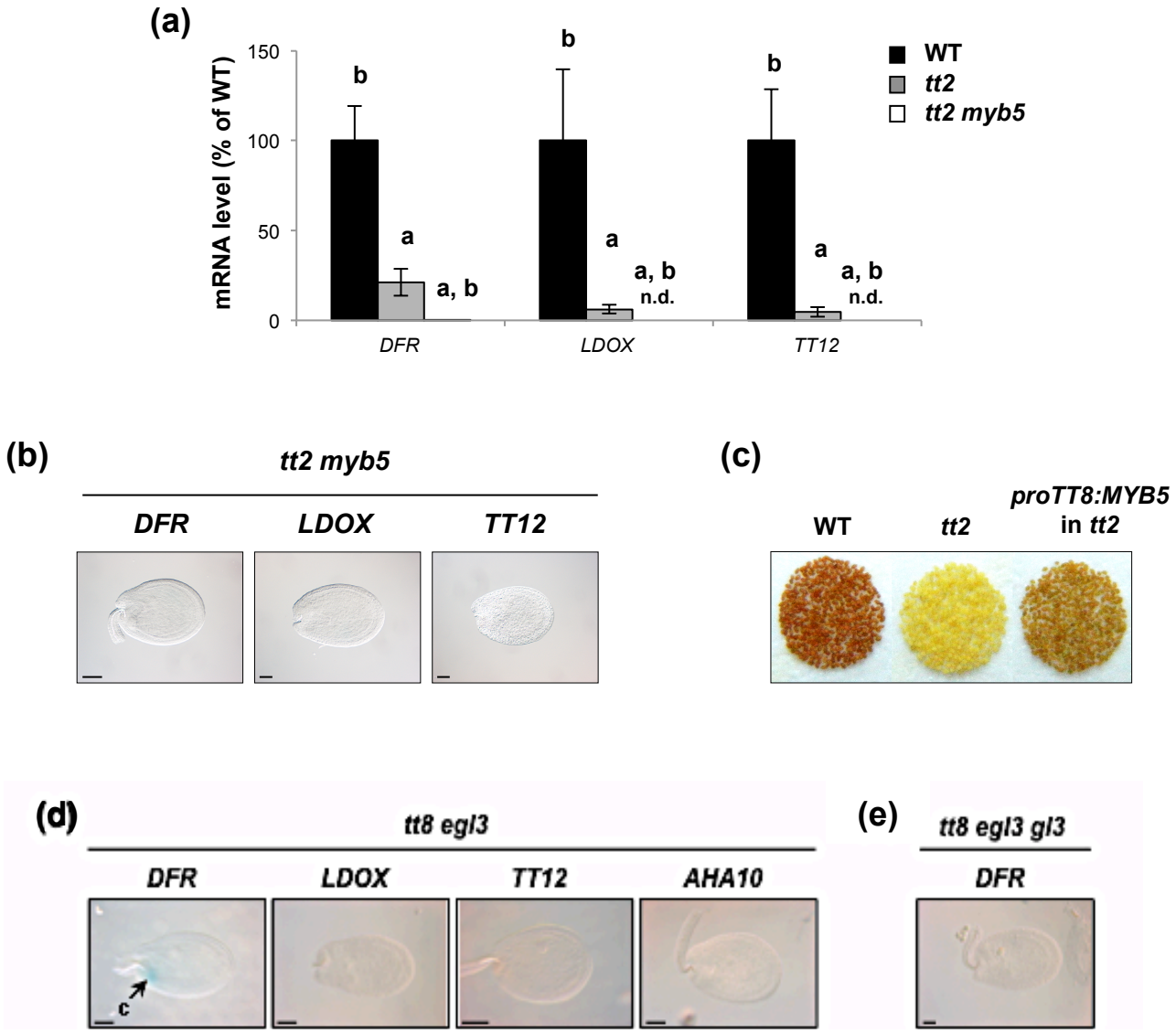


Figure 3-8. Effect of different R2R3-MYB (TT2 and MYB5) and R/B-like bHLH (TT8, EGL3 and GL3), with TTG1, on *DFR*, *LDOX*, *TT12* and *AHA10* expression in *Arabidopsis thaliana*. (a) Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis of the *DFR*, *LDOX* and *TT12* mRNA levels measured in seeds of wild-type (WT) (black bars), *tt2* (grey bars) and *tt2 myb5* (white bars) loss-of-function mutants at 4 d post-fertilization. Measurements are expressed as percentage of the WT level. Differences observed between the three genotypes were all significant. *t*-test significant difference: ***, $P<0.001$ (i.e. a *vs* WT and b *vs tt2*). Error bars, ±SE from three biological repetitions. nd, not detected. **(b)** *β-Glucuronidase* (GUS) activity driven by the promoters of the proanthocyanidin (PA) biosynthetic genes showing remaining activity in *tt2* investigated in *tt2 myb5* double mutant. **(c)** Complementation of the *transparent testa* phenotype of the *tt2* mutant by overexpression of *MYB5* in seeds. (**d, e**) GUS activity driven by the promoters of the PA biosynthetic genes showing remaining activity in *tt8* investigated in double and triple R/B-like bHLH mutants, respectively. c, chalaza.

LDOX, *TT12*, *TT19* and *AHA10* promoters, the DNA regions (modules) involved in the regulation of their activity by the MBW complexes. The TT2-TT8-TTG1 complex was chosen as a model. First, a 5'-end deletion series was generated and assayed using GFP as reporter gene (Fig. 3-9). The different promoter fragments used in this experiment were selected in accordance to the position of the putative R2R3-MYB and bHLH binding sites that were present on the studied promoters, with the aim to remove one putative target of the MBW complex in each of the generated deletions. This search was carried out using the PLACE web tool (http://www.dna.affrc.go.jp/PLACE/; (Higo *et al.*, 1999). For this search, only the MYB-core ($CNGTT^{A}/_{G}$) and *AC*-rich DNA motifs were considered as putative TT2 (or MYB5) binding sites, as similar *cis*-elements have been identified in the promoter of several genes regulated by specific R2R3-MYB involved in the secondary metabolism, or by the TT2-TT8-TTG1 complex (Grotewold *et al.*, 1994; Patzlaff *et al.*, 2003; Hartmann *et al.*, 2005; Prouse & Campbell, 2012; Thevenin *et al.*, 2012; Xu *et al.*, 2013). By contrast, all the consensus bHLH binding sequences (*CANNTG*) were searched. The effect of the MBW complex on each deletion was quantified relative to the activity of the longest promoters. From this approach, key deletion fragments (corresponding to drastic changes in GFP intensity) were then selected and assayed *in planta* using the GUS reporter gene. This strategy was chosen as it has been used successfully for the study of the *BAN* and *TT8* promoters, for which *cis*-target motifs have been identified and characterized (Debeaujon *et al.*, 2003; Thevenin *et al.*, 2012; Xu *et al.*, 2013).

Deletions of the *DFR* promoter fragment did not have a significant impact on its activity until -302bp (*proDFR-3*) before the TIS site. Removal an additional fragment of 84 bp (*proDFR-4)* led to a strong decrease (≈45%) in the activity (Fig. 3-9a). Consistent with this result, when assayed *in planta*, *proDFR-4* did not display any GUS activity. By contrast, blue staining was still found in PA-accumulating cells when *proDFR-3* was tested, indicating that this fragment contains the minimal information necessary to specifically drive transcriptional activity in this tissue. Similarly, *LDOX* promoter deletion analysis led to the identification of a 336-bp promoter fragment upstream of the ATG, namely *proLDOX-3* (Fig. 3-9b). Intriguingly, *proTT12-3* (-464 bp) was identified as containing the minimal nucleotide

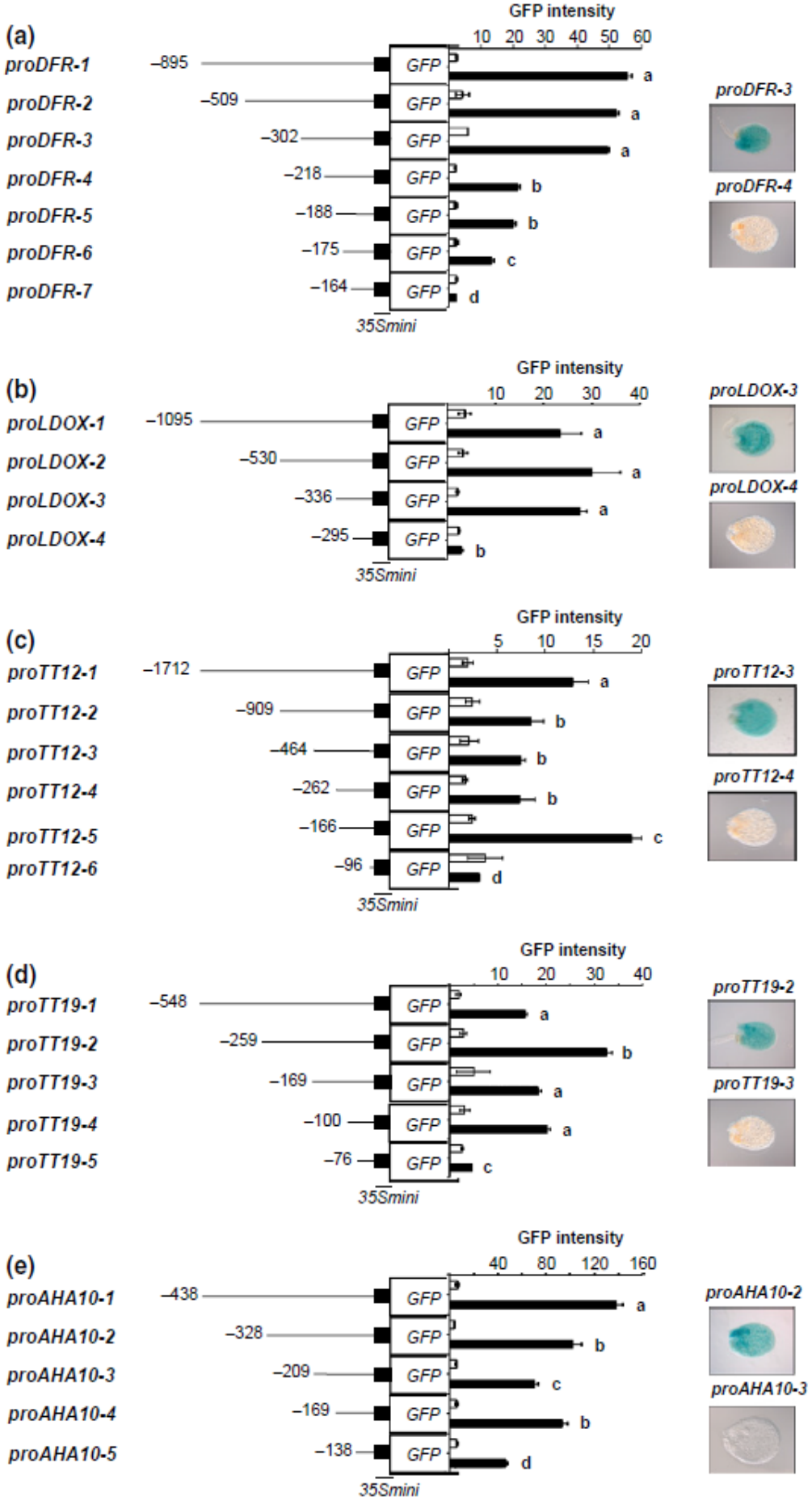


Figure 3-9. Molecular dissection of the *DFR*, *LDOX*, *TT19*, *TT12* and *AHA10* promoters. (**a**) *DFR*, (**b**) *LDOX*, (**c**) *TT12*, (**d**) *TT19* and (**e**) *AHA10* promoter analysis. Left panels: the different promoters issued from the 5'-end deletion series were fused to the *35S* cauliflower mosaic virus minimal promoter upstream of the green fluorescent protein (GFP) reporter gene, and then assayed in transient expression assays using *P. patens* protoplasts, in the presence (closed bars) or absence (open bars) of the TT2, TT8 and TTG1 proteins. Transactivation activity was monitored by GFP fluorescence in *Arabidopsis thaliana*. Identical letters represent levels of statistically equivalent green fluorescence ($P < 0.05$). Error bars, ±SE from three biological repetitions. Right panels: *β-Glucuronidase* (GUS) activity driven by the different promoters in seeds (at globular stage of embryo development).

sequence that specifically drives *TT12* expression in PA-accumulating cells, in contrast with the observations in transient expression assays (Fig. 3-9c). Such results suggest that some domains upstream of *proTT12-5* have a negative impact on *TT12* promoter activity in *P.patens* protoplasts (i.e. the occurrence of some specific repressors able to recognize *proTT12-6* in *P. patens*). A similar hypothesis can be made concerning the *proTT19-2* (259 bp before TIS site) fragment, which confers seed expression and displays a higher activity than *proTT19-1* in transient expression assays (Fig. 3-9d). Finally, this strategy has allowed the identification of a 328-bp promoter fragment (*proAHA10-2*) that specifically drives *AHA10* expression in PA-accumulating cells (Fig. 3-9e).

As the longest *DFR*, *LDOX* and *TT19* promoters were also active in WT seedlings, *proDFR-3*, *proLDOX-3* and *proTT19-2* were assayed in this tissue. As expected, these three promoter fragments were sufficient to drive GUS activity in the cotyledon margins and the upper part of the hypocotyls in cells in which the anthocyanins accumulate, whereas shorter promoter versions (i.e. *proDFR-4*, *proLDOX-4* and *proTT19-3*) were not (Fig. 3-10).

3.4.6 Identification of putative cis-regulatory elements

The characterization of short promoter domains able to confer MBW regulation in seeds and seedlings revealed the presence of various R2R3-MYB and bHLH binding sites (Fig. 3-11). When the sequences of the different classes of *cis*-regulatory motifs were compared with each other (Fig. 3-11), two subgroups per class were identified as follow: MYB core, $C^{C}/_{T}GTTA$ and $C^{A}/_{C}GTTG$; *AC*-rich element, ${}^{A}/_{C}CCAAC^{C}/_{G}$ and $ACCTA^{A}/_{C}$ (*ACI* or *ACIII*); bHLH binding site, *CANNTG* (E-box) and *CACGTG* (G-box). At least one conserved DNA motif belonging to the three classes (i.e. MYB core, *AC*-rich element and bHLH binding site) of *cis*-regulatory elements was found for all the short functional promoters, with the exception of *proLDOX-3*, for which no bHLH binding site was found (Table 3-2).

3.5 DISCUSSION

Our main objective was to carry out a comprehensive functional analysis of the MBW

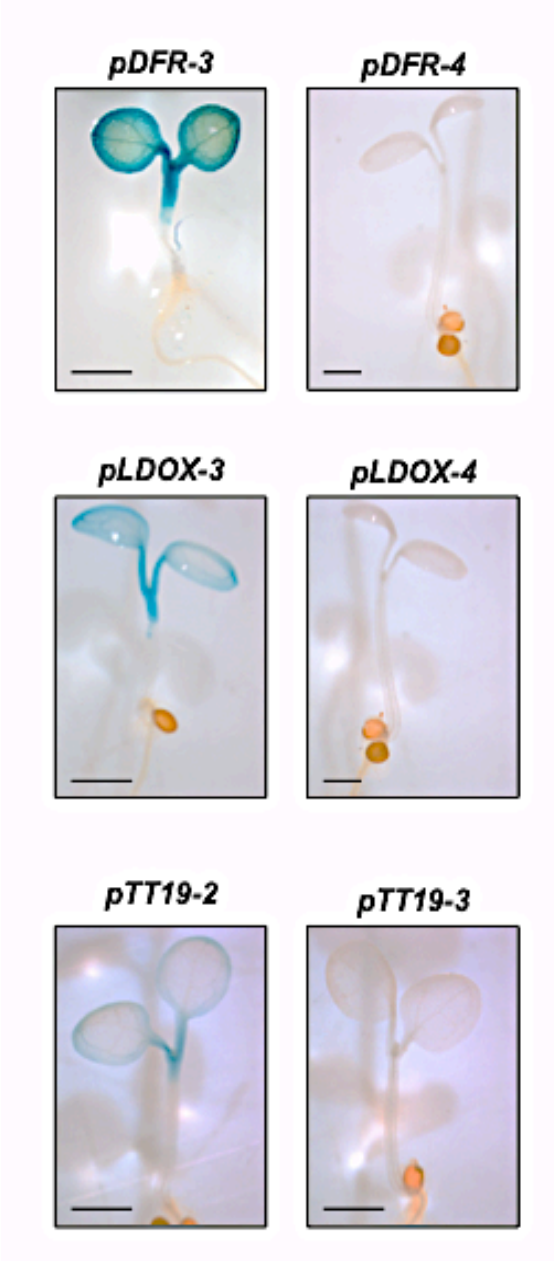


Figure 3-10. Activity of *DFR*, *LDOX* and *TT19* promoters, in 10-day-old WT seedlings. Scale bars: 0.2 mm. Both, the shortest active deletion and the first inactive deletion are shown for each promoter.

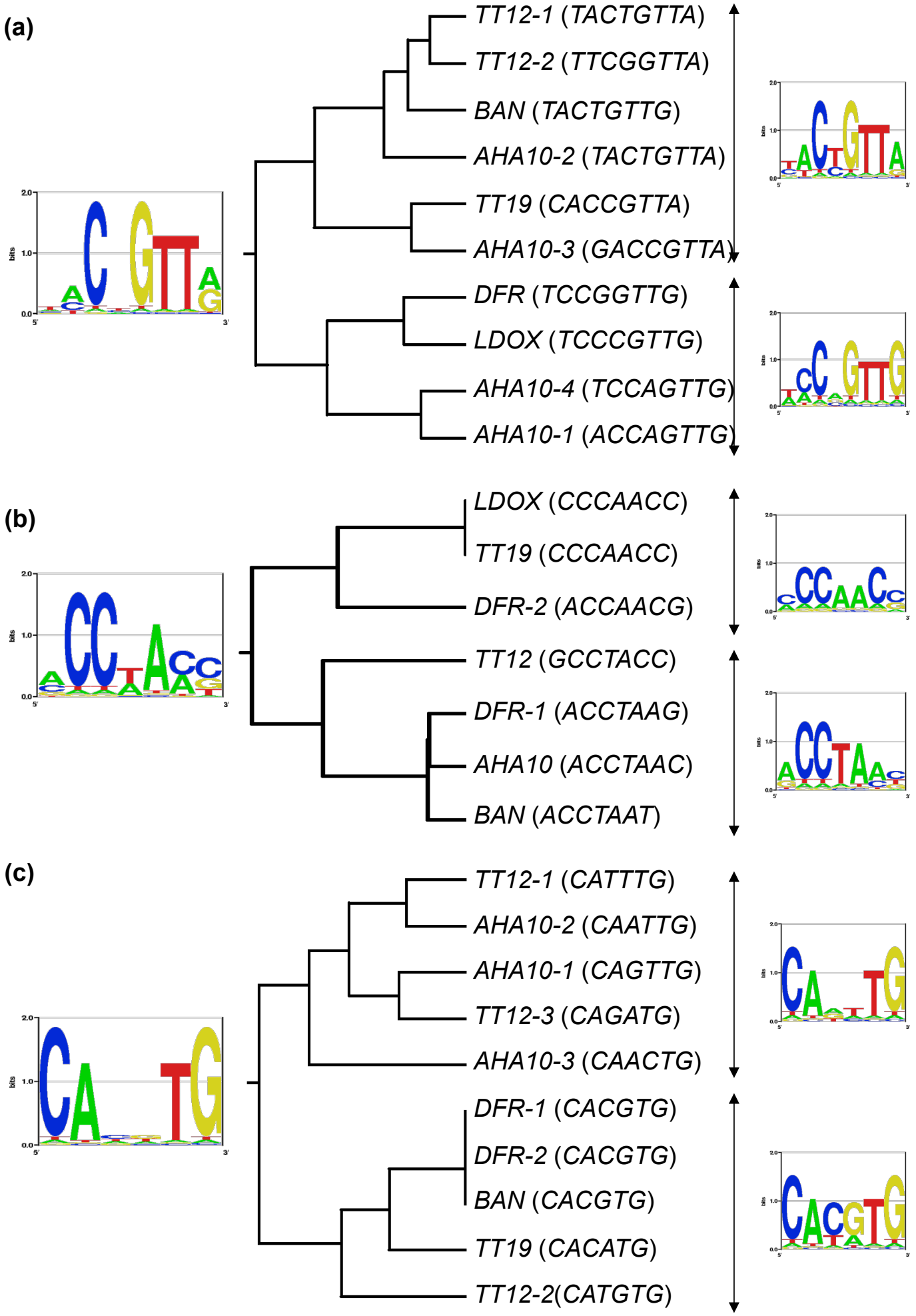


Figure 3-11. R2R3-MYB and R/B-like bHLH *cis*-regulatory elements identified in the minimal "*LBG*" promoters. (a) MYB-core, (b) *AC*-rich sequence, and (c) bHLH binding site. Left panel, consensus motif. Central panel, consensus dendrogram constructed from parsimony and distance analysis based on the *cis*-regulatory elements actually identified (1000 bootstraps). Right panel, subgroups of the different *cis*-regulatory elements.

complexes involved in PA biosynthesis and their target gene promoters in *Arabidopsis thaliana* seeds. Because most of the studied structural genes are also involved in the anthocyanin pathway, we looked by extension for similar regulatory mechanisms in seedlings (as there is no anthocyanin accumulation in seeds). For this purpose, genetic, molecular and biochemical approaches were used. In this study, we have identified the direct targets of the complexes, functionally characterized their promoters and determined the minimal promoters involved in the regulation by the MBW complexes. The results presented here demonstrated that the *LBGs*, as well as *TT12*, *TT19* and *AHA10*, are direct targets of the TT2-TT8-TTG1 complex and other *TTG1*-dependent transcriptional regulatory complexes involving MYB5, GL3 and/or EGL3. These results are discussed below.

3.5.1 PA biosynthesis is regulated by more than one MBW complex in Arabidopsis

GUS staining analysis gained from the study of the *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* promoters in immature seeds showed that the promoter of these six genes is active in PA-accumulating cells (i.e. endothelium, micropyle and chalaza) in a *TTG1*-dependent manner (Fig. 3-2). This result confirms that the tissue-specific expression of these six genes is largely controlled at the transcriptional level. Interestingly, the weak promoter activities remaining in the *tt2* and/or *tt8* mutants, but not in *ttg1*, suggest that (with the exception of the *TT19* promoter) functionally redundant R2R3-MYB and R/B-like bHLH proteins are involved in the transcriptional regulation of this specific set of promoters. Such *TTG1*-dependent redundant transcriptional activities have been reported previously in the regulation of trichome and root hair patterning, anthocyanin and PA biosynthesis and mucilage production in seed (Heim *et al.*, 2003; Zhang *et al.*, 2003; Lepiniec *et al.*, 2006; Gonzalez *et al.*, 2008; Ishida *et al.*, 2008; Gonzalez *et al.*, 2009; Pesch & Hulskamp, 2009; Schiefelbein *et al.*, 2009; Dubos *et al.*, 2010a; Feller *et al.*, 2011; Xu *et al.*, 2013).

Taking into account all of these results, we made the assumption that the functional homologues of *TT2* and *TT8* in the regulation of the *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* promoter activities could be *MYB5*, and *EGL3*/*bHLH002* and/or *GL3*/*bHLH001*,

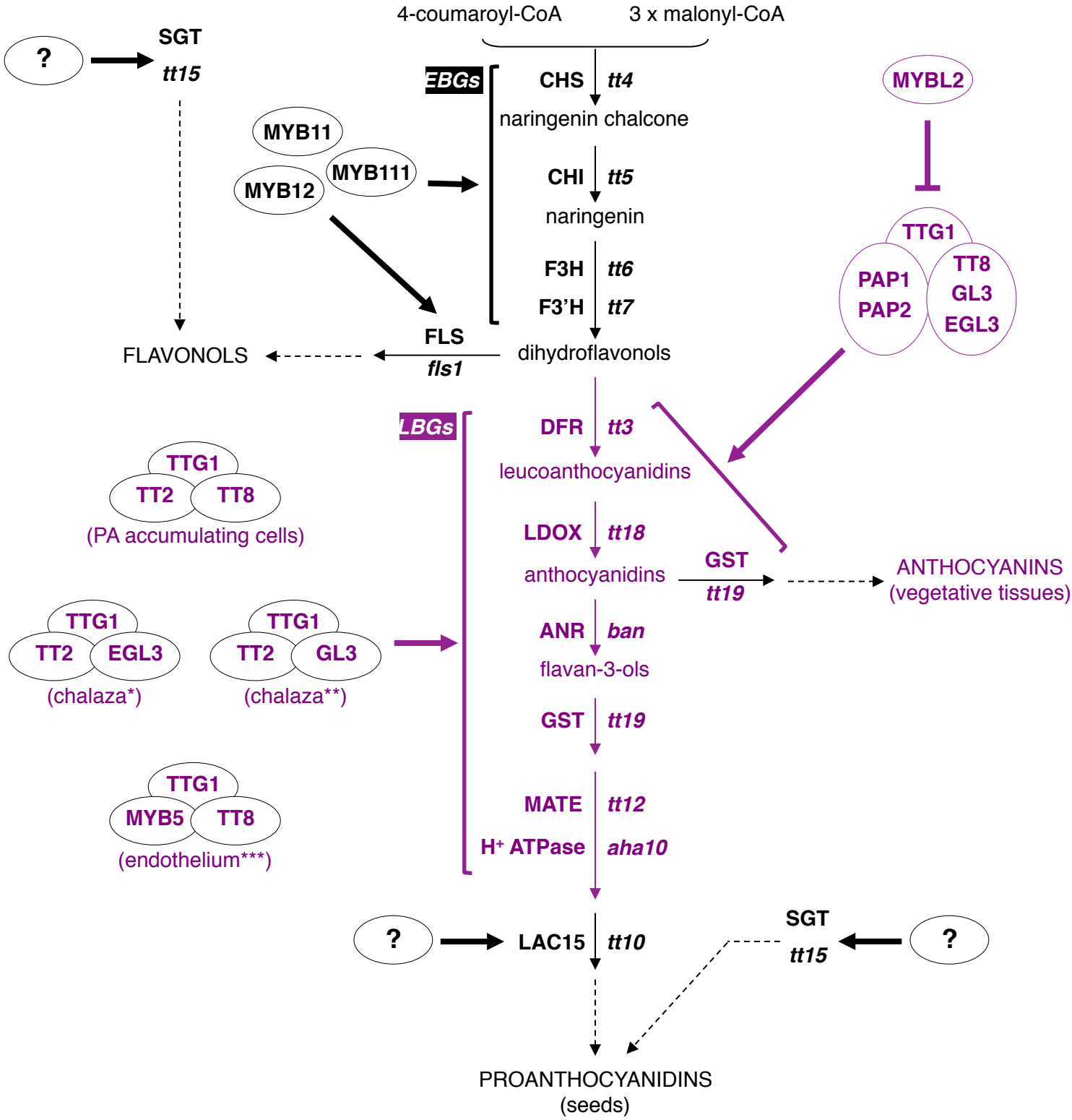


Figure 3-12. Schematic representation of late biosynthetic gene (LBG) regulations by MBW complexes in *Arabidopsis thaliana* seed. The names of the structural proteins are indicated in capital letters and corresponding mutants in lower-case italics. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavonol 3-hydroxylase; F3'H, flavonol 3'-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol-4-reductase; LDOX, leucoanthocyanidin dioxygenase; ANR, anthocyanidin reductase; MATE, multidrug and toxic efflux transporter; GST, glutathione-S-transferase; SGT, UDP-glucose:sterol-glucosyltransferase; LAC15, laccase 15. Arrows indicate the different steps leading to the formation of flavonoids in Arabidopsis; dashed lines indicate multiple steps; circles correspond to transcription factors; *EBG*, early biosynthetic gene; *LBG*, late biosynthetic gene. In purple, part of the pathway leading to proanthocyanidin (PA) (condensed tannins) and anthocyanin biosynthesis analysed in the present study, and directly regulated by the activity of different MBW complexes. In black, part of the flavonoid pathway that is not regulated by the activity of the MBW complexes. Asterisk(s) indicate the involvement of the MBW complexes mentioned above in the transcriptional regulation of: *, *DFR*, *LDOX*, *BAN* and *AHA10*; **, *DFR*, alone; or ***, *DFR*, *LDOX* and *TT12*.

respectively (Figs 3-7, 3-8). By combining transient expression assays, quantification of mRNA accumulation and promoter analysis in multiple loss-of-function mutants, we demonstrated that EGL3 was able to fully replace TT8 in the MBW complex to activate the *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* promoters, whereas GL3 was only able to slightly activate the *DFR* promoter. Interestingly, EGL3 has also been shown to play a predominant role in the regulation of anthocyanin biosynthesis when plants were grown in non-stressful conditions, with TT8 and GL3 being involved to a lesser extent (Zhang *et al.*, 2003; Cominelli *et al.*, 2008; Feyissa *et al.*, 2009). MYB5 was also able to induce the *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* promoter activities, but to a lesser extent than TT2 (Fig. 3-7). Nevertheless, this result is consistent with genetic analyses showing that *TT2* is necessary for PA biosynthesis in seeds and that, by contrast, *MYB5* plays a secondary role (i.e. no PAs are synthesized in *tt2*, whereas decreased PA accumulation is observed in *myb5*; (Nesi *et al.*, 2001; Gonzalez *et al.*, 2009). This observation is strengthened by the lack of GUS activity driven by the promoters of *DFR*, *LDOX* and *TT12* in the seeds of the *tt2 myb5* double mutant, and by the ability of *MYB5* to partially complement the *transparent testa* phenotype of *tt2* when its expression is increased in cells that specifically accumulate PAs (Fig. 3-8).

Altogether, these data show that the TT2-TT8-TTG1 complex plays the main role in the regulation of *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* expression in developing seeds, although three additional MBW complexes have partially overlapping transcriptional regulatory function. In the endothelium, the MYB5-TT8-TTG1 complex is involved in the regulation of *DFR*, *LDOX*, *TT12* and *AHA10* expression. In the chalaza, two complexes are at play, namely the TT2-EGL3-TTG1 complex, which induces *DFR*, *LDOX*, *BAN*, *TT12* and *AHA10,* and the TT2-GL3-TTG1 complex, which only controls the expression of *DFR* (Fig. 3-12). Interestingly, the above results are fully consistent with a recent tissue-specific transcriptomic analysis carried out on developing Arabidopsis seeds using laser capture microdissection coupled with GeneChip analysis (ATH1 Affymetrix; (Le *et al.*, 2010). EGL3 and MYB5 mRNA preferentially accumulate from pre-globular to heart stage of embryo development (i.e. when PA biosynthesis is occurring) in the chalaza and the endothelium,

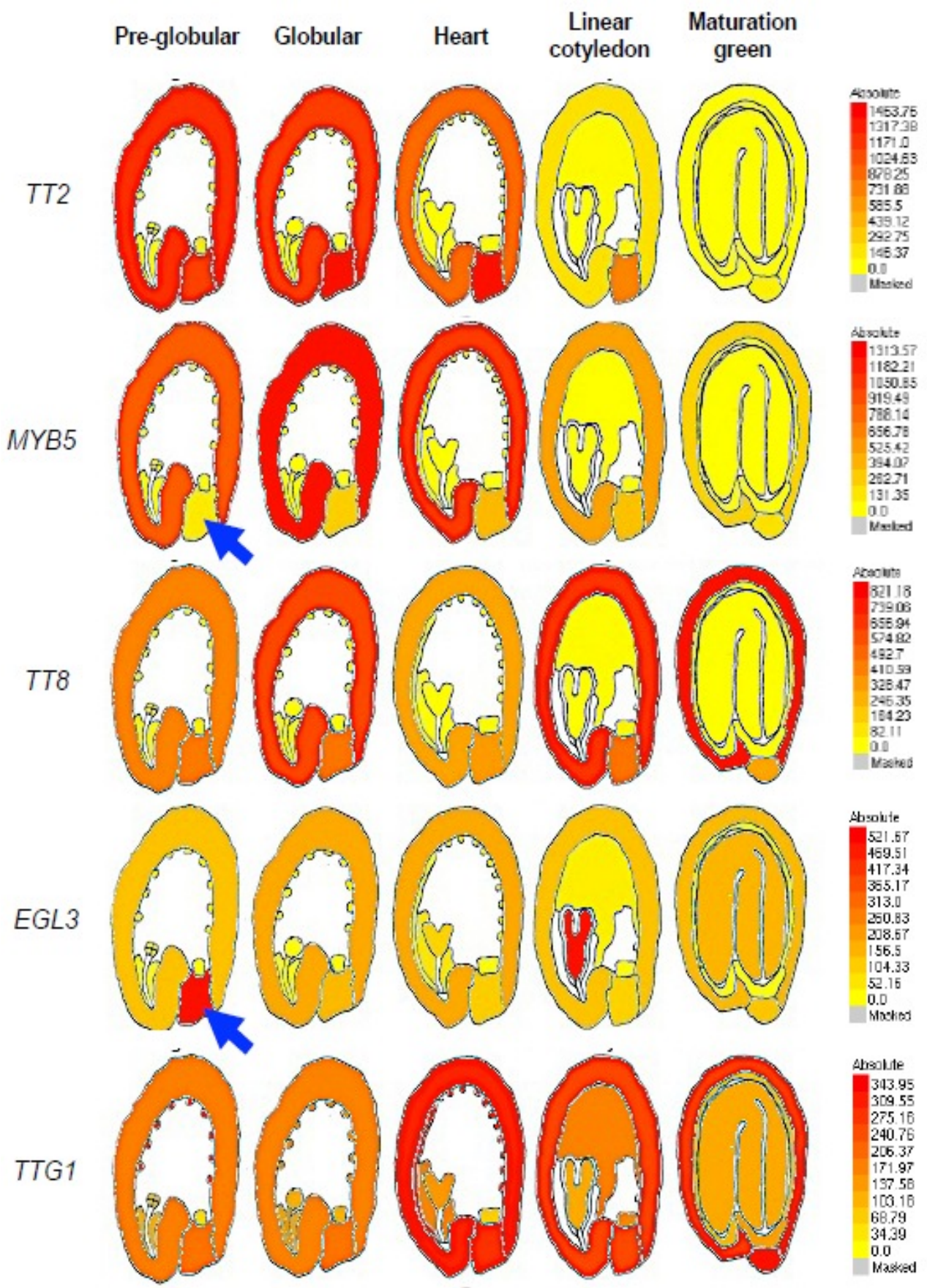


Figure 3-13. Steady state mRNA level of the MBW genes in developing Arabidopsis seeds. Data were generated by combining laser capture micro dissection coupled with ATH1 GeneChip analysis (Le *et al*., 2010). Schematic representations are available at the following URL: http://bar.utoronto.ca/efp/cgi-bin/efpWeb. Blue arrows indicate the absence or presence of mRNA accumulation in the chalaza for MYB5 and EGL3, respectively.

respectively (Figs 3-13, 3-14). Moreover, the results presented in this study are congruent with previously published data in which the promoter activity of the MBW complex members, as well as the mRNA or protein accumulation of their direct target (as revealed by *in situ* hybridisation or GFP fusion), was found to be specific to the PA accumulating cells in Arabidposis developing seeds (Devic *et al.*, 1999; Debeaujon *et al.*, 2001; Debeaujon *et al.*, 2003b; Baudry *et al.*, 2006; Gonzalez *et al.*, 2009; Li *et al.*, 2009; Kitamura *et al.*, 2010; Appelhagen *et al.*, 2011; Xu *et al.*, 2013).

Interestingly, the analyses carried out in seedlings suggest that similar redundancies occur in vegetative tissues (Fig. 3-4). This latter observation is fully consistent with previous reports showing that mutations of the three bHLHs (namely TT8, EGL3 and GL3) or two R2R3-MYBs (PAP1 and PAP2) are necessary for the complete prevention of anthocyanin accumulation in seedlings (Zhang *et al.*, 2003; Appelhagen *et al.*, 2011a).

3.5.2 Functional analysis and regulation of the MBW target gene promoters

We have shown that *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* are the direct targets of different MBW complexes. Nevertheless, how the complexes control the transcription of this defined set of target genes remains to be addressed. Functional analyses of the promoters were undertaken using both transient and stable expression assays, in moss (*P. patens*) protoplasts and transgenic plants (Arabidopsis), respectively (Fig. 3-9, Table 3-2). These approaches allowed the identification of a short promoter fragment (regulatory module) conferring a MBW-dependent expression in seed for each of the six genes (Fig. 3-9, Table 3-2).

Sequence analyses revealed that all the regulatory modules contain at least three types of conserved binding sites, a MYB core (*CNGTTR),* an *AC*-rich element ($[^{A}/_{C}]CC[^{A}/_{T}]A[^{A}/_{C}]$) and a bHLH binding sequence (*G*-box, $[^{C}/_{G}]ACGT[^{A}/_{G}]$ and/or E-box *CANNTG*), with the exception of *proLDOX-3*, for which no obvious bHLH binding site was found (Devic *et al.*, 1999; Debeaujon *et al.*, 2003; Patzlaff *et al.*, 2003; Hartmann *et al.*, 2005; Dare *et al.*, 2008; Prouse & Campbell, 2012; Thevenin *et al.*, 2012; Table 3-2). Interestingly, a similar finding has been reported recently for the promoter of *TT8*, another target of the MBW complexes

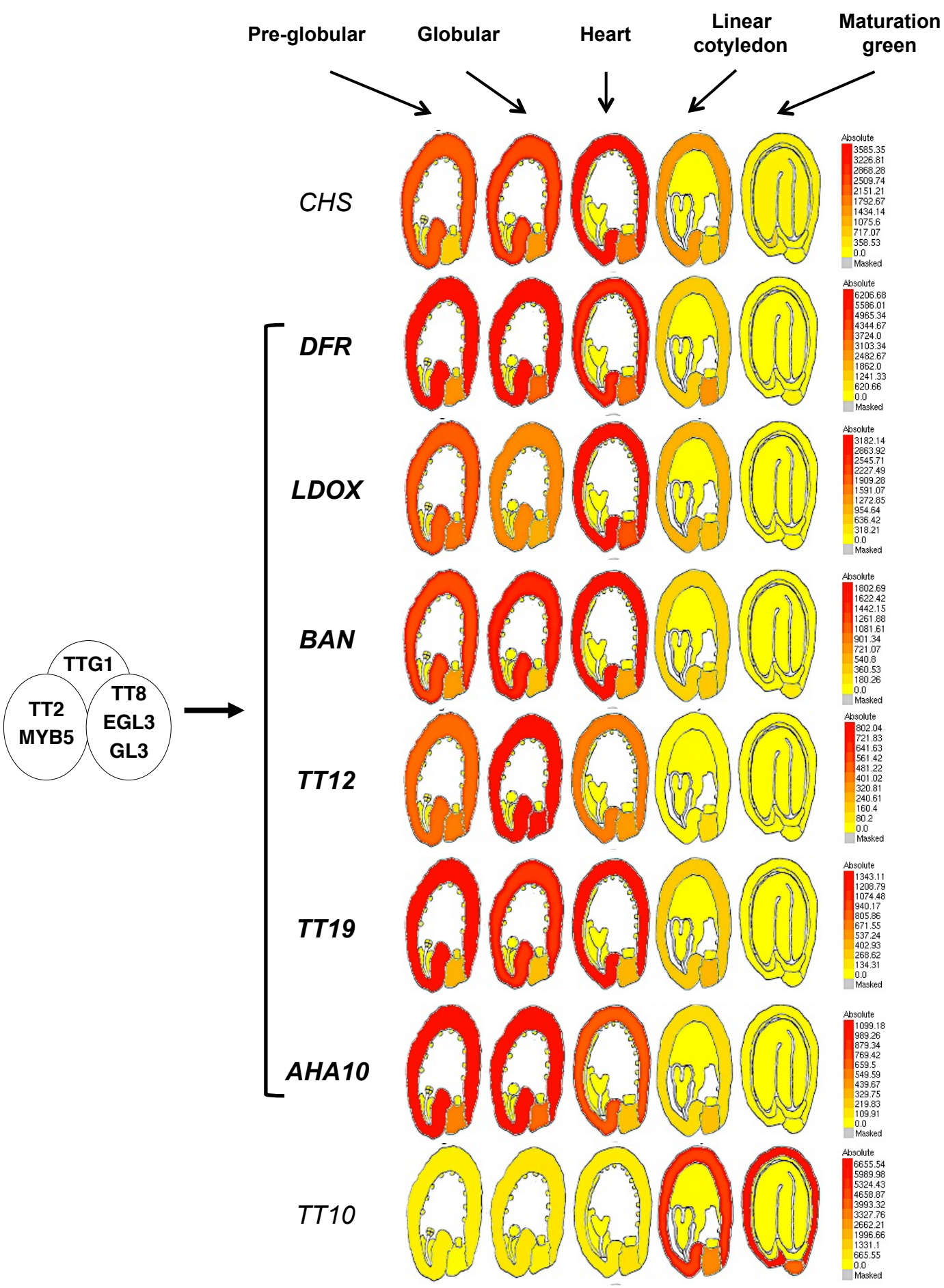


Figure 3-14. Steady state mRNA level of the structural genes involved in PA biosynthesis and directly regulated by the different MBW complexes in developing Arabidopsis seeds. CHS and TT10: negative controls. Data were generated by combining laser-capture micro dissection coupled with ATH1 GeneChip analysis (Le *et al*., 2010). Schematic representations are available at the following URL: http://bar.utoronto.ca/efp/cgi-bin/efpWeb.

(Xu *et al.*, 2013). Briefly, the promoter of *TT8* contains two domains through which the MBW complexes regulate its expression in seeds, with one of them not displaying any already described MYB and bHLH binding sites. Similarly, this study demonstrates that the *BAN* promoter also contains at least two MBW-regulated domains, with partially redundant functions. Indeed, although the minimal *BAN* promoter is totally inactive in *tt2*, *tt8* or *ttg1* (Fig. 3-2), a clear GUS activity was detected in the chalaza area of the *tt8* mutant when a 2-Kb promoter fragment was used (Debeaujon *et al.*, 2003).

3.5.3 Extending the LBG group to TT12, TT19 and AHA10, in Arabidopsis thaliana seed?

The results presented here showed that *TT12*, *TT19* and *AHA10* display the same patterns of mRNA accumulation, promoter activity and regulation by the TT2-TT8-TTG1 complex as the *LBGs* (Figs 3-1, 2, 3-14). Moreover, similar to the *LBGs*, *TT12*, *TT19* and *AHA10* have been shown to be direct targets of TTG1-dependent transcriptional regulatory complexes (Fig. 3-6). The *EBGs* and *LBGs* were initially categorized accordingly to their coordinate expression in response to environmental cues, such as light, at distinct developmental stages, in a species-dependent manner (Pelletier *et al.*, 1999; Lepiniec *et al.*, 2006). As such, the results presented above strongly support the idea that *TT12*, *TT19* and *AHA10* should be included in the *LBG* group, at least when referring to the Arabidopsis seed development. This assumption is reinforced by the analysis of the GUS pattern driven by the *DFR*, *LDOX*, *TT19* and *AHA10* promoters in seedlings, which is specific to the cotyledon margins, and strictly dependent on *TTG1* activity (Fig. 3-4).

This study also allowed us to refine our understanding of the transcriptional role played by the MBW complexes in the regulation of the flavonoid biosynthetic pathway in Arabidopsis seeds. Indeed, we found that the MBW complexes control specifically the transcription of the extended *LBG* group, but, more interestingly, we also found that this regulation occurs in a gene- and tissue-specific manner (Fig. 3-12). The TT2-TT8-TTG1 complex plays the major role in the development of seeds by regulating *DFR*, *LDOX*, *BAN*, *TT19*, *TT12* and *AHA10* in all the PA-accumulating cells (i.e. micropyle, chalaza and

			Conserved regulatory sequence					
Minimal promoter			MYB binding sites				bHLH binding sites	
			MYB-core		AC-element		E-box	G-box
Gene	Construct	size (bp)	*C(C/T)GTTA*	*C(A/C)GTTG*	*(A/C)CCAAC(C/G)*	*ACCTA(A/C)*	*CANNTG*	*CACGTG*
DFR	*proDFR-3*	302	0	1	1	1	0	2
LDOX	*proLDOX-3*	336	0	1	1	0	0	0
BAN	*proBAN236*	236	1	0	0	1	0	1
TT12	*proTT12-3*	464	2	0	0	1	2	1
TT19	*proTT19-2*	259	1	0	1	0	0	1
AHA10	*proAHA10-2*	328	2	2	0	1	3	0

Table 3-2. Summary of the R2R3-MYB and R/B-like bHLH *cis*-regulatory elements identified in the *Arabidopsis thaliana* minimal '*late biosynthetic gene*' ('*LBG*') promoters

endothelium), whereas the MYB5-TT8-TTG1 complex is only active in the endothelium where it regulates *DFR*, *LDOX* and *TT12* expression. Finally, the TT2-EGL3-TTG1 and TT2-GL3-TTG1 complexes regulate the expression of *DFR*, *LDOX*, *BAN* and *AHA10*, and *DFR*, respectively, in the chalaza.

Taken together, the results presented here highlight the complexity and robustness of the mechanisms involved. Several partially overlapping transcriptional regulations have been characterized involving functionally redundant DNA *cis*-elements and/or *trans*-acting factors. Further analyses of the structure-function relationships existing between the *DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10* promoters and the MBW complexes will be necessary to understand how they interact at the molecular level (e.g. 3'-end deletions and point mutation analyses of the promoters, chromatin immunoprecipitation experiments or electrophoresis mobility shift assay). This knowledge should help in identifying new direct target genes involved in flavonoid pathway that cannot be easily or directly detected by classical genetic approaches.

3.6 ACKNOWLEDGEMENTS

We thank Dr N. Terrier, Dr K. Hematy and Dr J.M. Routaboul for their useful comments. We thank the 'Plateforme de cytologie et imagerie végétale (PCIV)' from the Institut Jean-Pierre Bourgin (IJPB) for excellent technical support. The *egl3 tt8* and *egl3 gl3 tt8* mutant seeds were kindly provided by Dr Alan Lloyd. W.X.'s work was generously supported by the China Scolarship Council (CSC) of China.

CHAPTER IV

NEW INSIGHTS TOWARD THE TRANSCRIPTIONAL ENGINEERING OF PROANTHOCYANIDIN BIOSYNTHESIS

Article 3: New insights toward the transcriptional engineering of proanthocyanidin biosynthesis

Wenjia Xu[1, 2], Loïc Lepiniec[1, 2], Christian Dubos[1, 2]

Accepted for the publication in Plant Signaling & Behavior (2014, vol. 9)

[1] INRA, Institut Jean-Pierre Bourgin, Saclay Plant Sciences, RD10, F-78026 Versailles, France

[2] AgroParisTech, Institut Jean-Pierre Bourgin, Saclay Plant Sciences, RD10, F-78026 Versailles, France

Author for correspondence:

Christian Dubos

Tel: +33 499612613

Email: christian.dubos@supagro.inra.fr

4 Article 3

KEYWORDS: *Arabidopsis thaliana*, *Physcomitrella patens*, transient assay, overexpression, transcription factor, transparent testa, TT2, TT8, TTG1, cis-element

4.1 ABSTRACT

Flavonoids are secondary metabolites that play important roles throughout the plant life cycle and have potential human health beneficial properties. Flavonols, anthocyanins and proanthocyanidins (PAs or condensed tannins) are the three main class of flavonoids found in Arabidopsis thaliana. We have previously shown that PA biosynthesis (occurring exclusively in seeds) involves the transcriptional activity of four different ternary protein complexes composed of different R2R3-MYB and bHLH factors together with TRANSPARENT TESTA GLABRA 1 (TTG1), a WD repeat containing protein. We have also identified their direct targets, the late biosynthetic genes. In this study, we have further investigated the transcriptional capacity of the MBW complexes through transactivation assays in moss protoplast and overexpression in Arabidopsis siliques. Results provide new information for biotechnological engineering of PA biosynthesis, as well as new insights into the elucidation of the mechanisms that govern the interactions between MBW complexes and the DNA motifs they can target.

4.2 INTRODUCTION

Flavonoids are one of the largest classes of plant secondary metabolites, which include flavonols, anthocyanins and proanthocyanidins (PAs or condensed tannins). Flavonoids play important roles throughout the plant life cycle and have potential human health beneficial properties (Lepiniec *et al.*, 2006; Petroni & Tonelli, 2011). In *Arabidopsis thaliana*, PAs are specifically synthesized and accumulated in the inner integument of the seed coat, in which they are thought to play an important role in protecting the embryo against biotic and abiotic stresses (Debeaujon *et al.*, 2000) and conferring brown colour to mature seeds once oxidized

(Pourcel *et al.*, 2005). Arabidopsis mutants impaired in flavonoid accumulation have been identified through visual screenings for altered seed pigmentation, which results in the *transparent testa* (*tt*) phenotype (Koornneef, 1990). Most of these mutants have been characterized at the molecular level allowing to decipher the core of this complex biosynthetic pathway (Routaboul *et al.*, 2012). In Arabidopsis, it is composed of at least two distinctly co-regulated groups of genes, namely the early and late biosynthetic genes (*EBGs* and *LBGs*) (Koornneef, 1990). The *EBGs* are involved in biosynthesis of the common precursor of the three classes of flavonoids (*i.e.* dihydroflavonols), whereas the *LBGs* are specific to anthocyanin and PA biosynthesis (Fig. 4-1). Therefore, *LBGs* include *DFR* (*dihydroflavonol-4-reductase*), *LDOX/ANS* (*leucoanthocyanidin dioxygenase/ anthocyanidin synthase*), *BAN/ANR* (*BANYULS /anthocyanidin reductase*), *TT12* (*MATE transporter*), *TT19/GST26/GSTF12* (*glutathione-S-transferase*), and *AHA10* (*H+-ATPase isoform 10*). However, TT10/LAC15 (laccase 15) and TT15/UGT80B1 (UDP-glucose:sterol-glucosyltransferase) which are both critical for PA biosynthesis (Pourcel *et al.*, 2005; DeBolt *et al.*, 2009) could not been classified as *EBG* nor as *LBG*.

We have recently demonstrated that PA biosynthesis relies on the transcriptional activity of R2R3-MYB (TT2/MYB123 and MYB5) and R/B-like basic helix-loop-helix (bHLH: GL3/bHLH001, EGL3/bHLH002 and TT8/bHLH042) proteins which form four different ternary complexes (MBW) with TRANSPARENT TESTA GLABRA1 (TTG1), a WD repeat-containing protein(Xu *et al.*, 2013a). These four MBW complexes (namely TT2-TT8-TTG1, MYB5-TT8-TTG1, TT2-EGL3-TTG1 and TT2-GL3-TTG1) directly regulate the expression of the *LBGs*, in a tissue-specific manner (Xu *et al.*, 2013a). The TT2-TT8-TTG1 complex is the main complex in regulating PA biosynthesis (Xu *et al.*, 2013a). Similar MBW protein complexes are involved in the transcriptional regulation of anthocyanin biosynthesis, in which PRODUCTION OF ANTHOCYANIN PIGMENT 1 (PAP1/MYB75) and 2 (PAP2/MYB90) ensure most of the R2R3-MYB function (Appelhagen *et al.*, 2011). MBW complexes regulating PA and anthocyanin biosynthesis have been characterized in various plant species such as maize (*Zea mays*), petunia (*Petunia hybrida*), grapes (*Vitis vinifera*), apples (*Malus domestica*) or strawberry (*Fragaria x*

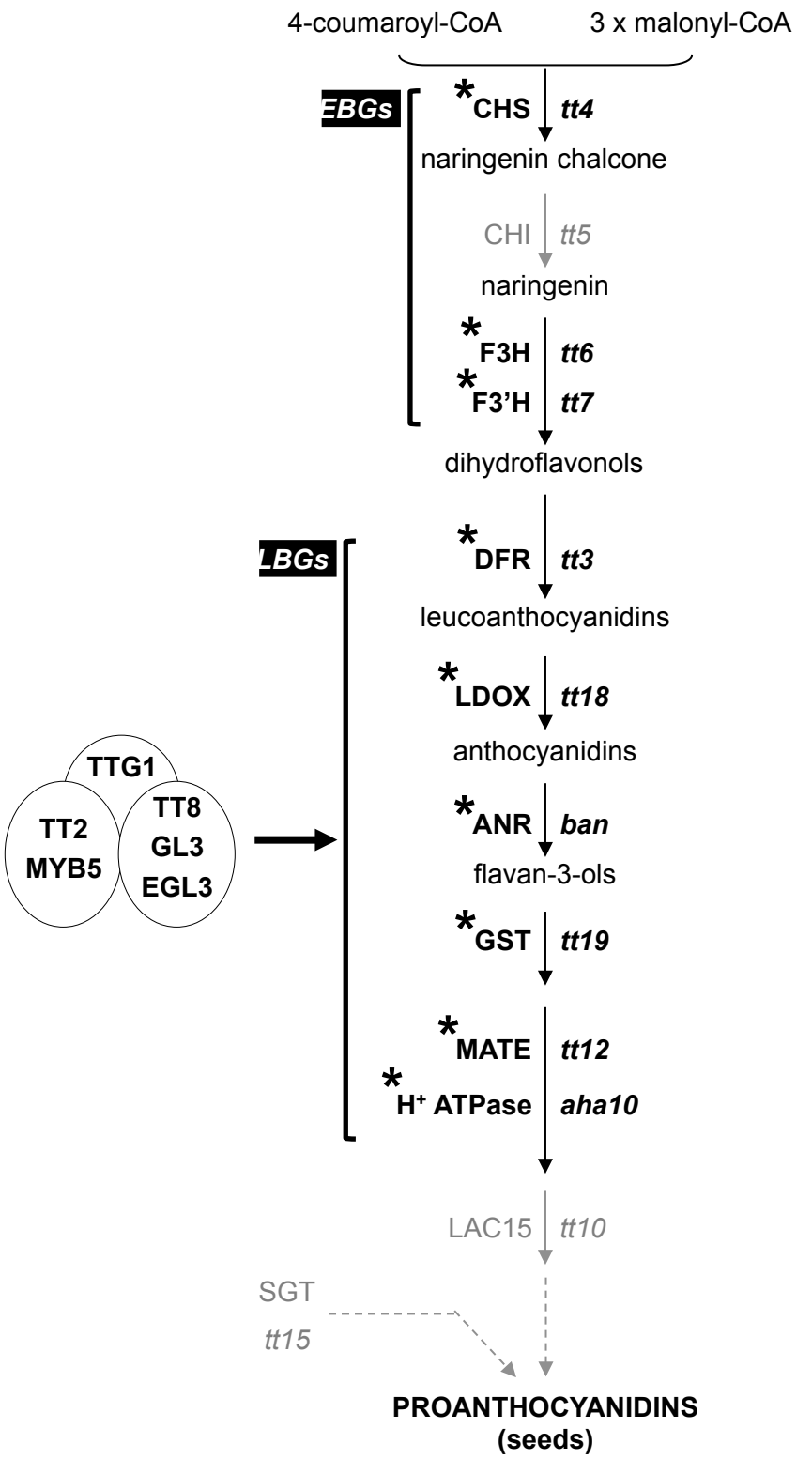


Figure 4-1. Schematic representation of the *Arabidopsis thaliana* proanthocyanidin biosynthetic pathway. Four different ternary protein complexes composed of R2R3-MYB and R/B-like bHLH transcription factors together with the WD repeat-containing protein TTG1 directly regulate the expression of the *LBGs* (late biosynthetic genes) tissue-specific manner. These (MBW) complexes are TT2-TT8-TTG1, MYB5-TT8-TTG1, TT2-EGL3-TTG1 and TT2-GL3-TTG1. *, genes whose expression is directly induced in Arabidopsis siliques when *TTG1* is overexpressed. In grey, genes whose expression is not affected in Arabidopsis siliques when *TTG1* is overexpressed. The names of the structural proteins are indicated in capital letters and the corresponding mutants in lower-case italics. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavonol 3-hydroxylase; F3'H, flavonol 3'-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol-4-reductase; LDOX, leucoanthocyanidin dioxygenase (also called ANS: anthocyanidin synthase); ANR, anthocyanidin reductase; MATE, multidrug and toxic efflux transporter; GST, glutathione-S-transferase; SGT, UDP-glucose:sterol-glucosyltransferase; LAC15, laccase 15. Arrows indicate the different steps leading to the formation of flavonoids in Arabidopsis; dashed lines indicate multiple steps; circles correspond to transcription factors; EBGs, early biosynthetic genes; LBGs, late biosynthetic genes.

ananassa) fruits (Hichri *et al.*, 2011; Petroni & Tonelli, 2011; Schaart *et al.*, 2013).

PAP1 overexpression in Arabidopsis as observed in the *PAP1-D* activation tagging mutant led to both over- and ectopic accumulation of anthocyanin in various plant organs, including leaves, stems, flowers and silique valves (Borevitz *et al.*, 2000). Similar observations were made when the Arabidopsis *PAP1* gene was overexpressed in tobacco (*Nicotiana tabacum cv xanthi*) (Borevitz *et al.*, 2000). At the molecular level, *PAP1* overexpression was sufficient to transcriptionally induce the flavonoid biosynthetic pathway leading to anthocyanin accumulation, from biosynthesis to storage, which includes the *EBGs* and most of the *LBGs* (Fig. 4-1) (Borevitz *et al.*, 2000; Shi & Xie, 2010; Shi & Xie, 2011). In contrast, the ectopic expression of *TT2* can trigger PA accumulation in Arabidopsis seed coat, but not in vegetative parts of the plant (Nesi *et al.*, 2001). Nevertheless, the expression of *BAN* is sufficient, in addition to *PAP1*, for accumulating PA in tobacco and *Medicago truncatula* leaves (Xie *et al.*, 2006). Similarly, PAs accumulation has been observed is some vegetative tissues when *TT2* was overexpressed in *PAP1-D* plants (Sharma & Dixon, 2005). Interestingly, the overexpression of the Arabidopsis *TT2* gene was sufficient to trigger PA accumulation in hairy roots of *M. truncatula* (Pang *et al.*, 2008). Taken together, these results demonstrated that PAP1 and TT2 have different targets *in planta*. In this regard, a recent study has shown that DNA binding differences between TT2 and the PAP proteins *in planta* can be explained by a few amino acids present in the R2 and R3 domains of these TFs (Heppel *et al.*, 2013). Altogether these results also showed that TT2 (and PAP1) could activate different targets depending on the "cellular" context (e.g. Arabidopsis seed coat, vegetative tissues, cell cultures or *M. truncatula* hairy roots).

In order to better understand the underlying molecular mechanisms, some transactivation assays were carried out in moss (*Physcomitrella patens*) protoplasts (Thevenin *et al.*, 2012). For this purpose the promoters of 12 PA biosynthetic genes (Fig. 4-1) were fused to the green fluorescent protein (GFP) reporter gene, and then co-transfected in moss protoplasts alone or in combination with one, two or three of the studied transcriptional regulators (*i.e. TT2*, *TT8* and *TTG1*). These promoter fragments have been already shown to be active in PA and anthocyanin accumulating cells (Xu *et al.*, 2013a). Transactivation

activities were monitored by GFP fluorescence (Thevenin *et al.*, 2012).

4.3 RESULTS

An induction of GFP fluorescence (*i.e.* above the fluorescence of the promoter alone) was observed for all the tested promoters but *TT10/LAC15,* when TT2, TT8 and TTG1 were expressed simultaneously (Fig. 4-2). A strong fluorescence signal was detected for the *CHS* (*chalcone synthase*), *CHI* (*chalcone isomerase*), *F3H* (*flavonol 3-hydroxylase*), *DFR*, *LDOX*, *BAN*, *TT15/UGT80B1*, *TT19* and *AHA10* promoters, whereas a slight but significant induction was observed for the promoters of *F3'H* (flavonol 3'-hydroxylase) and *TT12*. Interestingly, this experiment revealed that the TT2-TT8-TTG1 complex could activate the expression of both the *EBGs* and *TT15,* in addition to the *LBGs*, at least in moss protoplasts. These data are consistent with previous experiments carried out in grapes cells in which the authors have shown that the overexpression of the grapes *VvMYBPA1* (grapes homologue of TT2) alone or in combination with *EGL3*, *TT8* or *VvMYC1* (grapes homologue of TT8) induces *VvCHI* (*EBG*) or *VvANR* (*LBG*) promoter activities (Hichri *et al.*, 2010).

A slight increase in GFP intensity was also observed with the *CHI*, *F3H*, *F3'H*, *DFR*, *TT12*, *TT15*, *TT19* and *AHA10* promoters, when TT2 was assayed with TT8 but without TTG1 (Fig. 4-2). This later finding was consistent with yeast two hybrid analyses in which TT2 and TT8 were able to bind *BAN* promoter when expressed simultaneously (Baudry *et al.*, 2004). Nevertheless, *in planta* and in Arabidopsis protoplasts, the presence of TTG1 is required for the activation of the *BAN* promoter (Debeaujon *et al.*, 2003; Baudry *et al.*, 2004). Finally, the lack of induction for the promoter of *TT10* was fully consistent with previous finding showing that *TT10* is not a target of the MBW complexes *in planta* (Xu *et al.*, 2013a).

The ability of the MBW complexes to induce the expression of the *EBGs* and *TT15* was then investigated *in planta* by using transgenic plants that constitutively express the inducible TTG1:GR chimeric protein (a translational fusion between TTG1 and the glucocorticoid receptor (Baudry *et al.*, 2004; Xu *et al.*, 2013a). In these plants, the translocation of the TTG1:GR chimeric protein into the nucleus occurs only in the presence of dexamethasone (DEX).

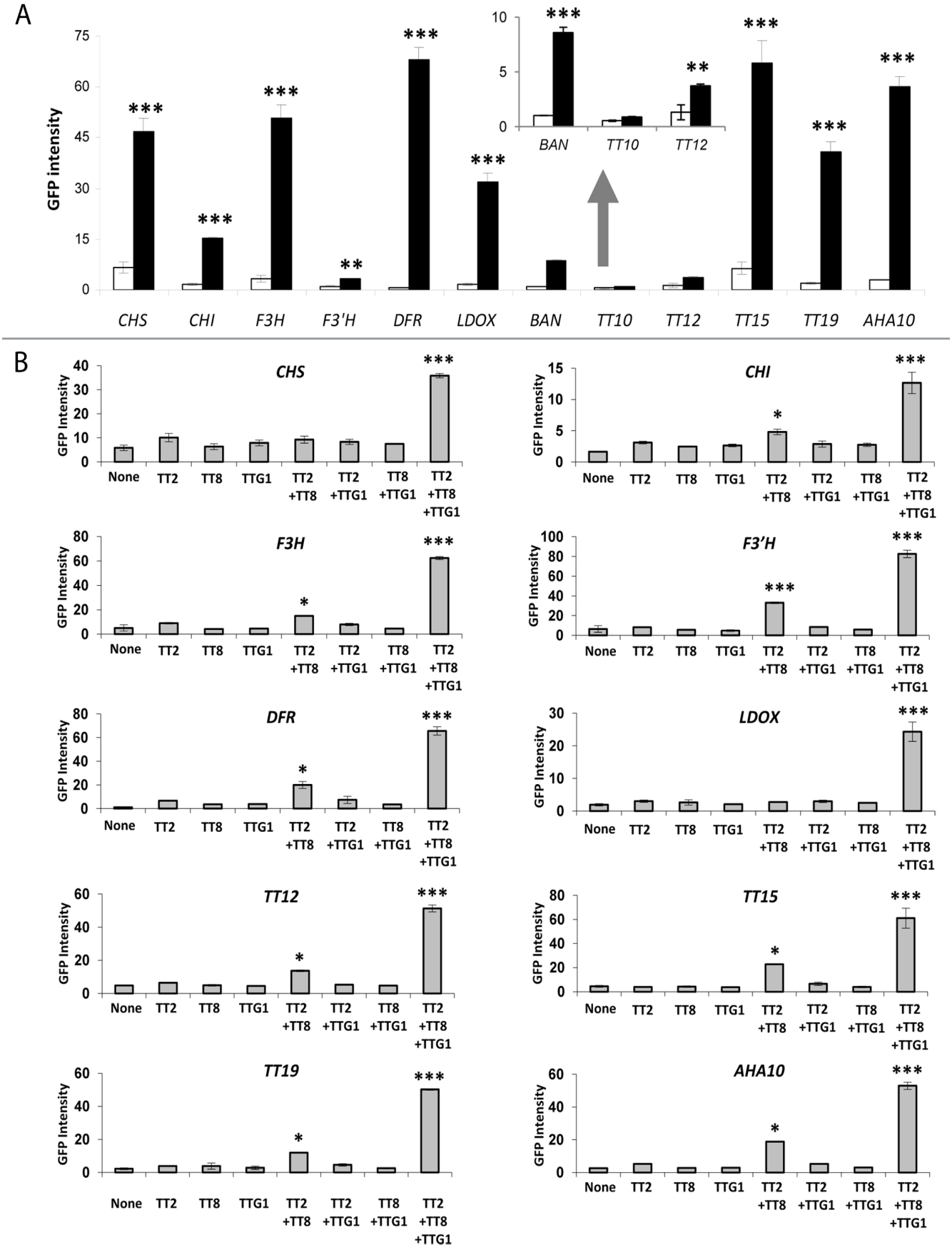


Figure 4-2. Transactivation assays in moss (*Physcomitrella patens*) protoplasts. (**A**) The TT2-TT8-TTG1 protein complex activate the transcription from the promoters of the *EBGs* (*CHS*, *CHI*, *F3H* and *F3'H*), LBGs (*DFR*, *LDOX*, *BAN*, *TT12*, *TT19* and *AHA10*) and *TT15*. Transactivation activities were measured by GFP intensity of arbitrary units of protoplasts co-transfected with the studied promoters in combination with the TT2-TT8-TTG1 protein complex. (**B**) TT2 and TT8 are sufficient to activate the transcription from all the studied promoters but *CHS* and *LDOX*. In this analysis, *F3'H* and *TT12* were fused to the 35S cauliflower mosaic virus (CaMV) minimal promoter sequence as described previously.8 Student *t*-test significant difference: *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$. Error bars, ±SE from three biological repetitions. none, protoplasts transformed with the assayed promoters alone.

Quantitative RT-PCR experiments revealed that *CHS*, *F3H* and *F3'H* mRNA steady state levels were directly increased in response to TTG1 induction (Fig. 4-3), as this increase was observed when 4-day-old siliques were treated with DEX alone or in combination with cycloheximide (DEX/CHX), an inhibitor of protein translation. In contrast to what has been observed in transient assays carried out in moss protoplasts *CHI* and *TT15* mRNA levels were unaffected after TTG1:GR inductions. We deduced that *CHI* gene activation in Arabidopsis might require other factors in addition to *TTG1*. This experiment suggested that the induction of *CHI* expression may be one of the limiting steps for the enhancement of PA accumulation, at least in Arabidopsis seed. This limiting step could be species specific, as the ectopic expression of the grape *TT2* homologues *VvMYBPA1* (seed specific) and *VvMYBPA2* (expressed in exocarp of young berries) in grapevine hairy roots induces both PA accumulation and *EBGs* (*VvCHS*, *VvCHI*, *VvF3H* and *VvF3'H*) expression (Terrier *et al.*, 2009).

The molecular mechanisms that control *TT15* expression are still not understood. Our results suggested that *TT15* expression that is sensitive to CHX treatment could be under the control of a transcriptional repressor.

4.4 DISCUSSION

Transcriptional modulation of gene expression relies on specific interactions between DNA motifs (*cis*-regulatory elements generally located in their promoter) and transcriptional regulators (*trans* factors). In this regard, how TTG1-dependent complexes regulate the spatio-temporal pattern of expression of the PA and anthocyanin biosynthetic genes is still an open question. Nevertheless, because TTG1 is not absolutely required in all biological systems used for inducing *BAN* expression, it suggested that TTG1 role might be stabilised the interaction between TT2 and TT8 or to counteract the activity of a negative regulator of this interaction.

Some *cis*-elements gathered from the characterization of *BAN* and *TT8* promoters led to the identification of three types of regulatory sequences that play a critical role in the regulation of their expression (Thevenin *et al.*, 2012; Xu *et al.*, 2013b). Two of these DNA

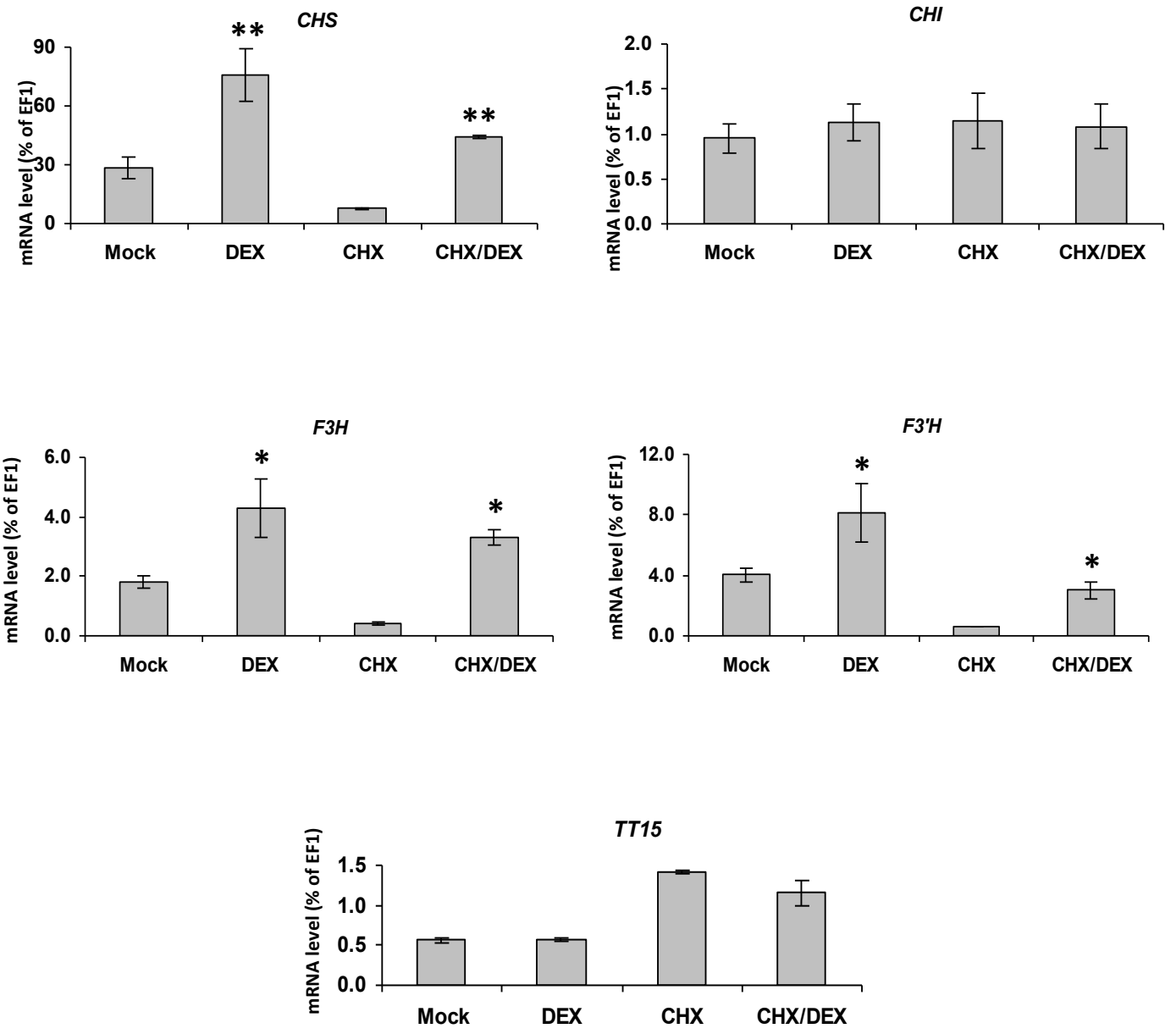


Figure 4-3. *CHS*, *F3H*, *F3'H* and *TT15* expression in Arabidopsis siliques is directly induced by the overexpression of *TTG1*. Experiments were carried out using plants that constitutively express the inducible TTG1:GR chimeric protein (a translational fusion between TTG1 and the glucocorticoid receptor). The steady state level of *CHS*, *CHI*, *F3H*, *F3'H* and *TT15* mRNA was measured by quantitative RT-PCR on 4-day-old siliques. Values are expressed as a percentage of the constitutively expressed *EF1αA4* gene. Student *t*-test significant difference: *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$. Error bars, ±SE from three biological repetitions. Mock, buffer; DEX, dexamethasone; CHX, cycloheximide.

motifs belong to the MYB-core (*CNGTTR*) and E-box (*CANNTG*) class of *cis*-regulatory sequences that are targeted by the R2R3-MYB (*i.e.* TT2 and MYB5) and bHLH (*i.e.* GL3, EGL3 and TT8) transcription factors, respectively. The third class corresponds to the *AC*-rich *cis*-regulatory sequences (also called *AC*-elements) and was shown to be necessary for *TT8* expression in both seeds and vegetative tissues.(Xu *et al.*, 2013b) This later group is well known to be the target of R2R3-MYBs (Prouse & Campbell, 2012; Romano *et al.*, 2012; Prouse & Campbell, 2013); however, to date it is still unclear which transcription factors are actually regulating PA biosynthesis through these *AC*-elements (Xu *et al.*, 2013b).

MYB-core, *AC*-element and E-box motifs are all present in the minimal promoter (upstream from the translational start site) of all the *LBGs* but *LDOX*, which has no E-box (Table 4-1) (Xu *et al.*, 2013a). The absence of an E-box in the *LDOX* promoter suggests that EGL3 and TT8 may recognise a *cis*-regulatory sequence that has not yet been characterized or bind indirectly to the promoter through the interaction with another transcription factor. These three regulatory sequences are also found in the promoter of the other studied genes albeit with some restrictions. *CHI* and *F3'H* promoters do not display any MYB-core or *AC*-rich *cis*-regulatory motifs (both motifs being targeted by R2R3-MYBs), respectively, whereas none of these two elements is present on the promoter of *TT10* (Table 4-1). Altogether, sequence analyses indicate that all the *EBGs* as well as *TT15* possess at least one R2R3-MYB and one bHLH binding site in their promoter. This finding could explain why *trans*-activations were observed in moss protoplasts in the presence of TT2, TT8 and TTG1 (Fig. 4-2). Similar results have been obtained in Arabidopsis protoplasts showing that two maize TFs orthologous of TT2 and TT8 (namely C1 and Sn, respectively) are able to trans-activate *CHS* promoter through two *cis*-regulatory sequences that correspond to an *AC*-rich motif (named MYB-recognition element or MRE) and an E-box (named R response element or RRE) (Hartmann *et al.*, 2005). Similarly, trans-activation of *BAN* promoter by C1 and Sn has been observed in Arabidopsis protoplasts (Baudry *et al.*, 2004). However, the analysis of siliques overexpressing the TTG1:GR chimeric protein highlights that transcriptional regulation of PA biosynthesis in Arabidopsis cannot be resumed to the sole interaction between the MBW complexes and some *cis*-regulatory elements, as *CHI* and

	Promoters		Conserved regulatory sequence		
	Gene	size (bp)	MYB-core	AC-element	E-box
			CNGTTR	*(A/C)CC(T/A)A(A/C)(C/G)*	*CANNTG*
EBGs	*CHS*	1211	1	3	3
	CHI	1187	0	2	3
	F3H	901	1	2	3
	F3'H	1472	1	0	6
LBGs	*DFR*	302	1	2	2
	LDOX	336	1	1	0
	BAN	236	1	1	1
	TT12	464	2	1	3
	TT19	259	1	1	1
	AHA10	328	4	1	3
	TT10	1473	0	0	1
	TT15	2819	2	2	4

Table 4-1. Summary of the R2R3-MYB and R/B-like bHLH *cis*-regulatory elements identified in the promoters of the genes involved in proanthocyanidin biosynthesis in *Arabidopsis thaliana* that were analysed in this study.

TT15 expression was unaffected upon DEX or DEX/CHX induction (Fig. 4-3).

The ectopic expression of TTG1-dependent complexes confirms their ability to activate the expression of a subset of genes involved in flavonoid biosynthesis that they do not usually target (Xu *et al.*, 2013a). Interestingly, depending on the biological system used, moss protoplasts or siliques, expression of the *EBGs* and *TT15* or the *EBGs* without *CHI* are induced, respectively. This later result indicates that the limiting step in enhancing PA biosynthesis in seed resides at least partly in the lack of induction of *CHI* expression in response to the overexpression of *TTG1*. However, *CHI* expression is certainly not the sole step that limits the ectopic accumulation of PAs throughout the plant body and that additional factors remain to be characterized. It could be for instance hypothesized that in vegetative tissues specific co-activators or repressors are absent or present, respectively (Li, 2014). In order to identify new molecular actors involved in controlling PA accumulation pattern one could envisage a genetic screen aiming at identifying *TT2* overexpressing plants that ectopically accumulate PAs. The data presented herein also suggests that TTG1 overexpression is sufficient to increase the stability of the interaction occurring between the R2R3-MYB and the bHLH transcription factors that in turn enhances the expression of the targeted genes. Altogether this study provides some new information for biotechnological engineering of PAs biosynthesis. Moreover, it highlights the limits of the different methods based on the overexpression of transcriptional regulators (*i.e.* transactivation assays and transgenic plants) in identifying their direct physiological targets.

4.5 ACKNOWLEDGEMENTS

W.X.'s work was generously supported by the China Scholarship Council (CSC) of China.

CHAPTER V

CONCLUSIONS, DISCUSSIONS AND PERSPECTIVES

5 CONCLUSIONS, DISCUSSIONS and PERSPECTIVES

5.1 CONCLUSIONS

5.1.1 Transcriptional regulation of TT8 expression

The regulation of PA and anthocyanin biosynthesis by the MBW complexes is currently relatively well described and understood, however, the precise nature of the molecular mechanisms that govern the expression of these anthocyanin and PA biosynthesis regulators was still elusive. One of my PhD obejectives was to elucidate the regulation of *TT8* expression because of the key role of TT8 in the regulation of both anthocyanin and PA biosynthesis in different tissues, as we have seen throughout this manuscript. Several years ago it has been demonstrated by our laboratory that TT8 can control its own expression through a positive feedback regulation loop (Baudry *et al*., 2006). This discovery was an important step in the understanding of the transcriptional regulation of flavonoid biosynthesis in Arabidopsis. Within the frame of my PhD, we would like to decipher transcriptional regulation of *TT8* expression through further studies using molecular, genetic and biochemical approaches. During the course of my PhD, I have shown through an in-depth analysis of *TT8* expression suggesting that *TT8* expression is directly regulated by several MBW complexes. Moreover the functional dissection of the *TT8* promoter revealed its modular structures. Two modules were found to specifically drive *TT8* promoter activity in PA- and anthocyanin-accumulating cells, by differentially integrating the signals issued from different regulators, in a spatiotemporal manner and a third module is responsible for the strength of the promoter (Fig. 5-1). Interestingly, this regulation involves at least six different MBW complexes involving four R2R3-MYBs (PAP1, PAP2, TT2 and MYB5), three bHLHs (TT8, EGL3 and GL3) and TTG1, which are interacting with the different modules in a tissue-specific manner (Fig. 5-1).

The work carried out during my PhD also confirmed that TTG2, TT1 and TT16 are required for proper activity of the *TT8* promoter in the endothelium and micropyle (Fig. 5-1). These interesting results provide clues for next steps for further analysis of the regulation of *TT8* expression at the developmental level. Further analysis needs to be performed in order to

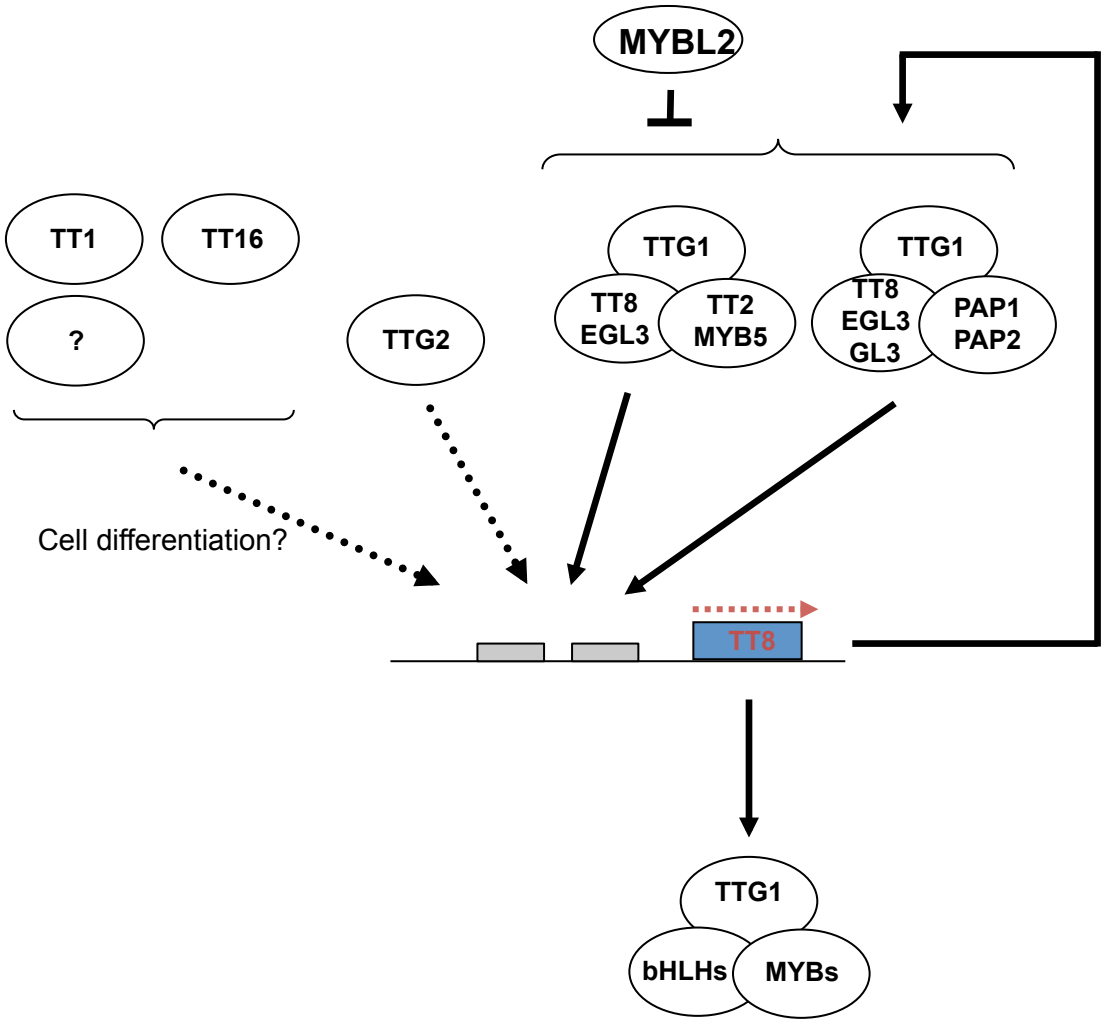


Figure 5-1. Schematic representation of the regulation of *TT8* expression in *Arabidopsis thaliana.* Solid arrows indicate the direct regulation; dashed lines indicate indirect regulation or multiple steps; circles correspond to transcription factors. circles correspond to transcription factors.

determine if the effects of these three regulators on *TT8* expression are direct or not.

5.1.2 Comprehensive analyses of MYB-bHLH-WDR complexes and their targets in Arabidopsis

In Arabidopsis, PAs are accumulated specifically in the innermost cell layer of the seed coat (i.e. endothelium, chalaza and micropyle) (Debeaujon *et al*., 2003). This specific pattern is conferred by the collective action of TT2 (MYB), TT8 (bHLH) and TTG1 (WDR) in regulation of the expression of the PA structural genes (Lepiniec *et al*., 2006). Specific PA regulators, TT8 and TT2, were characterized and identified successively by my laboratory in 2000 and 2001, respectively (Nesi *et al*., 2000; 2001). In 2004, my laboratory has demonstrated that this ternary complex, composed of TT2, TT8 and TTG1, is able to bind the *BAN* promoter to activate transcription of the *BAN* gene in *planta* (Baudry *et al*., 2004). Within the framework of my PhD thesis, one of the objectives was to develop the study of TT2-TT8-TTG1 complex regulation via the comprehensive analysis of target gene expression. The ultimate goal was to obtain an integrated model of the network structure and regulations involved in the control of PA biosynthesis, allowing the modification of flavonoid nature and content in various plant organs. During the course of my PhD I have shown that the specific accumulation of PAs in the innermost cell layer of the seed coat (i.e. endothelium, chalaza and micropyle), involves at least four MBW complexes, namely TT2-TT8/GL3/EGL3-TTG1, and MYB5-TT8-TTG1, with partially overlapping functions (Fig. 5-2). The complex made of TT2-TT8-TTG1 plays a major role controlling the expression the *LBGs* (*DFR*, *LDOX, TT19, TT12, AHA10* and *BAN*). The MYB5-TT8-TTG1 complex is active in the endothelium regulating *DFR*, *LDOX* and *TT12* expression, whereas the TT2-EGL3-TTG1 and TT2-GL3-TTG1 complexes regulate the expression of *DFR*, *LDOX*, *BAN*, *AHA10* and *DFR* in the chalaza, respectively (Fig. 5-2). Similarly, I have also confirmed that anthocyanin biosynthesis is also regulated by different sets of MBW complexes in vegetative tissues involving PAP1, PAP2, GL3, EGL3, TT8 and TTG1 (Zhang *et al.*, 2003; Baudry *et al.*, 2006; Gonzalez *et al.*, 2008; Feyissa *et al.*, 2009; Appelhagen *et al.*, 2011a). I have also confirmed that among the four MYB paralogs, PAP1 is dominant, at least in seedlings, as a loss of

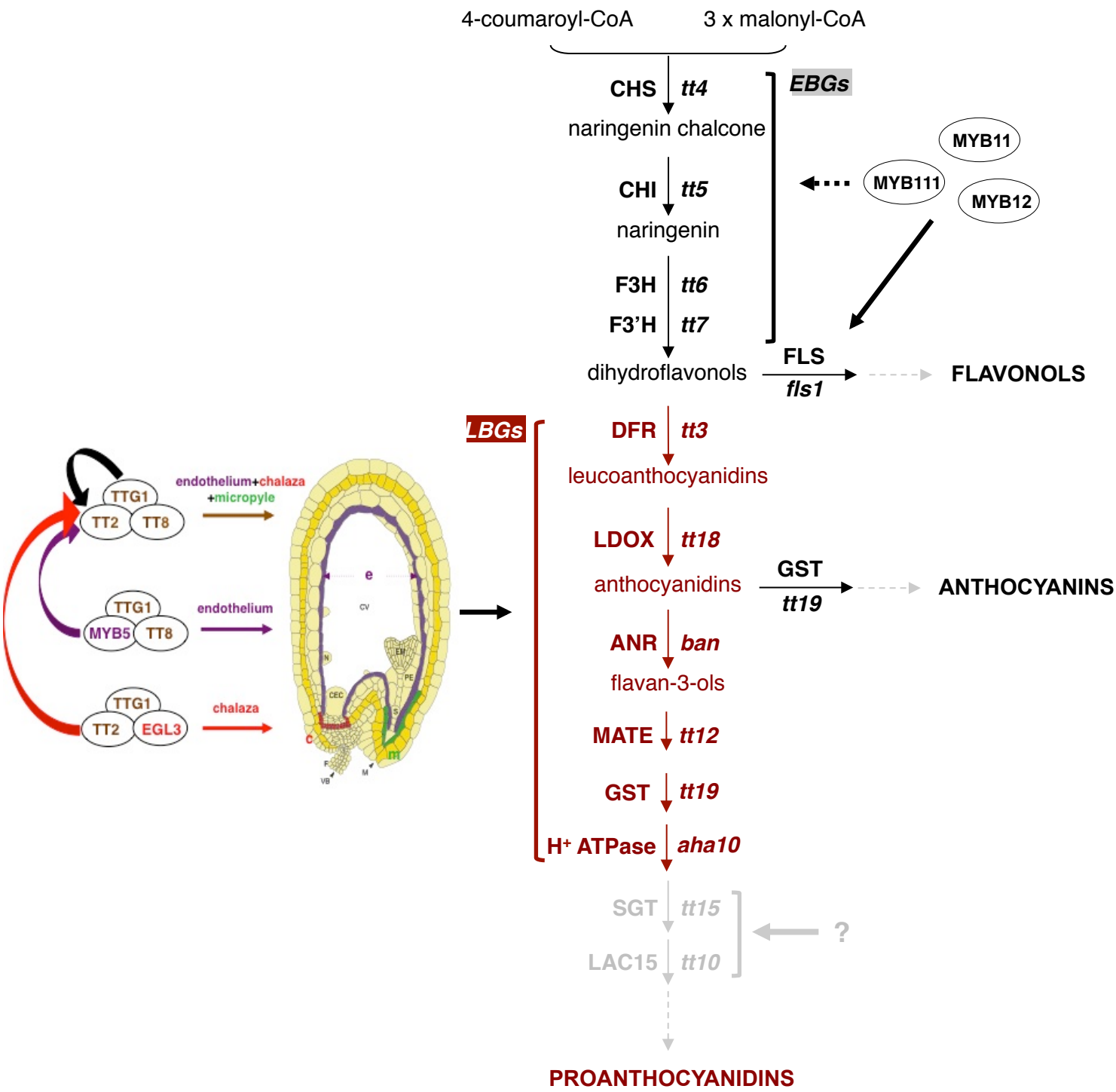


Figure 5-2. Schematic representation of the flavonoid biosynthetic pathway in *Arabidopsis thaliana*. The names of the structural proteins are indicated in capital letters and the corresponding mutants in lower-case italics. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavonol 3-hydroxylase; F3'H, flavonol 3'-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol-4-reductase; LDOX, leucoanthocyanidin dioxygenase; ANR, anthocyanidin reductase; MATE, multidrug and toxic efflux transporter; GST, glutathione-S-transferase; SGT, UDP-glucose:sterol-glucosyltransferase; LAC15, laccase 15. Arrows indicate the different steps leading to the formation of flavonoids in Arabidopsis; dashed lines indicate multiple steps; circles correspond to transcription factors; *EBG*, early biosynthetic gene; *LBG*, late biosynthetic gene. Arrows indicate the different steps leading to the formation and accumulation of flavonoids in Arabidopsis; dashed lines indicate multiple steps; circles correspond to transcription factors. In purple, part of the pathway leading to proanthocyanidin (PA) (condensed tannins) biosynthesis analysed in the present study, and directly regulated by the activity of different MBW complexes in seeds.

PAP1 function causes anthocyanin deficiencies and inhibits *TT8* promoter activity in young seedlings whereas the three *bHLHs* (*TT8*, *EGL3,* and *GL3*) display partially overlapping expression and redundant functions.

Various studies have indicated that different sets of MBW complexes regulate the biosynthesis of PAs and anthocyanins, by modulating the expression of particular genes specifically involved in these biosynthetic processes, which relies on specific interactions between transcriptional regulators and DNA *cis*-regulatory motifs. The MYB- and bHLH-binding sites gathered from comprehensive analyses of the promoters of two flavonoid genes, namely *BAN* and *TT8*, involved three types of conserved *cis*-regulatory sequences, MYB-core and "*AC*-rich" elements and bHLH binding sequences (Thevenin *et al.*, 2012; Xu *et al.*, 2013). Indeed, the analysis of *TT8* promoter led to the identification of two binding sites (MYB-core *5'-CAGTTA-3'* and G-box *5'-CACGTC-3'*) that display strong similarities with the *cis*-regulatory elements of the *BAN* promoter (MYB-core *5'-CTGTTG-3'*, and a G-box *5'-CACGTG-3'*) shown to be the target of the TT2-TT8-TTG1 complex (Debeaujon *et al.*, 2003; Thevenin *et al.*, 2012; Xu *et al.*, 2013). In addition, two *AC*-rich MYB binding sites were also found to be functionally redundant and necessary for the tissue-specific activity of the *TT8* promoter, but the TFs involved remain to be discovered (Xu *et al.*, 2013). Consistent with these findings, at least one element of each type (i.e. Myb-core, AC-rich and E-box) has been characterized in functional promoter fragments of four other *LBG*s (*DFR*, *TT19*, *TT12* and *AHA10*) able to trigger MBW-dependent expression in the seed coat (Xu *et al.*, 2014a). The two types of MYB binding site have also been found in the *LDOX* promoter, but no obvious bHLH binding sites, suggesting that other DNA motifs could be involved (directly or through other TFs).

The down-regulations of TTG1, as well as MYB homologs (i.e. PAP1 to PAP4) by RNAi or of the bHLHs in the triple mutant *gl3egl3tt8* inhibit only the expression of the *LBGs* (Baudry *et al.*, 2004; Gonzalez *et al.*, 2008; Xu *et al.*, 2014a). These results suggest that *in planta*, the activation of the *EBGs* does not directly involve the MBW complexes. Nevertheless, by the use of ectopic expressions and/or expression in heterologous systems (yeast or moss protoplasts) I have shown that the MBW complexes can activate the entire flavonoid pathway, including the EBGs, providing interesting information for

biotechnological application (Xu *et al.*, 2014b).

5.2 Discussions and perspectives

Flavonoid pathway and genes are well conserved in higher plants and mainly induced in response to biotic or abiotic stresses and/or expressed in specific tissues. MBW complexes are involved in both types of developmental and environmental regulations at the transcriptional level, mainly through the activation of flavonoid *LBG* expression (Fig. 5-3). In Arabidopsis as well as in other higher plants, several partially overlapping regulations have been characterized involving functionally redundant MBW partners and DNA-binding *cis*-elements. The complexity and robustness of these transcriptional regulatory networks seem astonishing and could simply reflect evolution footprints, gene duplications and neofunctionalizations. Nevertheless, they clearly allow cell specific accumulation of various flavonoids to fulfil different functions. For instance, a developmental regulation involving a positive feedback loop allows a very strong and specific expression of PA genes in a single cell layer of the seed coat. In the contrary, environmental regulations involving different negative regulatory feedbacks allow fine-tuned and reversible expression of flavonoid genes and flavonoid accumulation.

In depth functional analyses of the promoter of *TT8* and *BAN* demonstrated that the TTG1/TT8/TT2 complex needs both a MYB-core element and a *G*-box to activate the target promoters. Nevertheless, additional TFs able to interact with *AC*-rich MYB binding sites remain to be discovered. In this regard, some of the flavonoid genes, the expression of which seem to be co-regulated, lack known binding sites for the MBW partners. Therefore, some works are still necessary to identify the *cis*-regulatory motifs involved by the MBW complexes. These results emphasized that the precise role of the three partners and functioning of the MBW complexes are not yet fully understood (e.g. stereochemistry of the complexes, interactions with *cis*-elements, or posttranslational and chromatin regulations). Similarly, a lot of work remains to be carried out to decipher the control of their expression. For instance, concerning PA accumulation in the Arabidopsis seed coat, although it has been shown that TT1, TT16 and TTG2 control the activity of the MBW complexes at different

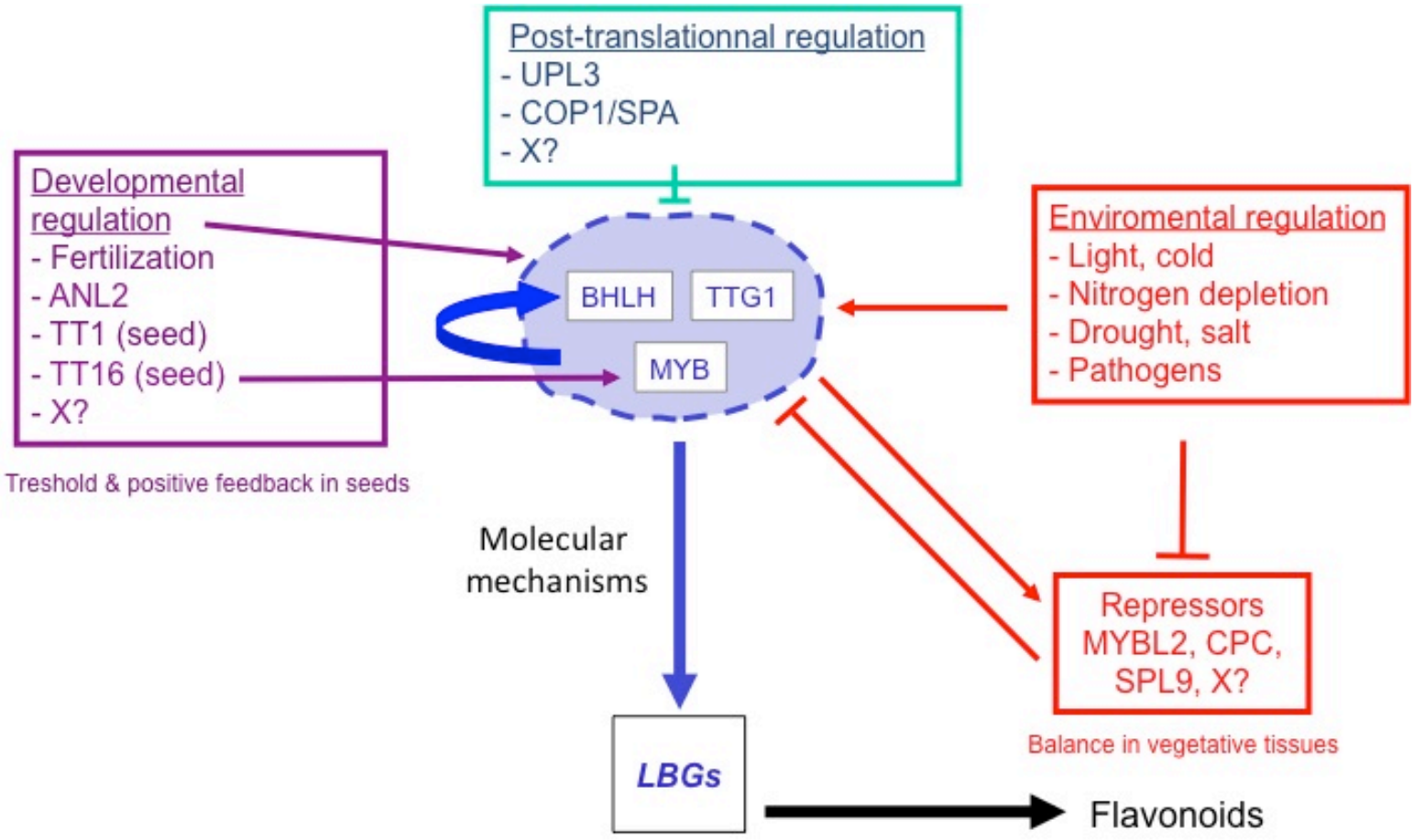


Figure 5-3. Current knowledge and future prospects about flavonoid transcriptional network

levels, their direct targets remain to be identified. Similarly the physiological function of *ATS* (that can bind the promoters of *TT2*, *TT8* and TTG1) should be investigated.

The use of inducible systems *in planta* allowed the characterization of putative targets of the MBW. In addition, we have recently developed a convenient *in vivo* heterologous system (moss protoplasts and GFP analysis by flow cytometry) allowing fast and high-throughput study of the interaction between regulatory proteins and DNA sequences. Nevertheless, the investigations currently suffer from several limitations such as, for instance, genetic and functional redundancies of the partners and *cis*-elements, difficulties in analysing the functions of TFs produced at very low levels and/or in specific cells, limited successes using tagged TFs in vivo (knowing that ectopic expression can lead to misleading interaction) or lack of positive results using EMSA with MBW complexes, *in vitro*. For instance, yeast-two-hybrid screening procedures was undertaken to isolate unknown regulators using TT2, TT8 and TTG1 as bait proteins with IJPB seed cDNA expression library. One of candidates, SPL7, was selected as a putative target of TT2 and as a negative regulator of the TT2-TT8-TTG1 complex to validate further in yeast and in planta. Unfortunately, ectopic expression of *SPL7* and several additional members of *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL1*, *SPL3*, *SPL5*, *SPL9* and *SPL10*), driven by the *TT8* promoter that possesses high expression in Arabidopsis seed, did not lead to *transparent testa* phenotype in Arabidopsis seeds, although a direct interaction between SPL3, SPL9 and TT2 has been detected in yeast. As the posttranscriptional and posttranslational levels mechanisms have been revealed in the regulation of the MBW complexes such as UPS-dependent proteolysis targeting GL3 and EGL3, which is mediated by the ubiquitin protein ligase 3 (UPL3). The post-translational regulation of other MBW regulators controlling PA and anthocyanin biosynthesis is missing, which suggest unknown regulators could be found. Global analyses of transcriptome and TF-DNA interactions in a cell specific manner are required. We are currently developing laser cell ablation and INTACT approaches to meet these challenges.

The regulation of flavonoid biosynthesis and more especially the MBW complexes, structure, function and regulation will continue to provide a powerful system to investigate basic mechanisms of plant gene expression. The interaction between TFs from various

different families makes it an ideal model to understand combinatorial control of transcription. In addition, conservation of the regulatory networks open the way to using this knowledge for crop improvement by classical breeding or genetic engineering.

CHAPTER VI

MATERIALS AND METHODS

6 MATERIALS and METHODS

6.1 MATERIALS

6.1.1 Plant materials

Arabidopsis thaliana belongs to the *Brasicaceae* family of the *dicotyledonous* class. As many features such as small genome, short time life cycle and producing many seeds, which facilitate Arabidopsis becoming a popular model plant used in the laboratory for plant biology and genetics. Plants were routinely grown in a greenhouse (16-hr photoperiod; 10 to 15°C night/20 to 25°C day temperature) on sterilized compost irrigated twice a week with mineral nutrient solution. For crosses and seed production, plants were grown in individual pots. For plant transformation, 20～30 seeds were directly sown on 10×15 cm pot, then were allowed to grow until the flowering.

For aseptic growth, seeds were surface-sterilized and plated on Murashige and Skoog (1962) medium. Petri dishes were first incubated at 4°C for 48 to 72 h to break dormancy and homogenize germination and then were kept in a growth chamber (16-hr photoperiod; 15°C night/20°C day temperature). For Selection of T-DNA mutants, appropriate antibiotics were added to the MS media. To obtain young seedling materials for GUS analysis, we grew plants *in vitro* on the MS media for 8-10 days.

6.1.2 Bacterial strains

DH10B (Escherichia coli)

Genotype: F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80d*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*Δ139 Δ(ara, leu) 7697 *gal*U *gal*Kλ$^-$ *rps*L (Str^R) (Invitrogen, Carlsbad, California). The DH10B strain is used for the amplification of vectors employed in the different cloning reactions.

DB3.1 (Escherichia coli)

Genotype: F^- *gyr* A462 *end*A1 Δ(*sr*1-*rec*A1) *mcr*B *mrr hsd*S20 (r_B-,m_B-) *sup*E44 *ara*-14 *gal*K2 *lac*Y1 *pro*A2 *rps*L20(Sm^r) *xy*1-5λ$^-$*leu mtl*-1 (Invitrogen, Carlsbad, California). The

DB3.1 strain is ccdB resistant (ccdB protein induces a dysfunction of the gyrase, which provokes DNA ruptures, blocking growth), which allows the multiplication of vectors containing this contra-selection gene. This kind of bacteria are commonly used for the entry- and destination-vectors used with the Gateway® recombination cloning method.

C58C1Rif pMP90 (Agrobacterium tumefasciens)

This strain described by Koncz and Schell (1986) is used for *A. thaliana* transforming by inflorescence agro-infiltration (Bechtold *et al*., 1993). It has rifampicin (C58C1) and gentamycin resistance gene presents in the Ti plasmid (pMP90).

The bacteria are cultured in a LB medium (yeast extract 5 $g.l^{-1}$, bacto-tryptone 10 $g.l^{-1}$, NaCl 10 $g.l^{-1}$, pH 7.2) at 37°C for *E. coli* and 28°C for *A. tumefasciens*. The medium is supplemented with one or several antibiotics (Duchefa, Haarlem, Pays-Bas) if necessary, following the conditions as table 6-1. The stock solutions are sterilized by filtration (0.22 um filter, Scheicher and Schuell, Dassel, Germany).

6.1.3 Yeast strains

YM4271

Genotype: *MAT*a*, ura3-52, his3-200, ade2-101, ade5, lys2-801, trp1-901, leu2-3, 112, tyr1-501, gal4Δ, gal80Δ, ade5::hisG.* This strain is used to study DNA-protein interactions by the one hybrid method (Clontech, Palo Alto, California). The target DNA is cloned upstream of the minimum promoter of *HIS3* reporter gene in the pHISi vector. Then, the vector is integrated by homologuous recombination at *URA3*-52 or *HIS3*-200 loci.

6.1.4 Moss strains

The Gransden wild-type strain of *Physcomitrella patens* was used in this study. Freshly fragmented protonema were inoculated on solid PPNH4 medium supplemented with 2.7 mM NH4-tartrate (Thevenin *et al.*, 2012), overlaid with sterile cellophane disks (Cellulose type 325P, A.A. Packaging Limited, Preston, Lancashire, UK) in 90-mm Petri dishes, sealed with 3M Micropore™. Protonemal cultures were grown for 7 d at 24°C with a light regime of 16 h

Antibiotics	Stock solution	Final concentration
Kanamycin	100 $mg.ml^{-1}$	50 $mg.l^{-1}$
Ampicillin	100 $mg.ml^{-1}$	100 $mg.l^{-1}$
Gentamicin	50 $mg.ml^{-1}$	50 $mg.l^{-1}$
Hygromycin	50 $mg.ml^{-1}$	50 $mg.l^{-1}$
Chloramphenicol	30 $mg.ml^{-1}$	30 $mg.l^{-1}$
Rifampicin	25 $mg.ml^{-1}$	25 $mg.l^{-1}$
Cefotaxime	200 $mg.ml^{-1}$	200 $mg.l^{-1}$

Table 6-1. The concentration of antibiotics used in this study

light/8 h darkness at 80 μmol $m^{-2}s^{-1}$ (Thevenin *et al.*, 2012).

6.1.5 Vectors of entry and destination Gateway™

pDONR207

This 5585 bp entry vector (Fig. 6-1a; Invitrogen, Carlsbad, California) contains an attP1/attP2 recombination cassette with a chloramphenicol resistance gene and a ccdB contra-selection gene. Out of this cassette there is a gentamicin resistance gene to make selections in *E. coli*.

pBI101-GTW

This 13948 bp binary destination vector contains an attR1/attR2 recombination cassette in the pBI101 plasmid (Fig. 6-1b; Fanchon Divol, Laboratoire de Biologie Cellulaire, INRA, Versailles); used for a transcriptional fusion between the reporter gene *uidA* (coding the *β*-glucoronidase or GUS), contains the kanamycin and chloramphenicol resistance genes to make selections in *E. coli*. *In planta*, the integration was verified by kanaycin resistance.

pGWB3

This 18275 bp binary destination vector contains an attR1/attR2 recombination cassette derived from the pBI101 plasmid (Fig. 6-1c; Nakagawa *et al.*, 2007); used for a transcriptional fusion between the reporter gene uidA (coding the β-glucoronidase-GUS), contains the kanamycin and chloramphenicol resistance genes to make selections in *E. coli*. In *planta*, the integration is verified by hygromycin resistance.

pR1R2ΔGR

This 13888 bp binary plasmid (Fig. 6-1d; Bertrand Dubreucq, Laboratoire de Biologie de Semences, INRA, Versailles) is a destination vector modified from pΔGR (Schena *et al.*, 1991), by introduction of an attR1/attR2 recombination cassette in the Xba1 site converted on blunt ended by the activity of the Klenow fragment of the DNA polymerase 1, from *E. coli*. It allows the expression of a translational fusion between a gene of interest and the glucocorticoid receptor (GR) (from 508 to 795 aa of the rat GR), controlled by the 35S

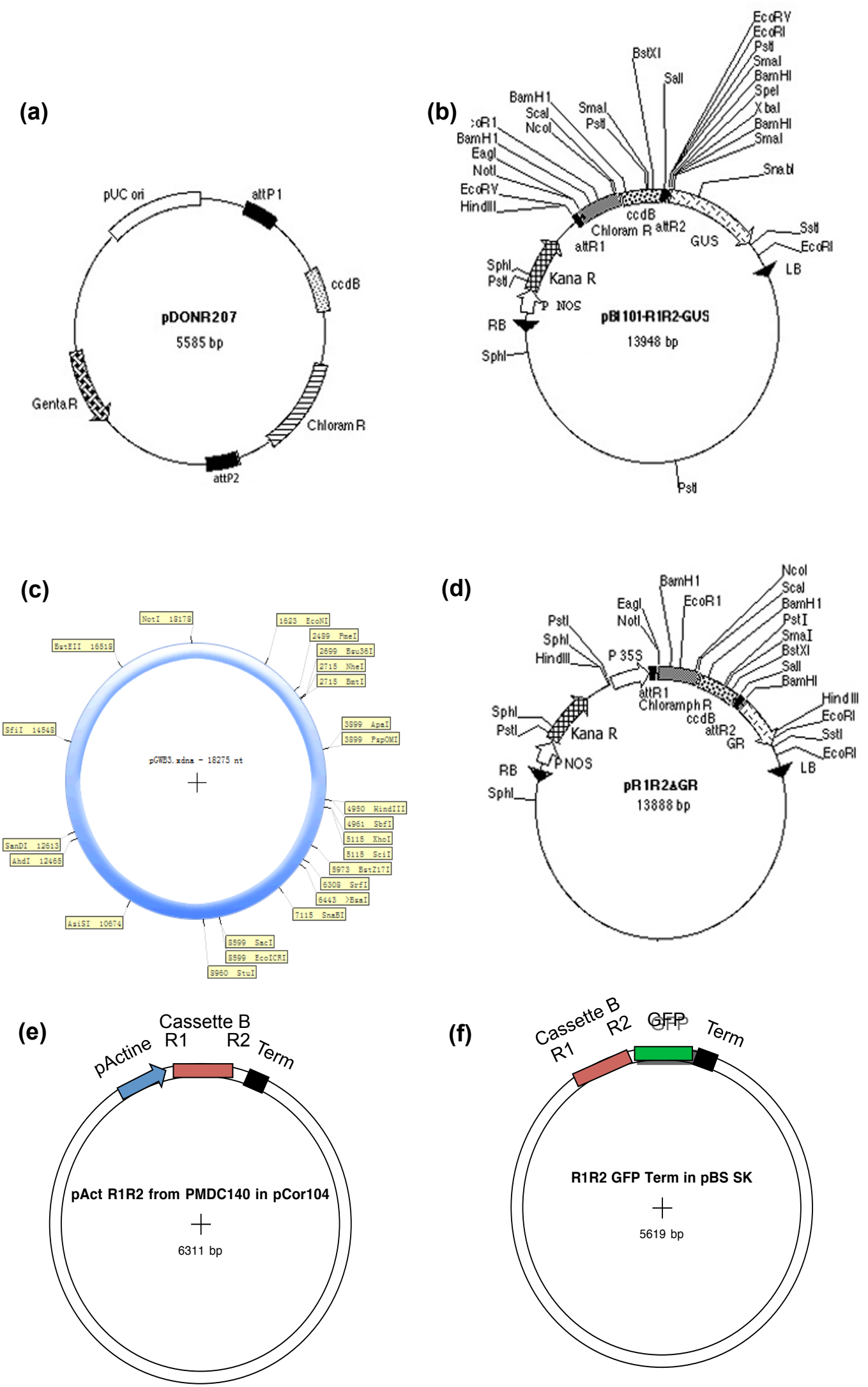


Figure 6-1. Map of vectors

promoter (CaMV). To be selected in *E. coli*, the vector contains the kanamycin and chloramphenicol resistance genes and *in planta* the integration is checked by the kanamycin resistance.

pBS TPp-A

A DNA fragment containing the Gateway® (Life Technologies, Carlsbad, CA, USA) recombination cassette, comprising the attR1 site, chloramphenicol resistance gene, ccdB counter selectable marker and attR2 site, was obtained after XbaI digestion of the pMDC140 vector (Curtis & Grossniklaus, 2003), and ligated into the XbaI-digested pCOR104-CaMVter plasmid (Proust *et al.*, 2011) between the rice actin promoter and the 35S cauliflower mosaic virus terminator (Fig. 6-1e). Ampicillin resistance gene is used for the selection in *E. coli*.

pBS TPp-B

A DNA fragment containing the Gateway® recombination cassette, comprising the attR1 site, chloramphenicol resistance gene, ccdB counter selectable marker, attR2 and GFP, was purified from the HindIII- and SacI-digested pGWB4 vector (Nakagawa *et al.*, 2007). The pCOR104-CaMVter vector was digested with SmaI, blunt ended with Klenow and digested again with HindIII. The insert was then ligated into the digested HindIII-Blunt pCOR104-CaMVter to generate the pBS TPp-B vector (Fig. 6-1f). Ampicillin resistance gene is used to the selection in *E. coli*.

pBIB-Hyg proTT8:GTW

This construct is based on the pBIB-Hyg binary vector (Fig. 6-2d) in which the Gateway® recombination cassette has been EcoRV-inserted at the multiple cloning site, giving the pBIB-Hyg-GTW vector. The orientation of the Gateway® recombination cassette was verified by SalI digestion. The 1518-bp TT8 promoter fragment was amplified from the pDONR207 vector (Gateway®) described in Baudry *et al*. (2006), with the primers proTT8-5'-HindIII and proTT8-3'-XbaI. The DNA fragment obtained was subsequently cloned into the pBIB-Hyg-GTW vector digested by HindIII and XbaI, giving the pBIB-Hyg

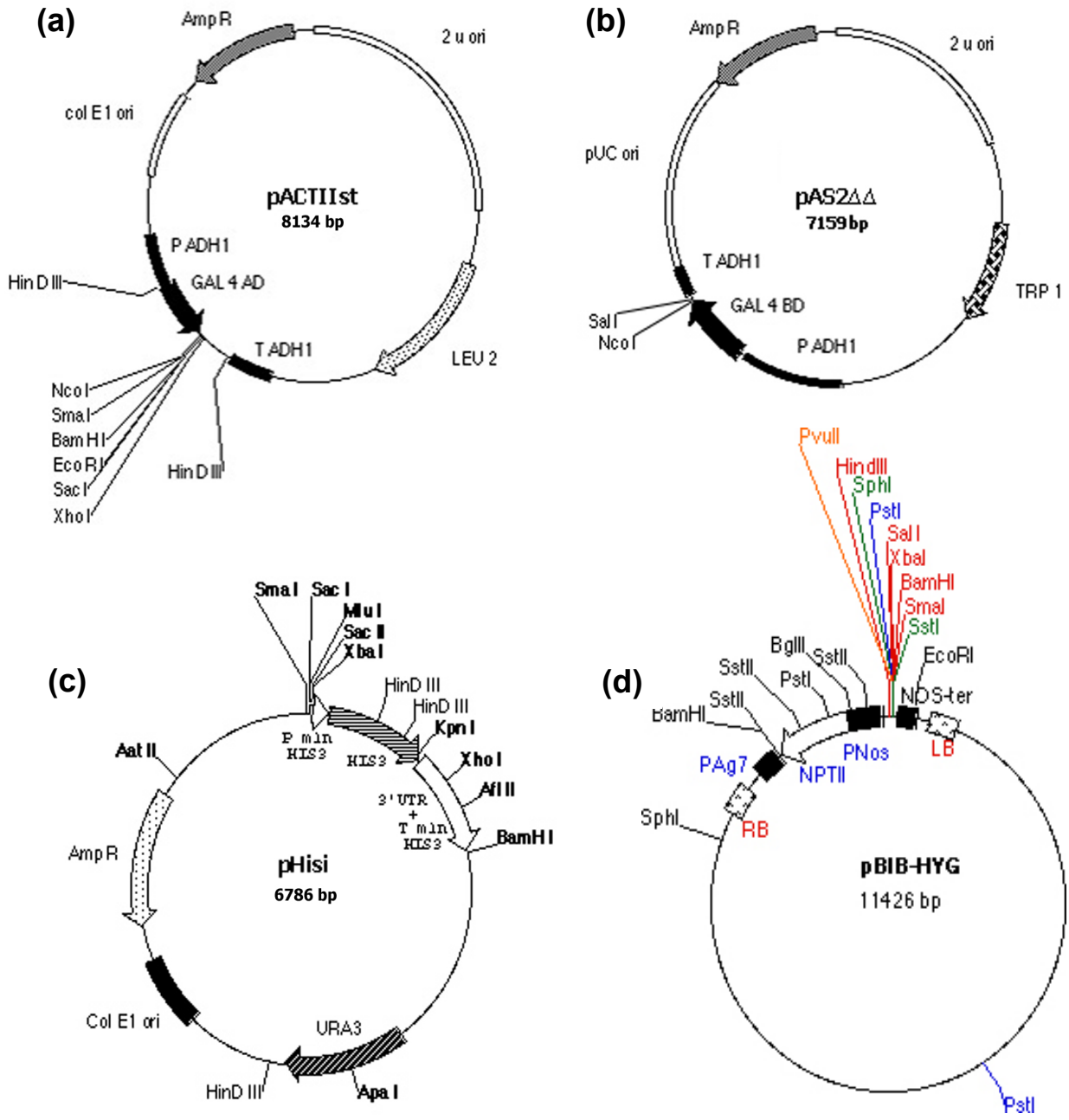


Figure 6-2. Map of vectors

proTT8:GTW vector.

6.1.6 Vectors of yeast one-hybrid

pACTIIst

This 8134 bp plasmid allows cloning a gene of interest in a translational fusion with the GAL4 activation domain (AD) letting the fusion protein expression in yeast (Fig. 6-2a) (Fromont-Racine *et al.*, 1997). This plasmid contains two selection genes, one for yeast use (*LEU2*) and the other for *E. coli* (ampicillin resistance). pACTIIst permits the constitutive expression of the fusion protein thanks to the *ALCOHOL DEHYDROGENASE 1* strong promoter (*ADH1*).

PAS2ΔΔplasmid

This 7159 bp plasmid allows cloning a gene of interest in a translational fusion with the GAL4 DNA bind domain (BD) letting the fusion protein expression in yeast (Fig. 6-2b) (Fromont-Racine *et al.*, 1997). This plasmid contains two selection genes, one for yeast manipulations (*TRP1* gene) and the other for *E. coli* (ampicillin resistance). pAS2ΔΔ permits constitutive fusion protein expression tanks to *ADH1* strong promoter.

pHISi plasmid

This 6786 bp plasmid is used to make transcriptional fusions between a promoter and a *HIS3* reporter gene (Fig. 6-2c; Clontech, Palo Alto, California). The promoter sequence is cloned upstream of *HIS3* minimal promoter and might be integrated in the YM4271 yeast genome by homologous recombination to *ura3* locus. Nevertheless, it cannot be replicated in yeast because of the absence of its replication origin (ORF). Ampicillin resistance gene is used for the selection in *E. coli*.

6.2 METHODS

6.2.1 Manipulation of plants and seeds

Seed sterilization

The sterilization solution is prepared by dissolution of an active chloride tablet (Inovchlore; Inov'chem, Brest, France) in 40 ml of distilled water, then a 1:10 dilution is made with 95° ethanol. 1 ml of sterilization solution is used to incubate 20 mg of seeds for 5 min, next seeds are rinsed with 95° ethanol and dried over-night under sterile air flux.

Arabidopsis crossing

For the most efficient crossings, we use a mother plant that have developed 5-6 inflorescences, and a father plants have formed siliques that indicate the pollen is fine. The emasculation of buds is performed with tweezer and manually pollinated with the mature anthers taken from the male plant. The seeds issues from the cross have to be harvest at maturity before silique dehiscence.

Genetic transformation of Arabidopsis

5 ml of LB liquid medium with the appropriate antibiotics are inoculated with an isolated agrobacterium colony and cultured for 24 h at 28°C in a shaker (200 RPM). 1 ml of this pre-culture is used to inoculate 500 ml of LB liquid medium. Incubation in a shaker (200 RPM) is performed to obtain a 0.8 DO_{600} (around 15 h). The culture is distributed into two 250 ml bottles to be centrifuged 10 min at 6500 RPM (rotor n. 12169, centrifuge 4K15). Each pellet is re-suspended in 250 ml of water to which 100 µl of Silwet-L77 and 5% sucrose are added (Witco Corporation, Friendly, Virginia). The transformation is made by 10 min of immersion of the aerial parts of the plants (when the first siliques have appeared) into the transformation solution ("barquette" containing around 30 individuals). The seeds harvested are sown in a selective solid medium to obtain the primary transformed plants (T1) (Bechtold *et al*., 1993).

6.2.2 DNA extraction, amplification and cloning

Plasmid extraction

Mini-preparations of plasmidic DNA (to obtain 10 ug of DNA) are made from 5 ml of bacterial cultures brought at saturation, using the NucleoSpin plasmid Isolation kit (Macherey-Nagel, Germany) following the recommendations of the manufacturer. In order to obtain bigger DNA quantities, the culture volume is adapted to the type of plasmid (from 25 to 500 ml depending of the number of plasmid copies by cell) and the Sigma Plasmid MaxiPrep kit (GenEluteTM HP, sigma-aldrich, USA) is used. In all cases, the extraction protocol consists in alkaline lyses as described by Birnboim and Doly (1979) in the presence of RNase-A and a separation of the genomic DNA by selective precipitation (Birnboim & Doly, 1979). Finally, plasmid purification is made using an ions interchange column.

DNA extraction for vegetal tissues

This method consists in a fast DNA extraction to be carried out genotyping (DNA is preferentially used during a month after extraction because of possible DNA degradation due to some impurities presence). Harvest a younger leaf and transfer it into an eppendorf tube. Process with a pestle and add 400 µl of extraction buffer (200 mM Tris-HCl, pH 7.5, 150 mM NaCl, 25 mM EDTA and 0.5% SDS) and vortex for 5 s. Then centrifuge for 1 min at 13000 RPM to precipitate the proteins. Transfer 300 µl of supernatant to a new eppendorf tube and add 300 µl of isopropanol and mix well. Incubate for 2 min. Centrifuge 5 min at 13000 RPM to precipitate the DNA. Then discard the supernatant and let pellet dry. Re-suspend in 50 µl of distilled water.

DNA amplification by Polymerase Chain Reaction (PCR)

The DNA amplification by PCR is made by the activity of a bacterial thermostable polymerase type Taq (*Thermus aquaticus*). Use 0.5 U of Thermo® (Thermo Scientific, USA) for 20 µl of reaction medium: 2 ul of 10X Taq buffer, 0.8 ul of 5 mM of the dNTP mix (Invitrogen) (final concentration at 0.2 mM), 1 ul of 10 uM each ologonucleotide (final concentration at 0.5 uM), 10 pg - 1 ug DNA template. Usually, the PCR reaction consist in a denaturation phase, 3 min to 94°C; followed by 35 cycles comprising: denaturation, 30s at

94°C; oligonucleotide hybridization, 30s at 55-60°C; elongation, 1min/1000bp at 72°C; then a final 10 min elongation step at 72°C. These conditions are adapted according to the GC/AT ratio.

DNA analyzis

DNA electrophoresis in an agarose gel allows qualitative and semi-quantitative analyzis. The electrophoresis is performed in 0.5X TAE buffer (Tris-acetate at 20 mM and EDTA at 0.5 mM; Eiromedex, Strasbourg, France), in a Mupid-21® cell (CosmoBio,Tokyo, Japon) with a power of 135V. The gel is made by hot agarose dissolution in 0.5% TAE, usually 1% (w/v), varying between 0.5 to 3% depending on the size of the DNA fragments to be separated (small fragments should use high concentration). BET is added at a 10 ng/µl final concentration just before polymerization (Amresco, Solo, Ohio). The samples containing the charge buffer (Bromophenol blue at 0.25 % (w/v), cyanol-xilene at 0.25% (w/v) and glycerol 30% (v/v)) in a 1/10 proportion are runed into the gel. After migration, the DNA is visualized under "UV radiation" and it is quantified if necessary by comparison with the Smart® ladder (0.2 to 10 kb; Eurogentec, Liège, Belgique).

Gateway® cloning by recombination

This type of recombination uses the specific sites, discovered in the λphage by Landy (1989) (Landy, 1989). The two steps of recombination are: i) introduction of the insert in the entry vector by a recombination type BP and ii) transfer of the insert to a destination vector by a recombination type LR (Fig. 6-3, adapted from Magnani *et al.*, 2006).

In this recombinant cloning technique, the sequence being cloned is amplified by PCR using Phusion High Fidelity DNA Polymerase (Thermo Scientific, USA).

The PCR amplification allows the introduction of the attB1 or attB2 recombination ends, in only one amplification or in two steps amplification, depending the efficacy. The reaction in two steps consist in: i) use of the specific oligonucleotides added of the 5'→ 3': AAAAAAGCAGGCT (a half attB1 site) or 5'→ 3': AAGAAAGCTGGGT (a half attB2 site) in 5'following the orientation of the wished insertion; ii) the second PCR reaction uses the

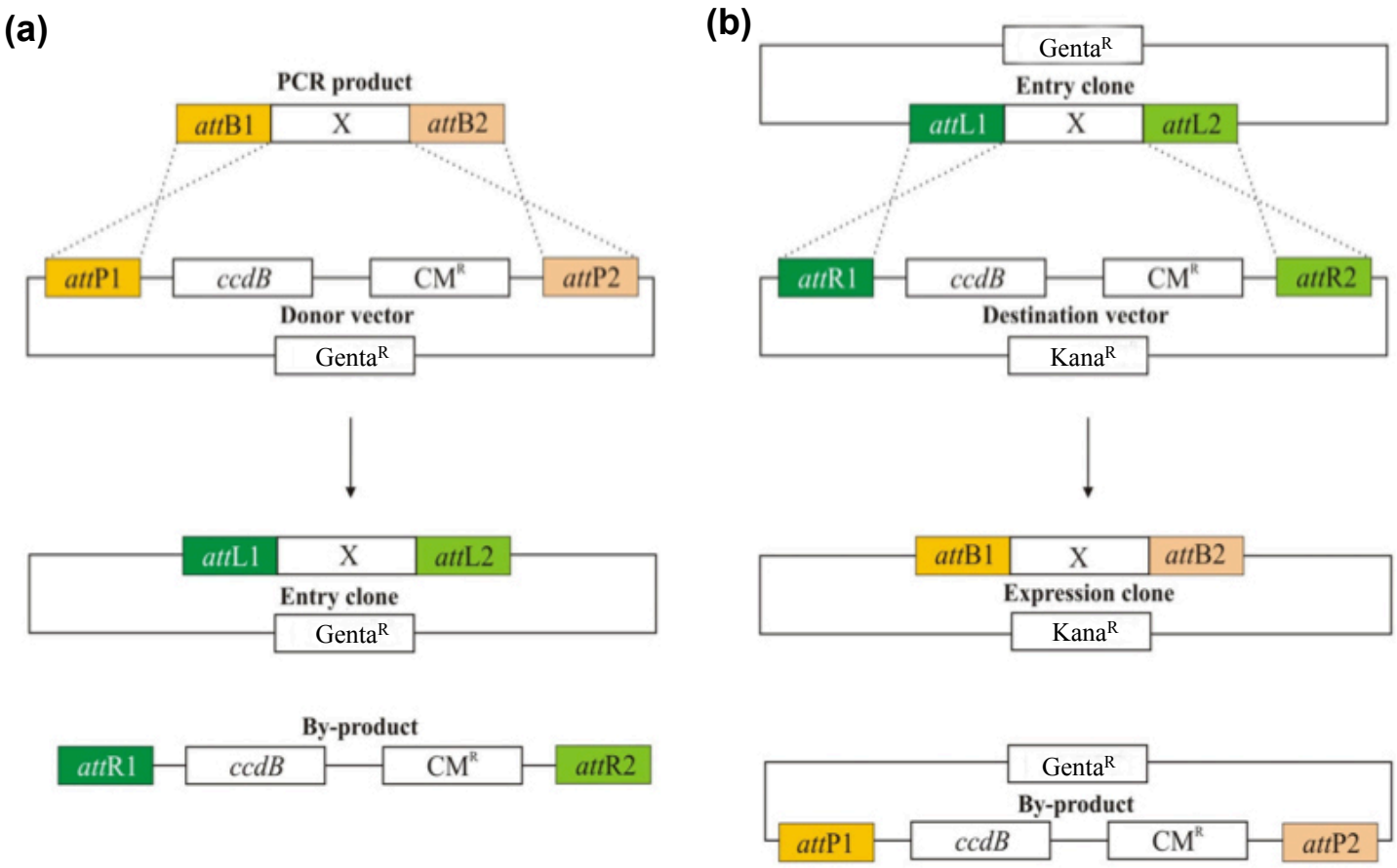


Figure 6-3. Gateway BP and LR reactions (Adapted from Magnani *et al*., 2006). (a) BP recombination of a PCR product "X" flanked by attB sites with a Gateway donor vector *pDNNOR207*. (b) LR recombination of an entry clone bearing a DNA fragment "X" with a Gateway destination vector *pGWB3.*

full attB1 and attB2 oligonucleotides and 1 µl of a 1/1000 dilution of the first amplification. The amplification in one step is made using the specific oligonucleotides added of the whole attB1 and attB2 sites. Then, 2.5 µl of the final product (made by one or two steps) is mixed with 0.5 ul of pDONR207 entry vector (at 200 $ng.l^{-1}$), 1 µl of BP buffer and 1 µl of BP Clonase® (Invitrogen, Carlsbad, California). Then the incubation is performed at 28°C for 2 h. The recombination reaction stops with the addition of 0.5 ul K proteinase (2 $ug.l^{-1}$; Invitrogen) and incubation for 10 min. 3 µl of this mix is used to transform 20 µl of DH10B electrocompetent bacteria. The selection of clones containing recombined vector is made on gentamycin. The vectors that do not contain an insert are not amplified due to the presence of ccdB gene whose product is toxic for DH10B strain. A clone is chosen when the correct sequence is obtained.

The transfer into the destination vector is made by LR type recombination. 100 ng of pDONR207 (2 µl) containing the insert is combined with 100 ng (1 µl) a destination vector, 1 µl of LR clonase buffer and 1 µl of LR Clonase® in 5 µl final volume (use distillated water if necessary). Then, the same procedure is followed as in BP reaction except for the selection antibiotic; the gentamicin is replaced by other appropriate antibiotics.

Bacterial transformation by electroporation

Competent bacterial cells are transformed with a recombinant vector by an electric shock (Potter, 1993). A mix of 3 µl of ligation reaction and 20 µl of defrost bacteria maintained in glass are deposited in an electroporation cell (1 mm; Biorad, Hercues; California). The electroporation cell is placed in the electroporator (MicroPulser®; Biorad, Hercules, California) and a power of 1.25 kV is applied. The bacteria are recovered in 1 ml of LB liquid medium and placed in a shaker at 220 RPM for 1 h at 37°C. The transformed clones are spread over LB solid medium plates containing the appropriate antibiotic.

Sequencing

Vectors are sent to Genoscreen (Campus de l'Institut Pasteur, Lille, France) for sequencing and sequence identification was performed through NCBI/TAIR nucleotide-nucleotide

BLAST searches.

6.2.3 Gene expression analysis by quantitative RT-PCR (Q-RT-PCR)

RNA extraction

The extraction is made through Genelute® Mamalian Total RNA kit (Sigma-Aldrich, Steinheim, Germany), which is added of a DNA degradation step. This step is made after the first cleaning of the RNA fixed in a column; 10 µl of "RNase-free" DNase (Qiagen, Hilden, Germany) are thoroughly taken in 70 µl of RDD buffer (Qiagen) and deposed in the column. After 15-min incubation at room temperature, the column clean up protocol continues and the RNA is eluted in 50 µl "RNase-free" water (Sigma-Aldrich).

Reverse-Transcription (RT) Reaction

RT reaction is made using the Superscript® II RNase H^- Reverse Transcriptase kit (Invitrogen, Carlsbad, California) following the recommendations of the manufacturer (Table 6-2). The simple strand cDNA is synthesized in 20 µl volume reaction: from 1 µg to 5 µg of total RNA denaturized by 5 min incubation at 65°C in $MgCl_2$ at 3mM, 0.25 mg of oligo dT primer oligo (Genosys, Cambridge; Grang-Britain), 0.25 mg of Random Hexamers, DTT at 10mM, dNTPs at 0.5 mM, 40 U of RNase OUT® (RNase inhibitor) and 200 U of Superscript® II and reaction buffer. The RT-PCR is performed at 42°C for 50 min, then the enzyme is inactivated by incubation at 70°C for 15 min (Table 6-2). The cDNA is diluted to serve as a matrix for PCR (usually from 1/10 to 1/250).

Quantitative RT-PCR (Q-RT-PCR) reaction

The FastStart DNA Master SYBR Green I (Roche) kit is used following the recommendations of the manufacturer. 10 µl of reaction medium are composed of 2.5 µl of DNA matrice (between 1/50 and 1/250 dilution of RT product), 0.5 µl of each oligonucleotide at 10 µM (Ganosys, Cambrdge, Grand-Bretain), 2 µl of FastStart Master SYBR Green I buffer (Roche), 2.5 µl distilled water. Q-PCR reaction system described as

cDNA synthesis (20ul)	Oligo dT 100mM	0.5 ul	65°C for 5min Then on ice
	Random Hexamers	0.5 ul	
	dNTP 5mM	1 ul	
	RNA	1ng-5ug	
	5X Buffer	4 ul	42°C for 50 min 70°C for 15 min
	DTT 0.1M	2 ul	
	RNase inhibitor	1 ul	
	Superscript II	0.25 ul	

Table 6-2. The reaction system of reverse transcription (RT)

Table 6-3.

Usually, the amplification reaction consists in a denaturation and Taq activation step at 95°C for 8 min and 40 times repeated cycle: 10 s of denaturation at 95°C, 7 s of oligonucleotides hybridization at 60°C, and 10 s of elongation at 72°C. Then the fusion curve of the reaction product is determined by denaturizing at 95°C, 30s of incubation at 65°C and the measure of the fluorescence diminution through a slowly temperature rise from 65 to 95°C (0.1°C/s). The last step allows determining the semi-denaturizing temperature (Tm), which is specific of each amplification product, Q-PCR running program described as Table 6-4.

Choice and validation of oligonucleotides

The oligonucleotide couples have to permit the amplification of a 150 bp specific fragment of the target to be quantified. The fusion curve profile allows confirmation over a first amplification, the specificity of the oligonucleotides (only one product amplified) and the absence of the match between the oligonucleotides. A second amplification is performed over different DNA matrix dilutions (from 1:10 to 1:100000). Each concentration corresponds to a defined number of cycles as the outlet point designates, a diminution of the concentration at each dilution provokes the rising of number of cycles to obtain the same quantity of amplification product. The representation of the correspondent number of cycles in function of the concentration logarithm is a linear curve, whose tangent allows determining the amplification efficiency of the oligonucleotides. This efficiency has to be comparable to that of the oligonucleotides used for the reference gene.

Quantification results

The quantification is expressed in a relative manner to a reference gene. The expression level of EF1αA4 (EF), which encodes for a translation elongation factor, whose level remains constant during seeds and seedlings development (Liboz *et al.*, 1990). The EF expression level is determined in parallel to the gene of interest. In each sample the outlet point of EF and of the gene of interest is determined to calculate the difference of the number of cycles

Q-PCR reaction (10ul)	Master mix	2 ul
	Primer-F	0.5 ul
	Primer-R	0.5 ul
	H2O	4.5 ul

Table 6-3. Q-PCR reaction system

Temperature	**Incubation time**	**Temperature transition rate**	
95°C	8 min	20°C/sec	Denaturation
95°C	10 sec	20°C/sec	
60°C	7 sec	20°C/sec	Cycles 40X
72°C	10 sec	20°C/sec	
95°C	0	20°C/sec	
65°C	30 sec	20°C/sec	Fusion
95°C	0	0.1°C/sec	
40°C	2 min	20°C/sec	Refroidissement

Table 6-4. Q-PCR running program

between each outlet point. Then this difference is used to calculate the relative expression level, which is termed in percent of the EF expression.

- Difference of the number of cycles: DC = EF outlet point – Y gene outlet point
- Relative expression level: $N = 1/(2^{DC})$ x 100% of EF for $DC \geq 0$
  $N = 2^{DC}$ x 100% of EF for $DC \leq 0$

6.2.4 One and two hybrid methods

The principle of Yeast transformation

Yeast transformation is made by the Gietz and Woods (2002) method, using lithium acetate, DNA of salmon sperm (Yeast maker carrier DNA; Clontech, Palo Alto, California) and polyethylene glycol (PEG) (Gietz & Woods, 2002). The lithium acetate weakens cell wall becoming permeable for the nucleic acids; the carrier DNA is supposed to act carrying plasmidic DNA into the cell or stimulating an active mechanism of DNA absorption, finally, PEG acts as a moderator of lithium acetate effect, preventing the cell lyses.

Preparation of the competent yeast

Two types of protocols were used to obtain competent yeasts. The first called high efficacy, allowing taking a high quantity of transformed yeasts, which is important for a cDNA library transformation with or for a co-transformation of several plasmids simultaneously. The second protocol can be used when not many transformed yeasts are needed.

① Competent yeast for transformation at high efficacy (10^5 transformed yeasts by μg of DNA)

This protocol allows obtaining enough competent yeasts to make 10 transformations. A 10 ml pre-culture in YPDA selective medium is made during the night in a shaker at 200 RPM and 28°C. The DO_{600nm} of this pre-culture is determined (using a 1:10 dilution) and the equivalent volume at 2.5 x 10^8 cells (DO_{600nm} = 0.1 is equivalent to 10^6 cell/ml) is used to inoculate 50 ml of medium. This culture rises at a DO_{600nm} between 1.5 and 2 (about 3 or 4 h). The culture is centrifuged at 5000 RPM for 5 min (rotor n. 12169, 4K15 centrifuge). The pellet is re-suspended in 25 ml of sterile distilled water and newly centrifuged. The pellet is

re-suspended in 1 ml of LiAc at 100 mM and incubated at 28°C for 10 min.

② *Competent yeast for transformation at low efficacy (200 transformants by* μg *of DNA)*

For each transformation, a quantity of cells growth over solid medium corresponding to about at 25 μl is put in a suspension in 1 ml sterile distillated water. The yeast are centrifuged for 10 s at 13 000 RPM (5415 D eppendorf centrifuge), washed a second time, then re-suspended in 1 ml of LiAc at 100mM and incubated at 28°C for 10 min.

Yeast transformation

The tranformation is made from the pellet obtained by centrifugation for 10 s (5415 D eppendorf centrifuge). Then the following compounds are added: 240 μl of PEG 3350 at 50%, 36 μl LiAc 1M, 10 μl of DNA of salmon sperm (10 $mg.ml^{-1}$) and 75 μl of plasmidic DNA (1 μg). After homogenization for 1 min, the yeast are incubated at 28°C for 30 min, then a 42°C-heat-shock is applied in a water-bath for 20 min. After centrifugation at 13 000 RPM for 10 s (5415 D eppendorf centrifuge), the pellet is re-suspended in 1 ml of sterile distilled water. 200 μl of this suspension and several dilutions are spread over selective solid medium (restrictive medium depending of the selection marker: auxotrophic gene). The incubation is made at 28°C from 2 to 4 days. In the case of low efficiency, the last pellet is re-suspended in 200 μl of sterile distilled water and completely spread.

Integration of a promoter/regulatory sequence-reporter gene transcriptional fusion is made by homologuous recombination in an YM4271 yeast strain. The transcriptional fusion between a DNA target and the reporter gene *HIS3* is made by cloning in the pHISi vector (Clontech, Palo Alto, California). A linearization by restriction at the site ApaI or NcoI of *URA3* gene is made. The restriction efficiency is controlled by electrophoresis, 2 of 20 μl of restriction reaction are used and the remaining is used for the transformation reaction. An aliquot of the competent yeast YM4271 strain (protocol at high efficacy) is added of 240 μl of PEG 3350 at 50%, 36 μl LiAc 1M, 10 μl of DNA of salmon sperm (10 $mg.ml^{-1}$) and 18 μl of linear vector and 57 μl of distilled water. The same protocol is followed as described in the precedent paragraph for transformation by a recombinant plasmid. The whole pellet is spread over medium without uracil to select the transformants that have integrated the regulatory

sequences by homologue recombination at *URA3* locus. It is important to note that the non-integrated vectors cannot be maintained because the pHISi do not contain the specific replication origin for yeast (usually the replication origin measures 2μ). The integration efficiency observed was between 10 and 100 transformants per μg of DNA (linear vector).

6.2.5 Histochemical detection of β-glucoronidase (GUS) activity

Plant material is placed in the infiltration buffer immediately after cut. The infiltration buffer is changed with the introduction of the substrate of the β-glucoronidase, the X-Gluc (5-bromo-4chloro-3indolyl-β-glucoronic acid; Duchefa, Harleem, Pays-Bas; Jefferson *et al*., 1987); supplemented by ferricyanide-$K_3[Fe(CN_6)]$-and potassium ferrocyanide-K_4FeCN_6, $3H_20$, which limits the diffusion of the reaction product, its concentrations being adjusted (usually from 0.5 to 5 mM) in function of the β-glucoronidase activity intensity (Table 6-5). To assure the penetration into the cells, vacuum is rapidly applied twice, and then maintained for 1 h. The samples are incubated at 37°C in the dark overnight. Finally, samples are destained by the successive rinsing with 70° ethanol. The results are observed under binocular and/or optic microscop.

Preparation of the samples to be qualitatively observed by optic microscopy for whole-mount analysis

To precisely localize GUS activity, the samples are observed on an Axioplan 2 optic microscop (Zeiss, Iena, Germany). After coloration, incubation with Chloral hydrate solution (100 g of chloral-hydrate, 25 ml of glycerol and 50 ml of distilled water) is made allowing observation under interferential contrast with the help of Nomarski optic.

Preparation of the samples to be qualitatively observed by optic microscopy for dissection analysis

For siliques, they are include in the Technovit 7100 resin (Heraeus Kulser, Wehrheim, Germany) after fixation in a solution of glutararldehyde-paraformaldehyde (2.5% and 4 % (w/v), respectively) in a sodium phosphate buffer, pH 7.2, 0.1% triton X-100 (v/v) by

	Composition	Final Concentration	For 100 ml
GUS buffer	Phosphate buffer	50 mM	50 ml 0.1 M pH 7.2
	EDTA	10 mM	2 ml 0.5 M
	Triton 100	0.1 %	1 ml 10%
	X-Gluc	2 mM	104 mg (dissolved in 2 ml DMF)

Table 6-5. The composition of GUS staining buffer

infiltration under vacuum for 1h, then incubated overnight at 4°C. The following day the tissues are progressively dehydrated through multiple ethanol both with increasing ethanol concentrations (10°, 30°, 50°, 70°, 85°, 95°). Then, ethanol is progressively replaced by Technovit 7100. After polymerization at room temperature in the Teflon moulds for 48 h, sections of 8 μm are cut with a microtome (Leica RM 2055, Rueil-Malmaison, France) and fixed over glass slides.

6.2.6 Dexamethasone (DEX) induction experiments

DEX (Sigma-Aldrich, Steinheim, Germany) induction for the monitoring of target gene expression in siliques was performed in 24-well plates. For each condition, four 4-day-old siliques (after pollination) were harvested from 35S:TT-GR transgenic plants, opened with a razor blade, and incubated in 10 μM DEX in 100 mM phosphate buffer, pH 7.2, 0.1% triton X-100, 10 mM Na_2-EDTA, and 100 μM CHX when necessary (Spelt *et al.*, 2000); CHX Ready made; Sigma-Aldrich). Vacuum was applied for 30 min to ensure effective penetration of CHX. Thereafter, DEX was added to a 10 μM final concentration, and the penetration was facilitated by a second round of 30 min vacuum (Fig. 6-4).

6.2.7 Transient expression system in Physcomitrella patens protoplasts

Moss protoplast preparation

Protoplasts were isolated from 6- to 7-day-old protonema by incubation for 1 to 2 h in 2% Driselase (Sigma D9515) dissolved in 0.47 M mannitol. The suspension was filtered successively through 63- and 40-μm stainless steel sieves. Protoplasts were sedimented by low-speed centrifugation (60 g for 5 min at 20°C) and washed twice with 0.47 M mannitol. Protoplasts were then resuspended at 0.6 to 1.2 $\times 10^6$ protoplasts ml^{-1} in MMM solution (0.47 M mannitol, 15 mM MgCl2 and 0.1% 2-(Nmorpholino) ethanesulfonic acid (MES), pH 5.6) for transformation (Fig. 6-5) (Thevenin *et al.*, 2012).

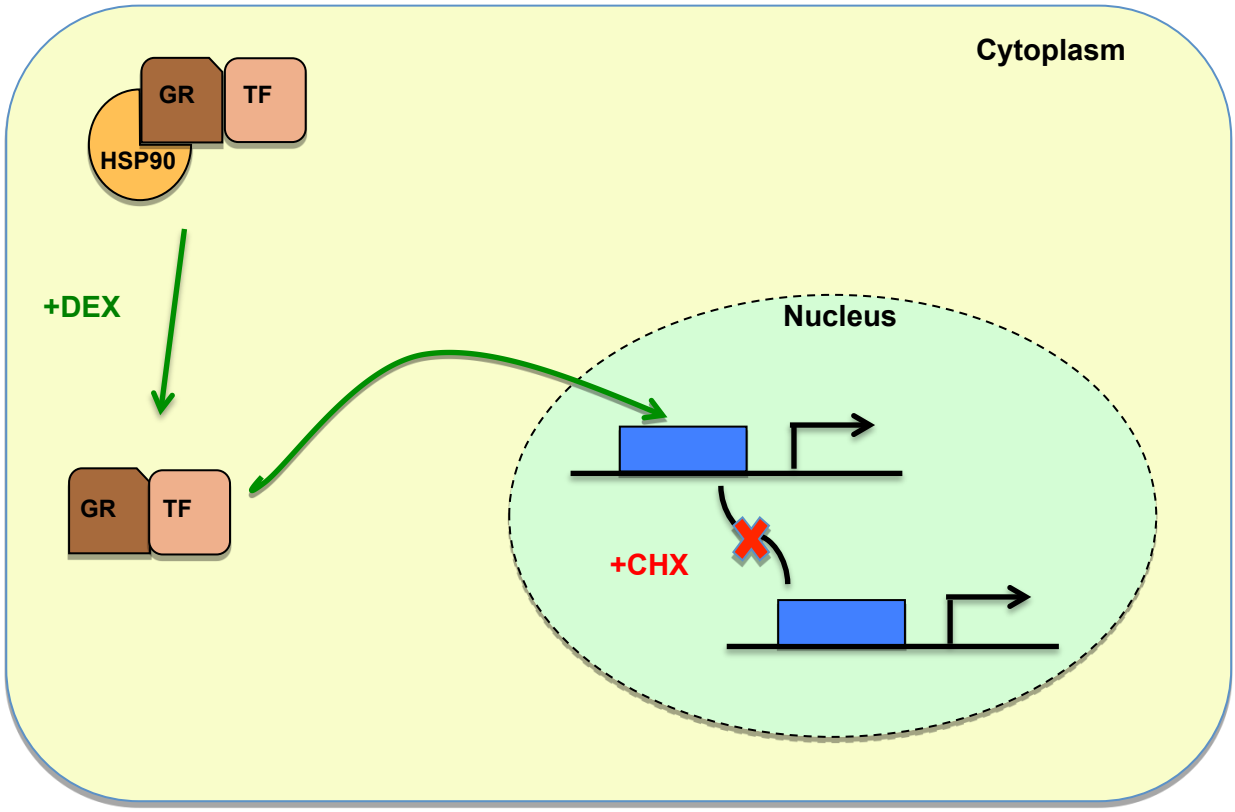


Figure 6-4. Principle of dexamethasone (DEX) induction. DEX, Synthetic glucocorticoid (steroid); CHX, Cycloheximide, inhibitor of protein synthesis; GR, Glucocorticoid Receptor; TF, Transcription factor.

Moss protoplast transformation

Transient transformations were carried out in 15 ml sterile tubes with 4.5 μg of DNA from each of the appropriate constructs, 300 μl of protoplast suspension and 300 μl of 40% (w/v) polyethylene glycol 4000 (PEG4000) solution, by heat shock for 7 min at 45°C. After 10 min at room temperature, the protoplasts were progressively diluted in 6.5 ml of liquid PPNH4 medium with 0.48 M mannitol, and left in the dark. Although a signal was observed after 24 h of incubation, in this study, the protoplasts were incubated for 48 h in order to obtain a better signal (Fig. 6-5).

Flow cytometry measurements

Before flow cytometry analysis, protoplast suspensions were passed through a 50 μm mesh. Flow cytometry of live protoplasts expressing GFP was performed on a PARTEC CyFlow Space instrument (Partec S.A.R.L., Sainte Geneviève des Bois, France), using FLOMAX (Partec S.A.R.L., Sainte Geneviève des Bois, France) acquisition and analysis software, and a 488 μm solid sapphire 20 mW laser for excitation. The instrument was calibrated with calibration beads (3 μm, PARTEC reference 05-4008). The sheath fluid used was PARTEC reference 04-4007. GFP fluorescence was detected with an FITC 527 μm / 30 μm band-pass filter (FL1 channel). Red chlorophyll-based fluorescence from living protoplasts was detected with a 610 μm / 30 μm band-pass filter in the FL2 channel. The side light scatter (SSC) detector high voltage was set to 161.5 V. The photomultiplier tube voltages were adjusted to 275 V for FL1 and 475 V for FL2 (logarithmic amplification mode, four decades range, speed 4). For each sample, 40 000 events were analysed on SSC to check the size of the protoplasts and to exclude debris. The protoplasts were then gated according to FL1 and FL2, as living GFP-expressing protoplasts should always show a red chlorophyll-based fluorescence. The weighted GFP fluorescence per population of cells was calculated as the product of the average fluorescence intensity of the population of cells above the background threshold. This threshold was set arbitrarily, based on a zero DNA transformed control compared with a positive control, as the fluorescence detection on FL1 shows a marked hook.

Gating was drawn in such a way that the gated output was contaminated by non-transformed protoplasts (blank). The gate was drawn along a line of maximum GFP

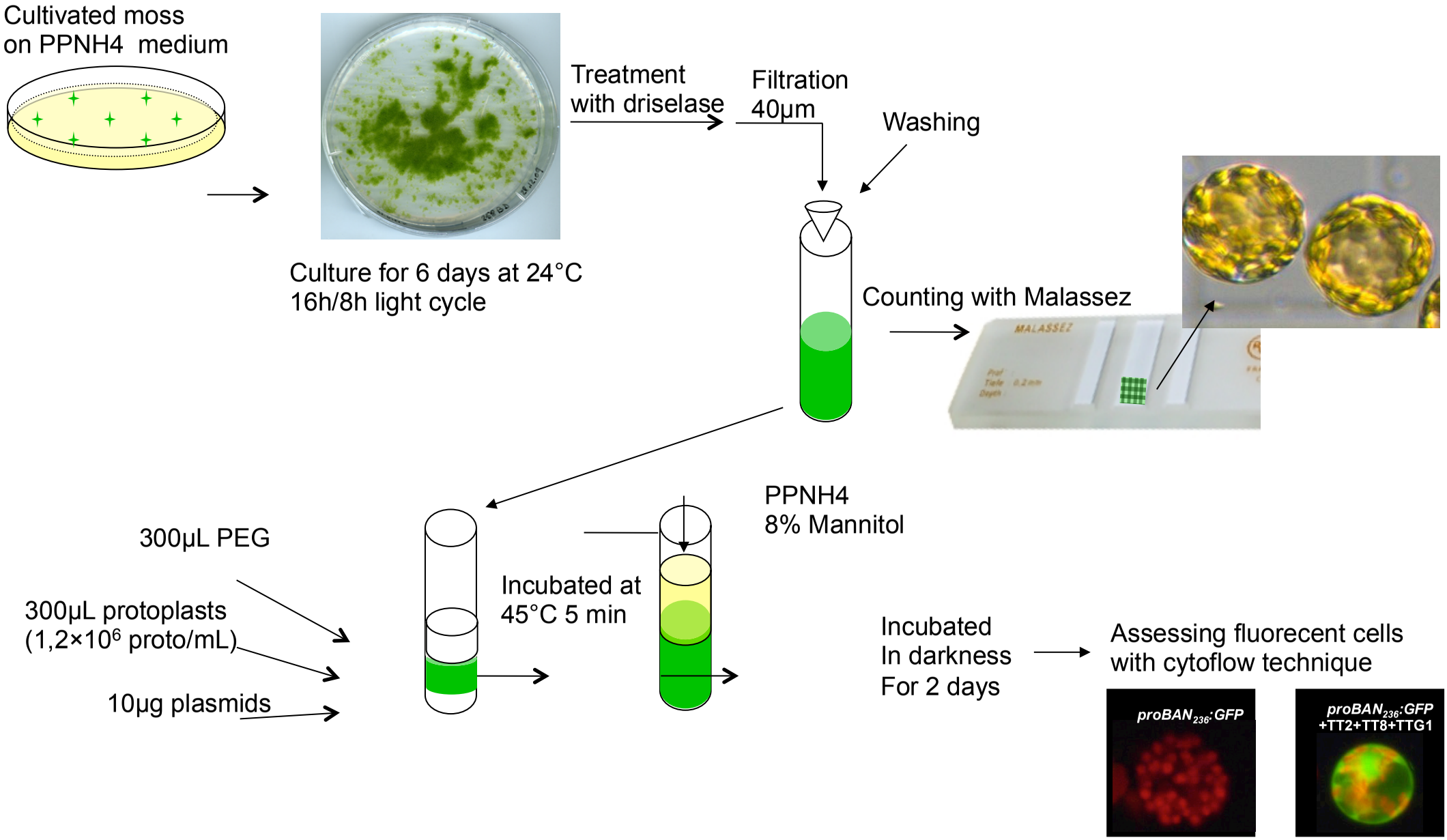


Figure 6-5. Protocol of transformation in *Physcomitrella patens* protoplasts (From JohanneTHÉVENIN & Bertrand DUBREUCQ)

intensities for positive samples, when compared with protoplasts that were only transfected with proBAN236:GFP or proBAN76:35Smini:GFP as negative controls. The gate was unchanged until all measurements in the same experiment were completed. No compensation was performed. GFP-positive cells were normally distributed.

REFERENCES

Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE. 2014. A Conserved Network of Transcriptional Activators and Repressors Regulates Anthocyanin Pigmentation in Eudicots. *Plant Cell* **26**(3): 962-980.

Albert NW, Lewis DH, Zhang H, Schwinn KE, Jameson PE, Davies KM. 2011. Members of an R2R3-MYB transcription factor family in Petunia are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning. *Plant J* **65**: 771-784.

Ali MB, McNear DH, Jr. 2014. Induced transcriptional profiling of phenylpropanoid pathway genes increased flavonoid and lignin content in Arabidopsis leaves in response to microbial products. *BMC Plant Biol* **14**: 84.

Allan AC, Hellens RP, Laing WA. 2008. MYB transcription factors that colour our fruit. *Trends Plant Sci* **13**(3): 99-102.

Appelhagen I, Huep G, Lu GH, Strompen G, Weisshaar B, Sagasser M. 2010. Weird fingers: functional analysis of WIP domain proteins. *FEBS letters* **584**: 3116-3122.

Appelhagen I, Jahns O, Bartelniewoehner L, Sagasser M, Weisshaar B, Stracke R. 2011a. Leucoanthocyanidin Dioxygenase in *Arabidopsis thaliana*: characterization of mutant alleles and regulation by MYB-BHLH-TTG1 transcription factor complexes. *Gene* **484**: 61-68.

Appelhagen I, Lu GH, Huep G, Schmelzer E, Weisshaar B, Sagasser M. 2011b. TRANSPARENT TESTA1 interacts with R2R3-MYB factors and affects early and late steps of flavonoid biosynthesis in the endothelium of *Arabidopsis thaliana* seeds. *Plant J* **67**: 406-419.

Balkunde R, Bouyer D, Hulskamp M. 2011. Nuclear trapping by GL3 controls intercellular transport and redistribution of TTG1 protein in Arabidopsis. *Development* **138**(22): 5039-5048.

Baud S, Guyon V, Kronenberger J, Wuilleme S, Miquel M, Caboche M, Lepiniec L, Rochat C. 2003. Multifunctional acetyl-CoA carboxylase 1 is essential for very long chain fatty acid elongation and embryo development in Arabidopsis. *Plant J* **33**:

75-86.

Baudry A. 2004. Analyse fonctionnelle des protéines TT2, TT8 et TTG1 :étude de leur rôle dans la régulation de la biosynthèse des flavonoïdes chez *Arabidopsis thaliana*. *PhD Thesis*.

Baudry A, Caboche M, Lepiniec L. 2006. TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. *Plant J* **46**(5): 768-779.

Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. 2004. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J* **39**(3): 366-380.

Baxter IR, Young JC, Armstrong G, Foster N, Bogenschutz N, Cordova T, Peer WA, Hazen SP, Murphy AS, Harper JF. 2005. A plasma membrane H+-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **102**(7): 2649-2654.

Bechtold N, Ellis J, Pelletier G. 1993. *In planta* infiltration of adult Arabidopsis plants. *C R Acad Sci Paris Life Sci* **316**: 1194-1199.

Bedon F, Bomal C, Caron S, Levasseur C, Boyle B, Mansfield SD, Schmidt A, Gershenzon J, Grima-Pettenati J, Seguin A, MacKay J. 2010. Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses. *J Exp Bot* **61**(14): 3847-3864.

Berger N, Dubreucq B, Roudier F, Dubos C, Lepiniec L. 2011. Transcriptional regulation of Arabidopsis LEAFY COTYLEDON2 involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *The Plant cell* **23**: 4065-4078.

Bedon F, Grima-Pettenati J, Mackay J. 2007. Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (Picea glauca). *BMC Plant Biol* **7**: 17.

Benítez M, Monk NAM, Alvarez-Buylla ER. 2011. Epidermal patterning in Arabidopsis: models make a difference. *Journal of Experimental Zoology Part B: Molecular and*

Developmental Evolution **316B**(4): 241-253.

Bernhardt C, Zhao M, Gonzalez A, Lloyd A, Schiefelbein J. 2005. The bHLH genes GL3 and EGL3 participate in an intercellular regulatory circuit that controls cell patterning in the Arabidopsis root epidermis. *Development* **132**: 291-298.

Birnboim HC, Doly J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7(6):** 1513-1523.

Bogs J, Jaffe FW, Takos AM, Walker AR, Robinson SP. 2007. The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. *Plant Physiol* **143**(3): 1347-1361.

Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C. 2000. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* **12**(12): 2383-2394.

Bouyer D, Geier F, Kragler F, Schnittger A, Pesch M, Wester K, Balkunde R, Timmer J, Fleck C, Hulskamp M. 2008. Two-dimensional patterning by a trapping/depletion mechanism: the role of TTG1 and GL3 in Arabidopsis trichome formation. *PLoS Biol* **6**(6): e141.

Broun P. 2005. Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in Arabidopsis. *Curr Opin Plant Biol* **8**(3): 272-279.

Brown BA, Headland LR, Jenkins GI. 2009. UV-B action spectrum for UVR8-mediated HY5 transcript accumulation in Arabidopsis. *Photochem Photobiol* **85**(5): 1147-1155.

Buer CS, Imin N, Djordjevic MA. 2010. Flavonoids: new roles for old molecules. *J Integr Plant Biol* **52**(1): 98-111.

Buer CS, Kordbacheh F, Truong TT, Hocart CH, Djordjevic MA. 2013. Alteration of flavonoid accumulation patterns in transparent testa mutants disturbs auxin transport, gravity responses, and imparts long-term effects on root and shoot architecture. *Planta* **238**(1): 171-189.

Burr FA, Burr B, Scheffler BE, Blewitt M, Wienand U, Matz EC. 1996. The maize repressor-like gene intensifier1 shares homology with the r1/b1 multigene family of transcription factors and exhibits missplicing. *Plant Cell* **8**(8): 1249-1259.

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EG, Hall RD, Bovy AG, Luo J, Martin C. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat Biotechnol* **26**(11): 1301-1308.

Carey CC, Strahle JT, Selinger DA, Chandler VL. 2004. Mutations in the pale aleurone color1 regulatory gene of the Zea mays anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in *Arabidopsis thaliana*. *Plant Cell* **16**(2): 450-464.

Caro E, Castellano MM, Gutierrez C. 2007. GEM, a Novel Factor in the Coordination of Cell Division to Cell Fate Decisions in the Arabidopsis Epidermis. *Plant Signal Behav* **2**(6): 494-495.

Catala R, Medina J, Salinas J. 2011. Integration of low temperature and light signaling during cold acclimation response in Arabidopsis. *Proc Natl Acad Sci U S A* **108**(39): 16475-16480.

Chang CS, Li YH, Chen LT, Chen WC, Hsieh WP, Shin J, Jane WN, Chou SJ, Choi G, Hu JM, Somerville S, Wu SH. 2008. LZF1, a HY5-regulated transcriptional factor, functions in Arabidopsis de-etiolation. *Plant J* **54**(2): 205-219.

Cheng Y, Zhu W, Chen Y, Ito S, Asami T, Wang X. 2014. Brassinosteroids control root epidermal cell fate via direct regulation of a MYB-bHLH-WD40 complex by GSK3-like kinases. *Elife* DOI: 10.7554/eLife.02525.

Cheynier V, Comte G, Davies KM, Lattanzio V, Martens S. 2013. Plant phenolics: recent advances on their biosynthesis, genetics, and ecophysiology. *Plant Physiol Biochem* **72**: 1-20.

Clegg MT, Durbin ML. 2000. Flower color variation: a model for the experimental study of evolution. *Proc Natl Acad Sci U S A* **97**(13): 7016-7023.

Cominelli E, Gusmaroli G, Allegra D, Galbiati M, Wade HK, Jenkins GI, Tonelli C. 2008. Expression analysis of anthocyanin regulatory genes in response to different light qualities in *Arabidopsis thaliana*. *J Plant Physiol* **165**(8): 886-894.

Curtis MD, Grossniklaus U. 2003. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* **133**(2): 462-469.

Dare AP, Schaffer RJ, Lin-Wang K, Allan AC, Hellens RP. 2008. Identification of a cis-regulatory element by transient analysis of co-ordinately regulated genes. *Plant Methods* **4**: 17.

Das PK1, Shin DH, Choi SB, Yoo SD, Choi G, Park YI. 2012. Cytokinins enhance sugar-induced anthocyanin biosynthesis in Arabidopsis. *Mol Cells* **34**(1): 93-101.

Datta S, Hettiarachchi C, Johansson H, Holm M. 2007. SALT TOLERANCE HOMOLOG2, a B-box protein in Arabidopsis that activates transcription and positively regulates light-mediated development. *Plant Cell* **19**(10): 3242-3255.

De Geyter N, Gholami A, Goormachtig S, Goossens A. 2012. Transcriptional machineries in jasmonate-elicited plant secondary metabolism. *Trends Plant Sci* **17**(6): 349-359.

De Masi F, Grove CA, Vedenko A, Alibés A, Gisselbrecht SS, Serrano L, Bulyk ML, Walhout AJM. 2011. Using a structural and logics systems approach to infer bHLH–DNA binding specificity determinants. *Nucleic Acids Res* **39**(11): 4553-4563.

de Vetten N, Quattrocchio F, Mol J, Koes R. 1997. The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes Dev* **11**(11): 1422-1434.

Debeaujon I, Leon-Kloosterziel KM, Koornneef M. 2000. Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. *Plant Physiol* **122**(2): 403-414.

Debeaujon I, Nesi N, Perez P, Devic M, Grandjean O, Caboche M, Lepiniec L. 2003. Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. *Plant Cell* **15**(11): 2514-2531.

Debeaujon I, Peeters AJ, Leon-Kloosterziel KM, Koornneef M. 2001. The TRANSPARENT TESTA12 gene of Arabidopsis encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *Plant Cell* **13**(4): 853-871.

DeBolt S, Scheible WR, Schrick K, Auer M, Beisson F, Bischoff V, Bouvier-Nave P, Carroll A, Hematy K, Li Y, Milne J, Nair M, Schaller H, Zemla M, Somerville C. 2009. Mutations in UDP-Glucose:sterol glucosyltransferase in Arabidopsis cause transparent testa phenotype and suberization defect in seeds. *Plant Physiol* **151**(1): 78-87.

Devic M, Guilleminot J, Debeaujon I, Bechtold N, Bensaude E, Koornneef M, Pelletier G, Delseny M. 1999. The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. *Plant J* **19**(4): 387-398.

Dixon RA, Paiva NL. 1995. Stress-Induced Phenylpropanoid Metabolism. *Plant Cell* **7**(7): 1085-1097.

Dixon RA, Steele CL. 1999. Flavonoids and isoflavonoids - a gold mine for metabolic engineering. *Trends Plant Sci* **4**(10): 394-400.

Doughty J, Aljabri M, Scott RJ. 2014. Flavonoids and the regulation of seed size in Arabidopsis. *Biochem Soc Trans* **42**(2): 364-369.

Downes BP, Stupar RM, Gingerich DJ, Vierstra RD. 2003. The HECT ubiquitin-protein ligase (UPL) family in Arabidopsis: UPL3 has a specific role in trichome development. *Plant J* **35**(6): 729-742.

Downey MO, Dokoozlian NK, Krstic MP. 2006. Cultural Practice and Environmental Impacts on the Flavonoid Composition of Grapes and Wine: A Review of Recent Research. *American Journal of Enology and Viticulture* **57**(3): 257-268.

Dubos C, Le Gourrierec J, Baudry A, Huep G, Lanet E, Debeaujon I, Routaboul JM, Alboresi A, Weisshaar B, Lepiniec L. 2008. MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. *Plant J* **55**(6): 940-953.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. 2010. MYB transcription factors in Arabidopsis. *Trends Plant Sci* **15**(10): 573-581.

El Refy A, Perazza D, Zekraoui L, Valay JG, Bechtold N, Brown S, Hulskamp M, Herzog M, Bonneville JM. 2003. The Arabidopsis KAKTUS gene encodes a HECT protein and controls the number of endoreduplication cycles. *Mol Genet Genomics* **270**(5): 403-414.

Esch JJ, Chen M, Sanders M, Hillestad M, Ndkium S, Idelkope B, Neizer J, Marks MD. 2003. A contradictory GLABRA3 allele helps define gene interactions controlling trichome development in Arabidopsis. *Development* **130**(24): 5885-5894.

Falcone Ferreyra ML, Rius SP, Casati P. 2012. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci* **3**: 222.

Fei H, Ellis BE, Vessey JK. 2013. Carbon partitioning in tissues of a gain-of-function

mutant (MYB75/PAP1-D) and a loss-of-function mutant (myb75-1) in *Arabidopsis thaliana*. *Botany* **92**(2): 93-99.

Feller A, Machemer K, Braun EL, Grotewold E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J* **66**(1): 94-116.

Feyissa DN, Lovdal T, Olsen KM, Slimestad R, Lillo C. 2009. The endogenous GL3, but not EGL3, gene is necessary for anthocyanin accumulation as induced by nitrogen depletion in Arabidopsis rosette stage leaves. *Planta* **230**(4): 747-754.

Fromont-Racine M, Rain JC, Legrain P. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat Genet* **16**(3): 277-282.

Gan L, Xia K, Chen JG, Wang S. 2011. Functional characterization of TRICHOMELESS2, a new single-repeat R3 MYB transcription factor in the regulation of trichome patterning in Arabidopsis. *BMC Plant Biol* **11**: 176.

Gao P, Li X, Cui D, Wu L, Parkin I, Gruber MY. 2010. A new dominant Arabidopsis transparent testa mutant, sk21-D, and modulation of seed flavonoid biosynthesis by KAN4. *Plant Biotechnol J* **8**(9): 979-993.

Garcia D, Fitz Gerald JN, Berger F. 2005. Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *Plant Cell* **17**(1): 52-60.

Gietz RD, Woods RA. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**: 87-96.

Gonzalez A, Mendenhall J, Huo Y, Lloyd A. 2009. TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. *Dev Biol* **325**(2): 412-421.

Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. 2008. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant J* **53**(5): 814-827.

Gou JY, Felippes FF, Liu CJ, Weigel D, Wang JW. 2011. Negative regulation of anthocyanin biosynthesis in Arabidopsis by a miR156-targeted SPL transcription factor. *Plant Cell* **23**(4): 1512-1522.

Grotewold E. 2006. The genetics and biochemistry of floral pigments. *Annu Rev Plant Biol* **57**: 761-780.

Grotewold E, Drummond BJ, Bowen B, Peterson T. 1994. The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. *Cell* **76**: 543-553.

Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, Walhout AJ. 2009. A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. *Cell* **138**: 314-327.

Grunewald W, De Smet I, Lewis DR, Lofke C, Jansen L, Goeminne G, Vanden Bossche R, Karimi M, De Rybel B, Vanholme B, Teichmann T, Boerjan W, Van Montagu MC, Gheysen G, Muday GK, Friml J, Beeckman T. 2012. Transcription factor WRKY23 assists auxin distribution patterns during Arabidopsis root development through local control on flavonol biosynthesis. *Proc Natl Acad Sci U S A* **109**: 1554-1559.

Harper JF, Manney L, Sussman MR. 1994. The plasma membrane H(+)-ATPase gene family in Arabidopsis: genomic sequence of AHA10 which is expressed primarily in developing seeds. *Mol Gen Genet* **244**(6): 572-587.

Hartmann U, Sagasser M, Mehrtens F, Stracke R, Weisshaar B. 2005. Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. *Plant Mol Biol* **57**(2): 155-171.

Haughn G, Chaudhury A. 2005. Genetic analysis of seed coat development in Arabidopsis. *Trends Plant Sci* **10**(10): 472-477.

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. 2003. The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol Biol Evol* **20**(5): 735-747.

Heppel SC, Jaffe FW, Takos AM, Schellmann S, Rausch T, Walker AR, Bogs J. 2013. Identification of key amino acids for the evolution of promoter target specificity of anthocyanin and proanthocyanidin regulating MYB factors. *Plant Mol Biol* **82**(4-5): 457-471.

Hernández I, Alegre L, Van Breusegem F, Munné-Bosch S. 2009. How relevant are flavonoids as antioxidants in plants? *Trends Plant Sci* **14**(3): 125-132.

Hernandez JM, Feller A, Morohashi K, Frame K, Grotewold E. 2007. The basic helix loop helix domain of maize R links transcriptional regulation and histone modifications by recruitment of an EMSY-related factor. *Proc Natl Acad Sci U S A* **104**(43): 17222-17227.

Hernandez JM, Heine GF, Irani NG, Feller A, Kim MG, Matulnik T, Chandler VL, Grotewold E. 2004. Different mechanisms participate in the R-dependent activity of the R2R3 MYB transcription factor C1. *J Biol Chem* **279**(46): 48205-48213.

Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V. 2011a. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J Exp Bot* **62**(8): 2465-2483.

Hichri I, Deluc L, Barrieu F, Bogs J, Mahjoub A, Regad F, Gallois B, Granier T, Trossat-Magnin C, Gomes E, Lauvergeat V. 2011b. A single amino acid change within the R2 domain of the VvMYB5b transcription factor modulates affinity for protein partners and target promoters selectivity. *BMC Plant Biol* **11**: 117.

Hichri I, Heppel SC, Pillet J, Leon C, Czemmel S, Delrot S, Lauvergeat V, Bogs J. 2010. The basic helix-loop-helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. *Mol Plant* **3**(3): 509-523.

Higo K, Ugawa Y, Iwamoto M, Korenaga T. 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic acids research* **27**: 297-300.

Hsieh LC, Lin SI, Shih AC, Chen JW, Lin WY, Tseng CY, Li WH, Chiou TJ. 2009. Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. *Plant Physiol* **151**(4): 2120-2132.

Ishida T, Hattori S, Sano R, Inoue K, Shirano Y, Hayashi H, Shibata D, Sato S, Kato T, Tabata S, Okada K, Wada T. 2007. Arabidopsis TRANSPARENT TESTA GLABRA2 is directly regulated by R2R3 MYB transcription factors and is involved in regulation of GLABRA2 transcription in epidermal differentiation. *Plant Cell* **19**(8): 2531-2543.

Ishida T, Kurata T, Okada K, Wada T. 2008. A genetic regulatory network in the development of trichomes and root hairs. *Annu Rev Plant Biol* **59**: 365-386.

Jaakola L. 2013. New insights into the regulation of anthocyanin biosynthesis in fruits.

Trends Plant Sci **18**(9): 477-483.

Jefferson RA, Kavanagh TA, Bevan MW. 1987. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO journal* **6**: 3901-3907.

Jeong SW, Das PK, Jeoung SC, Song JY, Lee HK, Kim YK, Kim WJ, Park YI, Yoo SD, Choi SB, Choi G, Park YI. 2010. Ethylene suppression of sugar-induced anthocyanin pigmentation in Arabidopsis. *Plant Physiol* **154**(3): 1514-1531.

Johnson CS, Kolevski B, Smyth DR. 2002. TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. *Plant Cell* **14**(6): 1359-1375.

Kang YH, Kirik V, Hulskamp M, Nam KH, Hagely K, Lee MM, Schiefelbein J. 2009. The MYB23 gene provides a positive feedback loop for cell fate specification in the Arabidopsis root epidermis. *Plant Cell* **21**(4): 1080-1094.

Kitamura S, Matsuda F, Tohge T, Yonekura-Sakakibara K, Yamazaki M, Saito K, Narumi I. 2010. Metabolic profiling and cytological analysis of proanthocyanidins in immature seeds of *Arabidopsis thaliana* flavonoid accumulation mutants. *Plant J* **62**(4): 549-559.

Kitamura S, Shikazono N, Tanaka A. 2004. TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in Arabidopsis. *Plant J* **37**(1): 104-114.

Kleindt CK, Stracke R, Mehrtens F, Weisshaar B. 2010. Expression analysis of flavonoid biosynthesis genes during *Arabidopsis thaliana* silique and seed development with a primary focus on the proanthocyanidin biosynthetic pathway. *BMC Res Notes* **3**: 255.

Koes R, Verweij W, Quattrocchio F. 2005. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci* **10**(5): 236-242.

Kong Q, Pattanaik S, Feller A, Werkman JR, Chai C, Wang Y, Grotewold E, Yuan L. 2012. Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. *Proc Natl Acad Sci U S A* **109**(30): E2091-2097.

Koornneef M. 1990. Mutations affecting the testa colour in Arabidopsis. *Arabidopsis*

Information Service **27**: 1-4.

Koshino-Kimura Y, Wada T, Tachibana T, Tsugeki R, Ishiguro S, Okada K. 2005. Regulation of CAPRICE transcription by MYB proteins for root epidermis differentiation in Arabidopsis. *Plant Cell Physiol* **46**(6): 817-826.

Kwon SJ, Choi EY, Seo JB, Park OK. 2007. Isolation of the Arabidopsis phosphoproteome using a biotin-tagging approach. *Mol Cells* **24**(2): 268-275.

Lai Y, Li H, Yamagishi M. 2013. A review of target gene specificity of flavonoid R2R3-MYB transcription factors and a discussion of factors contributing to the target gene selectivity. *Frontiers in Biology* **8**(6): 577-598.

Landy A. 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* **58**: 913-949.

Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamuro JK, Harada JJ, Goldberg RB. 2010. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci U S A* **107**(18): 8063-8070.

Lea US, Slimestad R, Smedvig P, Lillo C. 2007. Nitrogen deficiency enhances expression of specific MYB and bHLH transcription factors and accumulation of end products in the flavonoid pathway. *Planta* **225**(5): 1245-1253.

Lee MM, Schiefelbein J. 2001. Developmentally distinct MYB genes encode functionally equivalent proteins in Arabidopsis. *Development* **128**(9): 1539-1546.

Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M. 2006. Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol* **57**: 405-430.

Lev-Yadun S, Inbar M, Izhaki I, Ne èman G, Dafni A. 2002. Colour patterns in vegetative parts of plants deserve more research attention. *Trends Plant Sci* **7**(2): 59-60.

Lewis DR, Ramirez MV, Miller ND, Vallabhaneni P, Ray WK, Helm RF, Winkel BS, Muday GK. 2011. Auxin and ethylene induce flavonol accumulation through distinct transcriptional networks. *Plant Physiol* **156**(1): 144-164.

Li Q, Zhao P, Li J, Zhang C, Wang L, Ren Z. 2014. Genome-wide analysis of the

WD-repeat protein family in cucumber and Arabidopsis. *Mol Genet Genomics* **289**(1): 103-124.

Li S. 2014. Transcriptional control of flavonoid biosynthesis: Fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signal Behav* **8**(12).

Li S, Zachgo S. 2013. TCP3 interacts with R2R3-MYB proteins, promotes flavonoid biosynthesis and negatively regulates the auxin response in *Arabidopsis thaliana*. *Plant J* **76**(6): 901-913.

Li SF, Milliken ON, Pham H, Seyit R, Napoli R, Preston J, Koltunow AM, Parish RW. 2009. The Arabidopsis MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *Plant Cell* **21**(1): 72-89.

Li Y, Van den Ende W, Rolland F. 2014. Sucrose induction of anthocyanin biosynthesis is mediated by della. *Mol Plant* **7**(3): 570-572.

Liboz T, Bardet C, Le Van Thai A, Axelos M, Lescure B. 1990. The four members of the gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1 alpha are actively transcribed. *Plant Mol Biol* **14**(1): 107-110.

Lillo C, Lea US, Ruoff P. 2008. Nutrient depletion as a key factor for manipulating gene expression and product formation in different branches of the flavonoid pathway. *Plant Cell Environ* **31**(5): 587-601.

Liu Z, Shi MZ, Xie DY. 2014. Regulation of anthocyanin biosynthesis in *Arabidopsis thaliana* red pap1-D cells metabolically programmed by auxins. *Planta* **239**(4): 765-781.

Loreti E, Povero G, Novi G, Solfanelli C, Alpi A, Perata P. 2008. Gibberellins, jasmonate and abscisic acid modulate the sucrose-induced expression of anthocyanin biosynthetic genes in Arabidopsis. *New Phytol* **179**(4): 1004-1016.

Lovdal T, Olsen KM, Slimestad R, Verheul M, Lillo C. 2010. Synergetic effects of nitrogen depletion, temperature, and light on the content of phenolic compounds and gene expression in leaves of tomato. *Phytochemistry* **71**(5-6): 605-613.

Luceri C, Giovannelli L, Pitozzi V, Toti S, Castagnini C, Routaboul JM, Lepiniec L, Larrosa M, Dolara P. 2008. Liver and colon DNA oxidative damage and gene expression profiles of rats fed *Arabidopsis thaliana* mutant seeds containing

contrasted flavonoids. *Food Chem Toxicol* **46**(4): 1213-1220.

Luo QJ, Mittal A, Jia F, Rock CD. 2012. An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in Arabidopsis. *Plant Mol Biol* **80**(1): 117-129.

Maes L, Inze D, Goossens A. 2008. Functional specialization of the TRANSPARENT TESTA GLABRA1 network allows differential hormonal control of laminal and marginal trichome initiation in Arabidopsis rosette leaves. *Plant Physiol* **148**(3): 1453-1464.

Magnani E, Bartling L, Hake S. 2006. From Gateway to MultiSite Gateway in one recombination event. *BMC Mol Biol* **7**: 46.

Maier A, Schrader A, Kokkelink L, Falke C, Welter B, Iniesto E, Rubio V, Uhrig JF, Hulskamp M, Hoecker U. 2013. Light and the E3 ubiquitin ligase COP1/SPA control the protein stability of the MYB transcription factors PAP1 and PAP2 involved in anthocyanin accumulation in Arabidopsis. *Plant J* **74**(4): 638-651.

Marinova K, Pourcel L, Weder B, Schwarz M, Barron D, Routaboul JM, Debeaujon I, Klein M. 2007. The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H+ -antiporter active in proanthocyanidin-accumulating cells of the seed coat. *Plant Cell* **19**(6): 2023-2038.

Martens S, Preuss A, Matern U. 2010. Multifunctional flavonoid dioxygenases: flavonol and anthocyanin biosynthesis in *Arabidopsis thaliana* L. *Phytochemistry* **71**(10): 1040-1049.

Masclaux-Daubresse C, Clement G, Anne P, Routaboul JM, Guiboileau A, Soulay F, Shirasu K, Yoshimoto K. 2014. Stitching together the Multiple Dimensions of Autophagy Using Metabolomics and Transcriptomics Reveals Impacts on Metabolism, Development, and Plant Responses to the Environment in Arabidopsis. *Plant Cell* Doi: http://dx.doi.org/10.1105/tpc.114.

Matsui K, Umemura Y, Ohme-Takagi M. 2008. AtMYBL2, a protein with a single MYB domain, acts as a negative regulator of anthocyanin biosynthesis in Arabidopsis. *Plant J* **55**(6): 954-967.

McAbee JM, Hill TA, Skinner DJ, Izhaki A, Hauser BA, Meister RJ, Venugopala Reddy G, Meyerowitz EM, Bowman JL, Gasser CS. 2006. ABERRANT TESTA

SHAPE encodes a KANADI family member, linking polarity determination to separation and growth of Arabidopsis ovule integuments. *Plant J* **46**(3): 522-531.

Mehrtens F, Kranz H, Bednarek P, Weisshaar B. 2005. The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol* **138**(2): 1083-1096.

Mol J, Grotewold E, Koes R. 1998. How genes paint flowers and seeds. *Trends Plant Sci* **3**(6): 212-217.

Morishita T, Kojima Y, Maruta T, Nishizawa-Yokoi A, Yabuta Y, Shigeoka S. 2009. Arabidopsis NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. *Plant Cell Physiol* **50**(12): 2210-2222.

Morohashi K, Grotewold E. 2009. A systems approach reveals regulatory circuitry for Arabidopsis trichome initiation by the GL3 and GL1 selectors. *PLoS Genet* **5**(2): e1000396.

Nakabayashi R, Yonekura-Sakakibara K, Urano K, Suzuki M, Yamada Y, Nishizawa T, Matsuda F, Kojima M, Sakakibara H, Shinozaki K, Michael AJ, Tohge T, Yamazaki M, Saito K. 2014. Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. *Plant J* **77**(3): 367-379.

Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T. 2007. Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* **104**: 34-41.

Nemie-Feyissa D, Olafsdottir SM, Heidari B, Lillo C. 2014. Nitrogen depletion and small R3-MYB transcription factors affecting anthocyanin accumulation in Arabidopsis leaves. *Phytochemistry* **98**: 34-40.

Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L. 2000. The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. *Plant Cell* **12**(10): 1863-1878.

Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, Caboche M, Lepiniec L. 2002. The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of

the seed coat. *Plant Cell* **14**(10): 2463-2479.

Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec L. 2001. The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* **13**(9): 2099-2114.

Olsen KM, Slimestad R, Lea US, Brede C, Lovdal T, Ruoff P, Verheul M, Lillo C. 2009. Temperature and nitrogen effects on regulators and products of the flavonoid pathway: experimental and kinetic model studies. *Plant Cell Environ* **32**(3): 286-299.

Pang Y, Peel GJ, Sharma SB, Tang Y, Dixon RA. 2008. A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. *Proc Natl Acad Sci U S A* **105**(37): 14210-14215.

Patra B, Pattanaik S, Yuan L. 2013a. Proteolytic degradation of the flavonoid regulators, TRANSPARENT TESTA8 and TRANSPARENT TESTA GLABRA1, in Arabidopsis is mediated by the ubiquitin/26Sproteasome system. *Plant Signal Behav* **8**(10).

Patra B, Pattanaik S, Yuan L. 2013b. Ubiquitin protein ligase 3 mediates the proteasomal degradation of GLABROUS 3 and ENHANCER OF GLABROUS 3, regulators of trichome development and flavonoid biosynthesis in Arabidopsis. *Plant J* **74**(3): 435-447.

Patra B, Schluttenhofer C, Wu Y, Pattanaik S, Yuan L. 2013c. Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim Biophys Acta* **1829**(11): 1236-1247.

Pattanaik S, Xie CH, Yuan L. 2008. The interaction domains of the plant Myc-like bHLH transcription factors can regulate the transactivation strength. *Planta* **227**(3): 707-715.

Patzlaff A, Newman LJ, Dubos C, Whetten RW, Smith C, McInnis S, Bevan MW, Sederoff RR, Campbell MM. 2003. Characterisation of Pt MYB1, an R2R3-MYB from pine xylem. *Plant Mol Biol* **53**: 597-608.

Payne CT, Zhang F, Lloyd AM. 2000. GL3 encodes a bHLH protein that regulates trichome development in arabidopsis through interaction with GL1 and TTG1. *Genetics* **156**(3): 1349-1362.

Peer WA, Murphy AS. 2007. Flavonoids and auxin transport: modulators or regulators?

Trends Plant Sci **12**(12): 556-563.

Pelletier MK, Burbulis IE, Winkel-Shirley B. 1999. Disruption of specific flavonoid genes enhances the accumulation of flavonoid enzymes and end-products in Arabidopsis seedlings. *Plant Mol Biol* **40**(1): 45-54.

Peng M, Hudson D, Schofield A, Tsao R, Yang R, Gu H, Bi YM, Rothstein SJ. 2008. Adaptation of Arabidopsis to nitrogen limitation involves induction of anthocyanin synthesis which is controlled by the NLA gene. *J Exp Bot* **59**(11): 2933-2944.

Pesch M, Hulskamp M. 2009. One, two, three...models for trichome patterning in Arabidopsis? *Curr Opin Plant Biol* **12**(5): 587-592.

Petroni K, Tonelli C. 2011. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci* **181**(3): 219-229.

Pires N, Dolan L. 2010. Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol Biol Evol* **27**(4): 862-874.

Popper ZA, Fry SC. 2003. Primary cell wall composition of bryophytes and charophytes. *Ann Bot* **91**(1): 1-12.

Potter H. 1993. Application of electroporation in recombinant DNA technology. *Methods Enzymol* **217:** 461-478.

Pourcel L, Routaboul JM, Cheynier V, Lepiniec L, Debeaujon I. 2007. Flavonoid oxidation in plants: from biochemical properties to physiological functions. *Trends Plant Sci* **12**(1): 29-36.

Pourcel L, Routaboul JM, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I. 2005. TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. *Plant Cell* **17**(11): 2966-2980.

Pourtau N, Jennings R, Pelzer E, Pallas J, Wingler A. 2006. Effect of sugar-induced senescence on gene expression and implications for the regulation of senescence in Arabidopsis. *Planta* **224**(3): 556-568.

Prouse MB, Campbell MM. 2012. The interaction between MYB proteins and their target DNA binding sites. *Biochim Biophys Acta* **1819**(1): 67-77.

Prouse MB, Campbell MM. 2013. Interactions between the R2R3-MYB transcription factor, AtMYB61, and target DNA binding sites. *Plos One* **8**(5): e65132.

Proust H, Hoffmann B, Xie X, Yoneyama K, Schaefer DG, Yoneyama K, Nogue F, Rameau C. 2011. Strigolactones regulate protonema branching and act as a quorum sensing-like signal in the moss Physcomitrella patens. *Development* **138**(8): 1531-1539.

Qi T, Huang H, Wu D, Yan J, Qi Y, Song S, Xie D. 2014. Arabidopsis DELLA and JAZ Proteins Bind the WD-Repeat/bHLH/MYB Complex to Modulate Gibberellin and Jasmonate Signaling Synergy. *Plant Cell* **26**(3): 1118-1133.

Qi T, Song S, Ren Q, Wu D, Huang H, Chen Y, Fan M, Peng W, Ren C, Xie D. 2011. The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. *Plant Cell* **23**(5): 1795-1814.

Quattrocchio F, Baudry A, Lepiniec L, Grotewold E. 2006. The Regulation of Flavonoid Biosynthesis. *The Science of Flavonoids*, pp97-122.

Quattrocchio F, Verweij W, Kroon A, Spelt C, Mol J, Koes R. 2006. PH4 of Petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell* **18**(5): 1274-1291.

Ramsay NA, Glover BJ. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci* **10**(2): 63-70.

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G. 2000. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290**(5499): 2105-2110.

Rishmawi L, Pesch M, Juengst C, Schauss AC, Schrader A, Hulskamp M. 2014. Non-Cell-Autonomous Regulation of Root Hair Patterning Genes by WRKY75 in Arabidopsis. *Plant Physiol* **165**(1): 186-195.

Rock CD. 2013. Trans-acting small interfering RNA4: key to nutraceutical synthesis in grape development? *Trends Plant Sci* **18**(11): 601-610.

Rolland F, Baena-Gonzalez E, Sheen J. 2006. Sugar sensing and signaling in plants:

conserved and novel mechanisms. *Annu Rev Plant Biol* **57**: 675-709.

Romano JM, Dubos C, Prouse MB, Wilkins O, Hong H, Poole M, Kang KY, Li E, Douglas CJ, Western TL, Mansfield SD, Campbell MM. 2012. AtMYB61, an R2R3-MYB transcription factor, functions as a pleiotropic regulator via a small gene network. *New Phytol* **195**(4):774-786.

Roszak P, Kohler C. 2011. Polycomb group proteins are required to couple seed coat initiation to fertilization. *Proc Natl Acad Sci U S A* **108**(51): 20826-20831.

Roth BA, Goff SA, Klein TM, Fromm ME. 1991. C1- and R-dependent expression of the maize Bz1 gene requires sequences with homology to mammalian myb and myc binding sites. *Plant Cell* **3**(3): 317-325.

Routaboul JM, Dubos C, Beck G, Marquis C, Bidzinski P, Loudet O, Lepiniec L. 2012. Metabolite profiling and quantitative genetics of natural variation for flavonoids in Arabidopsis. *J Exp Bot* **63**(10): 3749-3764.

Rowan DD, Cao M, Lin-Wang K, Cooney JM, Jensen DJ, Austin PT, Hunt MB, Norling C, Hellens RP, Schaffer RJ, Allan AC. 2009. Environmental regulation of leaf colour in red 35S:PAP1 *Arabidopsis thaliana*. *New Phytol* **182**(1): 102-115.

Rubin G, Tohge T, Matsuda F, Saito K, Scheible WR. 2009. Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in Arabidopsis. *Plant Cell* **21**(11): 3567-3584.

Ryu KH, Kang YH, Park YH, Hwang I, Schiefelbein J, Lee MM. 2005. The WEREWOLF MYB protein directly regulates CAPRICE transcription during cell fate specification in the Arabidopsis root epidermis. *Development* **132**(21): 4765-4775.

Sagasser M, Lu GH, Hahlbrock K, Weisshaar B. 2002. A. thaliana TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. *Genes Dev* **16**(1): 138-149.

Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, Fernie AR. 2013. The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. *Plant Physiol Biochem* **72**: 21-34.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**(12): 5463-5467.

Savage NS, Walker T, Wieckowski Y, Schiefelbein J, Dolan L, Monk NA. 2008. A mutual support mechanism through intercellular movement of CAPRICE and GLABRA3 can pattern the Arabidopsis root epidermis. *PLoS Biol* **6**(9): e235.

Schaart JG, Dubos C, Romero De La Fuente I, van Houwelingen AM, de Vos RC, Jonker HH, Xu W, Routaboul JM, Lepiniec L, Bovy AG. 2013. Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (Fragaria x ananassa) fruits. *New Phytol* **197**(2): 454-467.

Schena M, Picard D, Yamamoto KR. 1991. Vectors for constitutive and inducible gene expression in yeast. *Methods Enzymol* **194**: 389-398.

Schiefelbein J, Huang L, Zheng X. 2014. Regulation of epidermal cell fate in Arabidopsis roots: the importance of multiple feedback loops. *Front Plant Sci* **5**: 47.

Schiefelbein J, Kwak SH, Wieckowski Y, Barron C, Bruex A. 2009. The gene regulatory network for root epidermal cell-type pattern formation in Arabidopsis. *J Exp Bot* **60**(5): 1515-1521.

Schneitz K, Hülskamp M, Pruitt RE. 1995. Wild-type ovule development in *Arabidopsis thaliana*: a light microscope study of cleared whole-mount tissue. *Plant J* **7**(5): 731-749.

Selinger DA, Chandler VL. 1999. A mutation in the pale aleurone color1 gene identifies a novel regulator of the maize anthocyanin pathway. *Plant Cell* **11**(1): 5-14.

Serna L, Martin C. 2006. Trichomes: different regulatory networks lead to convergent structures. *Trends Plant Sci* **11**(6): 274-280.

Sharma SB, Dixon RA. 2005. Metabolic engineering of proanthocyanidins by ectopic expression of transcription factors in *Arabidopsis thaliana*. *Plant J* **44**(1): 62-75.

Shi M, Xie D. 2010. Features of anthocyanin biosynthesis in *pap1-D* and wild-type *Arabidopsis thaliana* plants grown in different light intensity and culture media conditions. *Planta* **231**(6): 1385-1400.

Shi M, Xie D. 2011. Engineering of red cells of *Arabidopsis thaliana* and comparative genome-wide gene expression analysis of red cells versus wild-type cells. *Planta* **233**(4):787-805.

Shi M, Xie D. 2014. Biosynthesis and Metabolic Engineering of Anthocyanins in Arabidopsis. *Recent Pat Biotechnol* **8**(1): 1872-2083.

Shin DH, Choi M, Kim K, Bang G, Cho M, Choi SB, Choi G, Park YI. 2013. HY5 regulates anthocyanin biosynthesis by inducing the transcriptional activation of the MYB75/PAP1 transcription factor in Arabidopsis. *FEBS Lett* **587**(10): 1543-1547.

Shin J, Park E, Choi G. 2007. PIF3 regulates anthocyanin biosynthesis in an HY5-dependent manner with both factors directly binding anthocyanin biosynthetic gene promoters in Arabidopsis. *Plant J* **49**(6): 981-994.

Shirley BW. Flavonoid biosynthesis: 'new' functions for an 'old' pathway. *Trends Plant Sci* **1**(11): 377-382.

Sieber P, Gheyselinck J, Gross-Hardt R, Laux T, Grossniklaus U, Schneitz K. 2004. Pattern formation during early ovule development in *Arabidopsis thaliana*. *Dev Biol* **273**(2): 321-334.

Sobel JM, Streisfeld MA. 2013. Flower color as a model system for studies of plant evo-devo. *Front Plant Sci* **4**: 321.

Solfanelli C, Poggi A, Loreti E, Alpi A, Perata P. 2006. Sucrose-specific induction of the anthocyanin biosynthetic pathway in Arabidopsis. *Plant Physiol* **140**(2): 637-646.

Song SK, Ryu KH, Kang YH, Song JH, Cho YH, Yoo SD, Schiefelbein J, Lee MM. 2011. Cell fate in the Arabidopsis root epidermis is determined by competition between WEREWOLF and CAPRICE. *Plant Physiol* **157**(3): 1196-1208.

Spelt C, Quattrocchio F, Mol J, Koes R. 2002. ANTHOCYANIN1 of petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms. *Plant Cell* **14**(9): 2121-2135.

Spelt C, Quattrocchio F, Mol JN, Koes R. 2000. anthocyanin1 of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. *Plant Cell* **12**(9): 1619-1632.

Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B. 2007. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J* **50**(4): 660-677.

Stracke R, Jahns O, Keck M, Tohge T, Niehaus K, Fernie AR, Weisshaar B. 2010. Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in *Arabidopsis thaliana* plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. *New Phytol* **188**(4): 985-1000.

Stracke R, Werber M, Weisshaar B. 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol* **4**(5): 447-456.

Symonds VV, Hatlestad G, Lloyd AM. 2011. Natural allelic variation defines a role for ATMYC1: trichome cell fate determination. *PLoS Genet* **7**(6): e1002069.

Teng S, Keurentjes J, Bentsink L, Koornneef M, Smeekens S. 2005. Sucrose-specific induction of anthocyanin biosynthesis in Arabidopsis requires the MYB75/PAP1 gene. *Plant Physiol* **139**(4): 1840-1852.

Terrier N, Torregrosa L, Ageorges A, Vialet S, Verries C, Cheynier V, Romieu C. 2009. Ectopic expression of VvMybPA2 promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. *Plant Physiol* **149**(2): 1028-1041.

Thevenin J, Dubos C, Xu W, Le Gourrierec J, Kelemen Z, Charlot F, Nogue F, Lepiniec L, Dubreucq B. 2012. A new system for fast and quantitative analysis of heterologous gene expression in plants. *New Phytol* **193**(2): 504-512.

Tohge T, Fernie AR. 2012. Co-expression and co-responses: within and beyond transcription. *Front Plant Sci* **3**: 248.

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, Noji M, Yamazaki M, Saito K. 2005. Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant J* **42**(2): 218-235.

Tohge T, Watanabe M, Hoefgen R, Fernie AR. 2013. The evolution of phenylpropanoid metabolism in the green lineage. *Crit Rev Biochem Mol Biol* **48**(2): 123-152.

Toledo-Ortiz G, Huq E, Quail PH. 2003. The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* **15**(8): 1749-1770.

Tominaga-Wada R, Iwata M, Nukumizu Y, Sano R, Wada T. 2012. A full-length R-like basic-helix-loop-helix transcription factor is required for anthocyanin upregulation whereas the N-terminal region regulates epidermal hair formation. *Plant Sci* **183**: 115-122.

Tominaga-Wada R, Wada T. 2014. Regulation of root hair cell differentiation by R3 MYB transcription factors in tomato and. *Front Plant Sci* **5**: 91.

van Nocker S, Ludwig P. 2003. The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. *BMC Genomics* **4**(1): 50.

Vandenbussche F, Habricot Y, Condiff AS, Maldiney R, Straeten DVD, Ahmad M. 2007. HY5 is a point of convergence between cryptochrome and cytokinin signalling pathways in *Arabidopsis thaliana*. *Plant J* **49**(3): 428-441.

Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC. 1999. The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. *Plant Cell* **11**(7): 1337-1350.

Wang S, Chen JG. 2014. Regulation of cell fate determination by single-repeat R3 MYB transcription factors in Arabidopsis. *Front Plant Sci* **5**: 133.

Wang S, Hubbard L, Chang Y, Guo J, Schiefelbein J, Chen JG. 2008. Comprehensive analysis of single-repeat R3 MYB proteins in epidermal cell patterning and their transcriptional regulation in Arabidopsis. *BMC plant biology* **8**: 81.

Wang S, Kwak SH, Zeng Q, Ellis BE, Chen XY, Schiefelbein J, Chen JG. 2007. TRICHOMELESS1 regulates trichome patterning by suppressing GLABRA1 in Arabidopsis. *Development* **134**(21): 3873-3882.

Wester K, Digiuni S, Geier F, Timmer J, Fleck C, Hulskamp M. 2009. Functional diversity of R3 single-repeat genes in trichome development. *Development* **136**(9): 1487-1496.

Weston K. 1992. Extension of the DNA binding consensus of the chicken c-Myb and v-Myb proteins. *Nucleic acids research* **20**: 3043-3049.

Winkel-Shirley B. 2001a. Flavonoid biosynthesis. A colorful model for genetics,

biochemistry, cell biology, and biotechnology. *Plant Physiol* **126**(2): 485-493.

Winkel-Shirley B. 2001b. It takes a garden. How work on diverse plant species has contributed to an understanding of flavonoid metabolism. *Plant Physiol* **127**(4): 1399-1404.

Winkel-Shirley B. 2002. Biosynthesis of flavonoids and effects of stress. *Curr Opin Plant Biol* **5**(3): 218-223.

Winkel BS. 2004. Metabolic channeling in plants. *Annu Rev Plant Biol* **55**: 85-107.

Xie D, Sharma SB, Wright E, Wang Z, Dixon RA. 2006. Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. *Plant J* **45**(6): 895-907.

Xu W, Grain D, Le Gourrierec J, Harscoet E, Berger A, Jauvion V, Scagnelli A, Berger N, Bidzinski P, Kelemen Z, Salsac F, Baudry A, Routaboul JM, Lepiniec L, Dubos C. 2013. Regulation of flavonoid biosynthesis involves an unexpected complex transcriptional regulation of *TT8* expression, in Arabidopsis. *New Phytol* **198**(1): 59-70.

Xu W, Grain D, Bobet S, Le Gourrierec J, Thevenin J, Kelemen Z, Lepiniec L, Dubos C. 2014a. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in Arabidopsis seed. *New Phytol* **202**(1): 132-144.

Xu W, Lepiniec L, Dubos C. 2014b. New insights toward the transcriptional engineering of proanthocyanidin biosynthesis. *Plant Signal Behav* **9**(4).

Xue B, Charest PJ, Devantier Y, Rutledge RG. 2003. Characterization of a MYBR2R3 gene from black spruce (*Picea mariana*) that shares functional conservation with maize C1. *Mol Genet Genomics* **270**(1): 78-86.

Yang F, Cai J, Yang Y, Liu Z. 2013. Overexpression of microRNA828 reduces anthocyanin accumulation in Arabidopsis. *Plant Cell Tiss Organ Cult* **115**(2): 159-167.

Ye H, Li L, Guo H, Yin Y. 2012. MYBL2 is a substrate of GSK3-like kinase BIN2 and acts as a corepressor of BES1 in brassinosteroid signaling pathway in Arabidopsis. *Proc Natl Acad Sci U S A* **109**(49): 20142-20147.

Yonekura-Sakakibara K, Tohge T, Niida R, Saito K. 2007. Identification of a flavonol

7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. *J Biol Chem* **282**(20): 14932-14941.

Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A. 2003. A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. *Development* **130**(20): 4859-4869.

Zhao H, Wang X, Zhu D, Cui S, Li X, Cao Y, Ma L. 2012. A single amino acid substitution in IIIf subfamily of basic helix-loop-helix transcription factor AtMYC1 leads to trichome and root hair patterning defects by abolishing its interaction with partner proteins in Arabidopsis. *J Biol Chem* **287**(17): 14109-14121.

Zhao M, Morohashi K, Hatlestad G, Grotewold E, Lloyd A. 2008. The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. *Development* **135**(11): 1991-1999.

Zhou LL, Shi MZ, Xie DY. 2012. Regulation of anthocyanin biosynthesis by nitrogen in TTG1-GL3/TT8-PAP1-programmed red cells of *Arabidopsis thaliana*. *Planta* **236**(3): 825-837.

Zhu HF, Fitzsimmons K, Khandelwal A, Kranz RG. 2009. CPC, a single-repeat R3 MYB, is a negative regulator of anthocyanin biosynthesis in Arabidopsis. *Mol Plant* **2**(4): 790-802.

Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF. 2004. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J* **40**(1): 22-34.

ANNEXES

Article 4: A new system for fast and quantitative analysis of heterologous gene expression in plants

J. Thévenin[1*], C. Dubos[1*], W. Xu[1], J. Le Gourrierec[2], Z. Kelemen[1], F. Charlot[1], F. Nogué[1], L. Lepiniec[1] and B. Dubreucq[1]

Accepted for the publication in New Phytologist (2012, vol.193)

[1] INRA AgroParisTech, IJPB, UMR 1318, INRA centre de Versailles, route de Saint Cyr, 78026 Versailles CEDEX, France
[2] Université d'Angers, INRA, INH UMR A462 SAGAH, Morphogenèse des ligneux, 2 Bd Lavoisier, 49045 Angers CEDEX1, France

[*] These authors contributed equally to this work.

Author for correspondence:
B. Dubreucq
Tel: +33 (0) 1 3083 3066
Email: bertrand.dubreucq@versailles.inra.fr

New Phytologist

Research

Methods

A new system for fast and quantitative analysis of heterologous gene expression in plants

J. Thévenin[1]*, C. Dubos[1]*, W. Xu[1], J. Le Gourrierec[2], Z. Kelemen[1], F. Charlot[1], F. Nogué[1], L. Lepiniec[1] and B. Dubreucq[1]

[1]INRA AgroParisTech, IJPB, UMR 1318, INRA centre de Versailles, route de Saint Cyr, 78026 Versailles CEDEX, France; [2]Université d'Angers, INRA, INH UMR A462 SAGAH, Morphogenèse des ligneux, 2 Bd Lavoisier, 49045 Angers CEDEX1, France

Author for correspondence:
B. Dubreucq
Tel: +33 1 30833066
Email: bertrand.dubreucq@versailles.inra.fr

Received: *21 July 2011*
Accepted: *13 September 2011*

New Phytologist (2011)
doi: 10.1111/j.1469-8137.2011.03936.x

Key words: *cis*-element, cytometry, flavonoids, *Physcomitrella patens*, protoplasts, transcription factor.

Summary

- Large-scale analysis of transcription factor–*cis*-acting element interactions in plants, or the dissection of complex transcriptional regulatory mechanisms, requires rapid, robust and reliable systems for the quantification of gene expression.
- Here, we describe a new system for transient expression analysis of transcription factors, which takes advantage of the fast and easy production and transfection of *Physcomitrella patens* protoplasts, coupled to flow cytometry quantification of a fluorescent protein (green fluorescent protein). Two small-sized and high-copy Gateway® vectors were specifically designed, although standard binary vectors can also be employed.
- As a proof of concept, the regulation of *BANYULS* (*BAN*), a key structural gene involved in proanthocyanidin biosynthesis in *Arabidopsis thaliana* seeds, was used. In *P. patens*, *BAN* expression is activated by a complex composed of three proteins (TT2/AtMYB123, TT8/bHLH042 and TTG1), and is inhibited by MYBL2, a transcriptional repressor, as in *Arabidopsis*. Using this approach, two new regulatory sequences that are necessary and sufficient for specific *BAN* expression in proanthocyanidin-accumulating cells were identified.
- This one hybrid-like plant system was successfully employed to quantitatively assess the transcriptional activity of four regulatory proteins, and to identify their target recognition sites on the *BAN* promoter.

Introduction

Transcription factors (TFs) are regulatory proteins found in all prokaryotic and eukaryotic organisms, including plants. TFs have been shown to control numerous facets of plant growth and development through the coordinated regulation of gene expression (Dubos *et al.*, 2010). These regulatory mechanisms involve specific interactions between TFs and DNA motifs. To characterize TFs, their target DNA sequences are identified and their transcriptional activity is determined (i.e. activator or repressor). To identify the *cis*-regulatory elements that are targeted by specific TFs, a large number of constructs containing numerous versions of a promoter, with, in some cases, various combinations of regulators, are tested, as TFs can act alone or in combination with other protein partners.

The yeast one-hybrid system is a simple and efficient method for determining the relationships between proteins and target DNA. However, only a limited number of TFs can be tested at the same time because of the lack of selective markers (generally three; see Baudry *et al.*, 2004). Furthermore, as a heterologous system, some yeast post-translational protein modifications can differ from those occurring in plants. Transient plant systems have been used extensively to study plant TF–promoter interactions because these approaches are fast and independent of any integration events into the genome (no position effects). These methods can be grouped into two categories based on the transformation protocol: the first relies on protoplast transfection and the second on tissues/cell transformation using agrobacteria. Protoplast preparation is time consuming and labour intensive (Hartmann *et al.*, 1998). Leaf (*Nicotiana benthamiana*) or cotyledon infiltration is an efficient method, but is not high throughput (Berger *et al.*, 2007; Marion *et al.*, 2008). However, plant cultured cells may be suitable for high-throughput experiments (Berger *et al.*, 2007; Marion *et al.*, 2008), but this involves maintaining the cell cultures for periods of time, which is not a trivial task (Fukuda *et al.*, 1994).

*These authors contributed equally to this work.

Article 5: Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria* × *ananassa*) fruits

Jan G. Schaart[1*], Christian Dubos[2,3*], Irene Romero De La Fuente[1], Adèle M. M. L. van Houwelingen[4], Ric C. H.de Vos[4,5], Harry H. Jonker[4,5], Wenjia Xu[2,3], Jean-Marc Routaboul[2,3], Loïc Lepiniec[2,3] and Arnaud G. Bovy[1,5]

Accepted for the publication in New Phytologist (2013, vol.197)

[1] Wageningen UR Plant Breeding, Wageningen University and Research Centre, PO Box 16, 6700 AA, Wageningen, the Netherlands
[2] INRA, Institut Jean-Pierre Bourgin, Saclay Plant Sciences, RD10, F-78026, Versailles, France
[3] AgroParisTech, Institut Jean-Pierre Bourgin, Saclay Plant Sciences, RD10, F-78026, Versailles, France
[4] Plant Research International, Business Unit Bioscience, Wageningen University and Research Centre, PO Box 16, 6700 AA, Wageningen, the Netherlands;
[5] Centre for Biosystems Genomics, PO Box 98, 6700 AB, Wageningen, the Netherlands

[*] These authors contributed equally to this work.

Author for correspondence:
Jan G. Schaart
Tel: +31 317480963
Email: jan.schaart@wur.nl

New Phytologist

Research

Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria* × *ananassa*) fruits

Jan G. Schaart[1]*, Christian Dubos[2,3]*, Irene Romero De La Fuente[1], Adèle M. M. L. van Houwelingen[4], Ric C. H.de Vos[4,5], Harry H. Jonker[4,5], Wenjia Xu[2,3], Jean-Marc Routaboul[2,3], Loïc Lepiniec[2,3] and Arnaud G. Bovy[1,5]

[1]Wageningen UR Plant Breeding, Wageningen University and Research Centre, PO Box 16, 6700 AA, Wageningen, the Netherlands; [2]INRA, Institut Jean-Pierre Bourgin, Saclay Plant Sciences, RD10, F-78026, Versailles, France; [3]AgroParisTech, Institut Jean-Pierre Bourgin, Saclay Plant Sciences, RD10, F-78026, Versailles, France; [4]Plant Research International, Business Unit Bioscience, Wageningen University and Research Centre, PO Box 16, 6700 AA, Wageningen, the Netherlands; [5]Centre for Biosystems Genomics, PO Box 98, 6700 AB, Wageningen, the Netherlands

Author for correspondence:
Jan G. Schaart
Tel: +31 317480963
Email: jan.schaart@wur.nl

Received: *17 July 2012*
Accepted: *23 September 2012*

New Phytologist (2012)
doi: 10.1111/nph.12017

Key words: bHLH, flavonoid biosynthesis, MYB, strawberry (*Fragaria* × *ananassa*), WD40-repeat, yeast-two-hybrid.

Summary

- Strawberry (*Fragaria* × *ananassa*) fruits contain high concentrations of flavonoids. In unripe strawberries, the flavonoids are mainly represented by proanthocyanidins (PAs), while in ripe fruits the red-coloured anthocyanins also accumulate. Most of the structural genes leading to PA biosynthesis in strawberry have been characterized, but no information is available on their transcriptional regulation. In *Arabidopsis thaliana* the expression of the PA biosynthetic genes is specifically induced by a ternary protein complex, composed of AtTT2 (AtMYB123), AtTT8 (AtbHLH042) and AtTTG1 (WD40-repeat protein).
- A strategy combining yeast-two-hybrid screening and agglomerative hierarchical clustering of transcriptomic and metabolomic data was undertaken to identify strawberry PA regulators.
- Among the candidate genes isolated, four were similar to *AtTT2*, *AtTT8* and *AtTTG1* (*FaMYB9/FaMYB11*, *FabHLH3* and *FaTTG1*, respectively) and two encode putative negative regulators (*FaMYB5* and *FabHLH3*Δ). Interestingly, *FaMYB9/FaMYB11*, *FabHLH3* and *FaTTG1* were found to complement the *tt2-1*, *tt8-3* and *ttg1-1 transparent testa* mutants, respectively. In addition, they interacted in yeast and activated the *Arabidopsis BANYULS* (anthocyanidin reductase) gene promoter when coexpressed in *Physcomitrella patens* protoplasts.
- Taken together, these results demonstrated that *FaMYB9/FaMYB11*, *FabHLH3* and *FaTTG1* are the respective functional homologues of *AtTT2*, *AtTT8* and *AtTTG1*, providing new tools for modifying PA content and strawberry fruit quality.

Introduction

The popularity of strawberry as a fruit crop is mainly a result of its unique aroma, sweet taste, bright colour and nutritional value. These quality traits are largely determined by the metabolic composition of the fruit. For example, strawberry fruits are known as a valuable source of polyphenol compounds. The main polyphenol compounds found in strawberry fruits are flavonoids. About 70% of the total phenolics that accumulated in cv Queen Elisa ripe fruits correspond to proanthocyanidins (PAs), whereas anthocyanins, flavonols and the other phenolics (i.e. *p*-coumaric acid, ellagic acid) represent *c.* 20, 3 and 7%, respectively (Almeida *et al.*, 2007; Carbone *et al.*, 2009). Flavonoids have potential human health beneficial properties. They may act as antioxidants or as signalling molecules and were shown to be able to modulate some cellular functions (Nijveldt *et al.*, 2001; Battino *et al.*, 2009). Flavonoids such as PAs are also known to impact fruit flavour and astringency (Akagi *et al.*, 2011).

The flavonoid biosynthetic pathway has been extensively studied in numerous plant species at the genetic, biochemical and molecular levels, and was recently described in strawberry (Lepiniec *et al.*, 2006; Almeida *et al.*, 2007; Carbone *et al.*, 2009). This pathway (Fig. 1a) is initiated by the condensation of one molecule of 4-coumaroyl-CoA with three molecules of malonyl-CoA by the chalcone synthase (CHS) enzyme. The successive steps catalysed by the chalcone isomerase (CHI), flavonoid 3-hydroxylase (F3H/FHT), flavonoid 3′-hydroxylase (F3′H),

*These authors contributed equally to this work.