UNIVERSITA' DEGLI STUDI DI VERONA
FACOLTA' DI SCIENZE MM.FF.NN.
DIPARTIMENTO DI BIOTECNOLOGIE

DOTTORATO DI RICERCA IN
BIOTECNOLOGIE MOLECOLARI, INDUSTRIALI ED AMBIENTALI - XXII CICLO

AGR/07

CHARACTERIZATION OF THE *Arabidopsis thaliana MYB59* TRANSCRIPTION FACTOR

Coordinatore: Prof. Roberto Bassi

Tutor: Prof.ssa Antonella Furini

Dottoranda: Silvia Maistri

CONTENTS

ABBREVIATIONS

°C	Degree Celsius
35SCaMV	35S promoter of the Cauliflower Mosaic virus
aa	Aminoacid
ABA	Abscisic acid
AS	Alternative splicing
Asp	Aspartate
bp	Base pair
Cd	Cadmium
cDNA	Complementary deoxyribonucleic acid
Da	Dalton
DNA	Deoxyribonucleic acid
Ds element	Dissociator element
g	Gram
GA_3	Giberellin
GUS	β-glucuronidase
h	Hour
IAA	Indoleacetic acid
IPTG	Isopropil β-D-1-tiogalattopiranoside
l	Liter
LB	Left border
Lys	Lysine
M	Molarity
min	Minute
mRNA	Messenger RNA
NLS	Nuclear localization signal
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
RB	Right border
RdDM	RNA-mediated DNA methylation
RNA	Ribonucleic acid

rpm	Revolution per minute
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SE	Standard error
sRNA	small RNA
siRNA	small interfering RNA
T-DNA	Transfer-DNA
TF	Transcription factor
Tyr	Tyrosine
uORF	Upstream ORF
UTR	Untranslated region
v/v	Volume per volume
WT	Wild-type
PVDF	Polyvinylidene fluoride

RIASSUNTO

Le piante sono frequentemente sottoposte a stress, ossia a condizioni esterne che influiscono negativamente sulla crescita e sullo sviluppo. La percezione dello stress e la trasduzione del segnale che portano all'attivazione di risposte adattive sono passaggi cruciali nel determinare la sopravvivenza della pianta. Le piante possono rispondere ad uno stimolo esterno attivando meccanismi di difesa specifici per un determinato stress, oppure meccanismi in grado di rispondere a più stress. Infatti, le varie vie di trasduzione del segnale sono spesso interconnesse fra loro a vari livelli, possono condividere uno o più componenti o intermedi oppure avere output comuni (Chinnusamy *et al*., 2003). Le informazioni sul cross-talk comunque sono ancora limitate, in quanto poco è noto riguardo ai recettori e agli intermedi delle vie di trasduzione del segnale. A valle della cascata di reazioni si trovano i fattori di trascrizione che sono in grado di regolare l'espressione genica legandosi al DNA in maniera sequenza-specifica. Tali proteine sono normalmente espresse in maniera tessuto-specifica, stadio di sviluppo-specifica o in modo dipendente da uno stimolo esterno, come ad esempio uno stress ambientale, e sono responsabili della selettività nella regolazione genica (Zhang, 2003). Quindi, caratterizzare i fattori di trascrizione può essere un passo importante verso la comprensione dei meccanismi che stanno alla base dei processi di sviluppo e di risposta agli stress.

La famiglia di fattori di trascrizione più rappresentata in pianta è la famiglia MYB (Stracke *et al*., 2001). I membri di tale famiglia sono caratterizzati dalla presenza di un dominio strutturalmente conservato di circa 52 aminoacidi (dominio MYB), contenente tre residui di triptofano spaziati regolarmente. Tali proteine sono in grado di legare il DNA in maniera sequenza specifica adottando una conformazione *helix-turn-helix*. I diversi membri della famiglia vengono identificati sulla base del numero di *repeats* imperfette di tale dominio, che può essere ripetuto fino a tre volte. Sono state proposte diverse funzioni per i diversi fattori MYB: alcuni sono coinvolti nel controllo del metabolismo secondario, della proliferazione e del differenziamento di alcuni tipi di cellule, mentre altri sono necessari nelle vie di trasduzione del segnale che rispondono a differenti stimoli (Martin e Paz-Ares, 1997). In particolare, i fattori di trascrizione MYB sembrano essere importanti nella mediazione delle risposte a molecole segnale come l'acido salicilico, e ad ormoni come l'acido abscissico e l'acido gibberellico.

Lo scopo di questo progetto di dottorato è stato quello di caratterizzare il gene *MYB59* di *A. thaliana* che appartiene alla famiglia R2R3-MYB ed è presente in tre varianti di *splicing*, indicate per semplicità *MYB59.1*, *MYB59.2* e *MYB59.3*. È noto che il gene in esame è coinvolto nella

risposta allo stress da cadmio (Li *et al*., 2006) ed, in particolare, è stato dimostrato che l'espressione del gene omologo in *B. juncea* viene modulata dopo l'esposizione a tale metallo (Fusco *et al*., 2005). Tramite analisi *Real-Time* PCR, condotta su piante di *Arabidopsis WT* sottoposte a trattamento con Cd, è stato confermato un coinvolgimento di tale gene nella risposta a questo tipo di stress; infatti la sua espressione veniva indotta dopo 2 ore di esposizione al Cd. Per verificare se la trascrizione di questo fattore MYB subisce una modulazione anche in seguito ad altri stress di tipo abiotico, piante di *A. thaliana WT* sono state sottoposte a differenti trattamenti: alte e basse temperature (42 °C e 4 °C), stress idrico, stress salino (NaCl) e trattamenti ormonali (IAA, ABA, kin e GA_3). L'analisi *Real-Time* PCR ha mostrato che la variante *MYB59.2* rispondeva allo stress da ABA, freddo e disidratazione, mentre l'espressione della variante *MYB59.3* veniva indotta dalla disidratazione sia nelle foglie sia nelle radici. L'espressione della forma *MYB59.1*, invece, non presentava modificazioni significative in seguito a tali trattamenti. In seguito, è stata condotta un'analisi dell'espressione genica, tramite *Real-Time* PCR, in piante di *Arabidopsis WT* in diversi organi: foglie della rosetta, foglie caulinari, stelo, fiori chiusi, fiori aperti ,e radici. L'analisi ha previsto l'utilizzo di tre coppie di *primers*, ognuna specifica per una delle tre varianti di *splicing* del gene stesso. L'analisi ha mostrato che le tre varianti di *splicing* avevano pattern di espressione organo-specifica. La caratterizzazione del gene è proseguita con l'isolamento del cDNA *full length* di *MYB59.1*, *MYB59.2* e *MYB59.3* ed il clonaggio nel vettore pMD1, sotto il controllo del promotore CaMV35S, per l'espressione in pianta. Inoltre, è stato analizzato un mutante inserzionale *knock-out* per il gene *MYB59*. Sono state messe a confronto piante WT, piante sovraesprimenti le tre forme di *splicing* e piante mutanti *knock-out*. Si è notato che le piante sovraesprimenti *MYB59.1* presentavano un'area fogliare maggiore, mentre le piante mutanti avevano foglie con dimensioni molto minori rispetto a piante controllo. Anche l'espressione ectopica di *MYB59.2* e *MYB59.3* portava ad una diminuzione dell'area fogliare. Successivamente, è stato confermato il coinvolgimento di *MYB59* nella risposta al Cd. È stata eseguita la quantificazione del contenuto di tale metallo in foglie e radici di piante WT, sovraesprimenti le tre forme di *splicing* e piante che presentavano una riduzione del 98% dell'espressione genica (dovuta all'induzione del silenziamento genico). I dati indicavano che il gene potrebbe essere coinvolto nel trasporto del Cd dalle radici alle foglie, in quanto, nelle radici, le piante con espressione ectopica del gene presentavano un contenuto del metallo maggiore rispetto a quella riscontrata nelle piante controllo. Nelle foglie invece la quantità di Cd era maggiore nelle linee silenziate. Successivamente è stato eseguito lo studio delle sequenze promotrici. Sono state quindi amplificate tre regioni di circa 2.0 Kbp a monte degli ATG delle tre forme e sono state clonate in un vettore a monte del gene reporter GUS. Il clonaggio del promotore della variante *MYB59.1* ha previsto l'inserimento di due mutazioni

puntiformi all'interno della sequenza amplificata, in quanto l'ATG di tale forma si trova a valle rispetto agli ATG delle altre due. Attraverso un set di tre reazioni di PCR effettuate con *primers* contenenti le mutazioni volute, gli ATG delle varianti *MYB59.2* e *MYB59.3* sono stati modificati in TAG. Il saggio istochimico dell'attività GUS ha permesso di analizzare l'espressione delle tre varianti geniche nei diversi organi e tessuti vegetali. L'espressione di *MYB59.1* è stata localizzata soprattutto a livello delle nervature fogliari, mentre l'espressione di *MYB59.2* principalmente nelle antere immature, e per *MYB59.3* si è osservata l'espressione in tutti gli stadi vegetativi e nei sepali, ma non nelle antere. Si può quindi ipotizzare che le tre varianti di *splicing*, avendo tre diverse localizzazioni, possano avere anche distinte funzioni all'interno della pianta. È stato inoltre analizzato il pattern di metilazione di una regione ripetuta (*direct repeat*) che si trova nell'intorno del TATA box e della regione che contiene l'intera sequenza del primo introne. L'analisi è stata condotta in foglie e antere di piante di *Arabidopsis WT* usando il metodo del bisolfito, che converte le citosine non metilate in uracile. Le foglie presentavano una metilazione maggiore rispetto a quella delle antere. Per verificare se esiste una corrispondenza tra mRNA e proteine, e quindi per capire quale variante di *splicing* viene in effetti espressa in pianta, sia in condizioni di crescita standard sia in seguito a stress, sono state adottate due diverse strategie, che hanno previsto la preparazione di proteine di fusione con il FLAGtag e con l'HaloTag®. Sfortunatamente, in entrambi i casi l'analisi Western ha evidenziato la presenza di segnali aspecifici. Per ovviare questo problema, è in corso la preparazione di un anticorpo specifico per MYB59. È stato riportato che il gene *MYB59* presenta il 74.2% di identità di sequenza nella regione codificante con *MYB48*, un altro fattore R2R3-MYB. I due geni mostrano il meccanismo di *splicing* alternativo conservato e probabilmente risultano da un evento di duplicazione genica (Li *et al*., 2006). Visto che le piante mutanti *myb48* non presentano particolari differenze fenotipiche rispetto al *WT*, i singoli mutanti *knock-out* dei due geni sono stati incrociati per ottenere il doppio mutante *myb59myb48*, per capire meglio il ruolo che i due fattori di trascrizione svolgono all'interno della pianta.

Capire la funzione di tale fattore di trascrizione potrebbe essere molto importante per comprendere i meccanismi vegetali che stanno alla base di alcuni processi di sviluppo e di risposta agli stress.

SUMMARY

Plants are frequently subjected to stress, that is external conditions that negatively influence plant growth and development. The perception of stress and signal transduction, carrying to the activation of adaptive responses are critical steps in determining plant survival. Plants can respond to an external stimulus activating defence mechanisms specific for a particular stress, or mechanisms able to respond to different stresses. In fact, different signalling transduction pathways are often interconnected to various levels, can share one or more components or intermediates or have common outputs (Chinnusamy *et al*., 2003). However, the information about the cross-talk is limited, since little is known about signalling pathways receptors and intermediates. Downstream of the reaction cascade there are transcription factors, that are able to regulate gene expression binding DNA in a sequence-specific manner. These proteins are normally expressed in a tissue, development or stimulus-specific manner, such as an environmental stress, and are responsible for the selectivity in gene regulation (Zhang, 2003). So, understanding TFs function is an important step towards studying plant development and stress responses. MYB transcription factors family is the larger TFs family in plants (Stracke *et al*., 2001). Its members are characterized by a structurally conserved domain of about 52 aminoacid (MYB domain), containing three regularly spaced tryptophan residues. They are able to bind DNA in a sequence-specific manner adopting an helix-turn-helix conformation. MYB proteins can be classified into three groups depending on the number of adjacent repeats in the binding domain. Different functions was proposed for different MYB proteins: some are involved in the control of secondary metabolism, cell proliferation and differentiation, whereas others are needed in signal transduction pathways responding to different stimuli (Martin e Paz-Ares, 1997), such as salicylic acid, and abscisic and gibberellic acid.

This PhD work has been focused on the characterization of the *AtMYB59* gene, that belongs to the R2R3-MYB family and presents three splicing variants, called for simplicity *MYB59.1*, *MYB59.2* e *MYB59.3*. It is known that this gene is involved in the Cd stress response (Li *et al*., 2006) and, in particular, it has been demonstrated that the expression of its gene homolog in *B. juncea* was modulated after Cd exposure (Fusco *et al*., 2005). Through Real Time PCR, carried out on *Arabidopsis* WT treated with Cd, an involvement of the gene in the response to this type of stress was confirmed; in fact, the expression was induced after a 2 h Cd treatment. To verify if the gene transcription was modulated also after different abiotic stress, *Arabidopsis* WT plants were subjected to different treatments: high and low temperatures (42 °C and 4 °C), drought,

salinity (NaCl) and hormonal treatments (IAA, ABA, kin e GA_3). Real Time PCR analysis was shown that *MYB59.2* responded to ABA, cold and drought stress, whereas *MYB59.3* was induced by drought in leaves as well as in roots. *MYB59.1* expression, instead, did not show important modulations after these treatments. Furthermore, through Real Time PCR, gene expression analysis on different *Arabidopsis* organs was conducted: rosetta leaves, cauline leaves, stem, closed flowers, open flowers and roots. For this analysis primers designed on a unique region of each splicing variant were used. The results indicated that the three splicing variants had different organ specific expression. Gene characterization was carried out by isolating full-length cDNA of *MYB59.1*, *MYB59.2* and *MYB59.3* and cloning in pMD1 vector for plant expression, under the control of CaMV35S promoter. Moreover, an insertional knock-out mutant was analyzed. Comparison between WT, overexpressing and mutant plants showed that plants overexpressing *MYB59.1* had a leaf area higher and mutant plants lower than control plants. The overexpression of *MYB59.2* and *MYB59.3* also induced a decrease of leaf area in respect to WT plants. Afterwards, an involvement of *MYB59* in Cd response was confirmed. The quantification of Cd content in leaves and roots of WT, overexpressing and plant showing a gene expression reduction of 98% (due to gene silencing induction) was carried out. Data indicated that the gene may be involved in root-to-shoot Cd transport, since, in roots, overexpressing plants showed a metal content higher than that in control plants. In leaves, instead, Cd content was higher in silenced lines. Subsequently, a study of promoter sequences was carried out. Three region of about 2.0 Kbp upstream ATGs of the three forms were amplified and cloned in a vector upstream a GUS gene reporter. For the analysis of the promoter region regulating *MYB59.1*, mutagenesis of the ATG of *MYB59.2* and *MYB59.3*, that are upstream the starting codon of this variant, was performed. So, these two ATGs were converted in the TAG stop codons, through a set of three PCR reactions. GUS assay allowed to localized the expression of the three variants in different plant organs and tissues. The expression of *MYB59.1* was found mainly in leaf veins; the expression of *MYB59.2* was mainly in the immature anthers, whereas the expression of *MYB59.3* was localized in most vegetative tissues, sepals, but not in anthers. It can be hypothesized that the three splicing variants, having different localizations, may also play different roles in plant. Moreover, methylation pattern of a direct repeat around the TATA box and the region containing the whole first intron sequence was analyzed. The analysis was carried out in leaves and anthers of *Arabidopsis* WT using bisulfite method, that converts unmethylated C in U. In leaves, this repeat region is highly methylated, whereas not in anthers.

To verify if a correspondence between mRNA and proteins exists, and so to understand which splicing variants is actually translated in plant, in standard condition as well as after stress exposure, two protein fusion strategies were carried out using the FLAGtag and the HaloTag®.

Unfortunately, in both cases Western blot analysis showed aspecific signals. To avoid this problem, the preparation of an antibody anti-MYB59 is under preparation. It has been reported that *MYB59* share the 74.2% nucleotide sequence identity in their coding region with *MYB48*, another R2R3-MYB. The two genes show a conserved alternative splicing mechanism and probably are the result of a relatively recent duplication event (Li *et al*., 2006). Since *myb48* mutant plant did not show phenotypic differences respect to WT, single mutants knock out were crossed to obtain double mutant *myb59myb48*, to better understand the role of these two TFs in plants.

Understand *MYB59* gene function may be an important step to gain insight into plant development and stress response mechanisms.

Chapter 1

INTRODUCTION

1.1 Abiotic stresses

Plants have a sessile nature and so they must constantly adapt their growth and metabolism to changes in the environmental conditions. For this reason, plants have evolved the ability to rapidly regulate gene expression in order to survive environmental changes in light, temperature, salinity and availability of water and essential minerals. Stress conditions that plants encounter are not always as rare as we might think (Ferguson, 2004). The most common environmental variables, that are necessary for regulating normal plant processes of germination, growth and flowering, can impose significant stress on the plant. One example is light. Plants need light for photosynthesis and development, but an excess of light might impose serious damages that could ultimately bring to death, so mechanisms to dissipate unnecessary light energy have evolved, for example metabolites such as flavonoids, which avoid the occurrence of free radicals (Ferguson, 2004).

Stress resistance mechanisms can be grouped in two main classes: the ones that prevent stress exposure (avoidance), and those that allow plant to resist adverse situation (tolerance). Phreatophyte, for example, improve water access deepening roots and so they are able to survive during no rain periods. Xerophytes, instead, tolerate drought through morphological traits that facilitate survival in arid climate (Buchanan *et al*., 2003). Adaptation mechanisms are constitutive traits and genetically determined for stress resistance; their expression is independent of the stress appearance and contribute to increase the fitness of a population. An example is represented by ephemeral desert plants, that germinate and complete their life cycle when sufficient amount of water is available. Generally, plants adapted to arid climate show sunken stomata and deepened root apparatus and their trichomes are able to reflect the light. Other resistance mechanism are achieved through acclimation, that is obtained through an adaptive response to environmental changes. During this period, the organism can modified its own homeostasis in order to survive. An acclimation period before a stress can confer resistance to a plant otherwise vulnerable (Ferguson, 2004; Bartels and Sunkar, 2005).

Plant response to abiotic stress depend on species, genotype, developmental stage and organ or tissue affected. However, stress duration, intensity or rapidity also greatly influence plant

response. Organism responses to various stresses may be common or specific; moreover, a combination of different negative conditions may cause a different response from that obtained during a single type of stress (Buchanan *et al*., 2003; Chinnusamy *et al*., 2003). Therefore, plants respond to environmental stresses at cellular, molecular but also physiological level, in order to gain tolerance to the stress and survive. Cellular responses to stress can cause changes in cell cycle and cell division, alterations in the endomembrane system, and changes in cell wall architecture and composition (Moore *et al*., 2008). At the biochemical level, plants can alter their metabolism producing osmoregulatory compounds such as proline and glycine betaine (Taiz and Zeiger, 2006). However, changes in metabolism and development arise from modifications in gene expression. During stress response, proteins and molecules including transcription factors (TFs), hormones and second messengers act in concert in the stress-signal transduction pathways (**Fig. 1.1**) (par. 1.5 in this section).

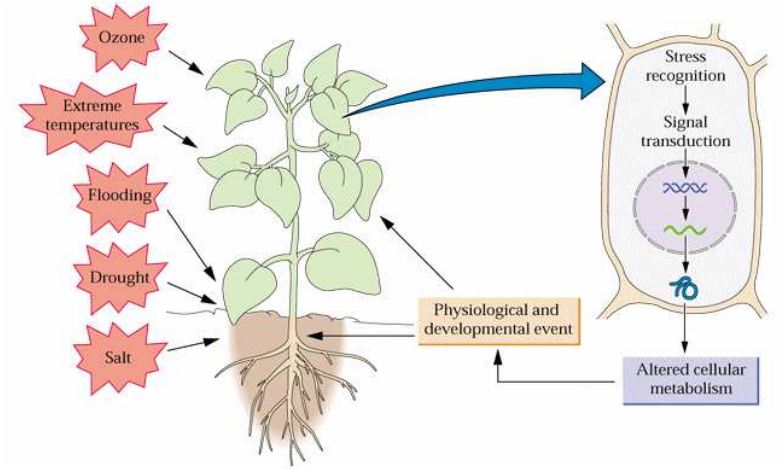


Fig. 1.1: Plants respond to stress both at cellular and whole organism level. From *Biochemistry & Molecular Biology of Plants*. Buchanan *et al*., 2003.

It is possible that one way to tolerate continuous environment changes is to combine and/or co-ordinate stress response mechanisms. Often one stress is closely associated with another, such as high temperature and drought. The activation of similar mechanisms to respond to different environmental stresses, is extremely important to ensure the maximum control maintaining plant homeostasis (Ferguson, 2004).

So, for sustainable agriculture development, future crops should have abiotic stress resistant traits and the mechanism for stress tolerance. The tolerance mechanisms can also be improved by the development of new techniques employing plant physiology and plant molecular biology tools.

1.2 Transcriptional regulation of gene expression in plants during abiotic stress

Microarray studies have been used to examine the plant gene expression in response to stress, demonstrating that changes in genes expression occur in plants after stress exposure. For example, in rice, the stress-controlled genes reflect up to 10% of the total genome (Kawasaki *et al*., 2001). In response to stress, some genes are overexpressed, whereas others are repressed.

Water stress is one of the most studied stress since its relevance on the agronomical productivity. Water deficit can be caused by prolonged no rain periods, high salt concentration and freezing temperature. During water stress, late embryogenesis abundant (LEA) proteins accumulate in vegetative tissues (Ramanjulu and Bartels, 2002; Goyal *et al*., 2005). These proteins have been known for many years to be present in maturing seeds (Delseny *et al.,* 2001). LEA proteins are associated with water stress tolerance resulting from desiccation and cold shock in plants and animals. However, although various functions of LEA proteins have been proposed, their precise role is still not completely understood. *In vitro* data suggest that these proteins prevent proteins aggregation due to water stress (Chakrabortee *et al*., 2007). Osmotin is another example of protein induced by drought. Nevertheless, the transcription of a tobacco osmotin gene is induced by numerous signals: ABA, ethylene, auxin, tobacco mosaic virus (TMV) infection, drought, cold and high salinity (Zhu *et al*., 1995). Furthermore, some stress-induced genes are regulated by ABA, a plant hormone that increase under water stress and cold stress (Rock, 2000). In the promoter regions of ABA-regulated genes, there are conserved *cis*-element, called ABA-responsive elements (ABREs), that control gene expression through bZIP-type AREB/ABF TFs. Three AREB/ABF TFs (AREB1, AREB2 and ABF3) cooperatively regulate ABRE-dependent ABA signalling involved in drought stress tolerance (Ramanjulu and Bartels, 2002; Yoshida *et al*., 2010). Another *cis*-element, dehydration responsive element (DRE), stimulates gene transcription in response to water stress and cold, but in a ABA-independent manner. DRE elements are recognized by CBF/DREB1 TFs, that are key regulators of plant freezing tolerance and members of the AP2/ERF (APETALA2/ ETHYLENE RESPONSE FACTOR) multi-gene family (Yamaguchi-Shinozaki and Shinozaki, 1994). Analyses in transgenic plants have shown that ectopic expression of CBFs is sufficient to activate the expression of *COLD RESPONSIVE* (COR) genes and induce cold acclimation (Miura *et al*., 2007). CBFs regulate the expression of genes involved in phosphoinositide metabolism, and osmolyte biosynthesis (Chinnusamy *et al*., 2007). Dehydrins are a family of plant proteins produced in response to low temperatures and water stress (Puhakainen *et al*., 2004). Their production is induced by ABA and salt stress. Dehydrins in barley and maize are extremely hydrophilic and glycine-rich. It seems that they may protect

membranes from damage (Puhakainen *et al*., 2004).

Responses to high-temperature stress cause a decrease in normal protein production and an increase in heat shock proteins (HSPs) synthesis. HSPs are molecular chaperones that regulate folding, localization, accumulation, and degradation of protein in both plant and animal (Feder and Hofmann, 1999). They are believed to play a key role in many cellular processes and in tolerance to various environmental stresses (Swindell *et al*., 2007).

Soil contamination by heavy metals such as cadmium (Cd) represents an emerging problem for plants, animal and human health. Plant responses to heavy metals depend on an intricate signal transduction pathway that converges in transcription regulation of metal-responsive genes (DalCorso *et al*., 2008). Cadmium can induce the synthesis of phytochelatins (PCs), small metal-binding peptides derived from glutathione, via a reaction catalyzed by the cytosolic PCs synthetase (PCS). Metallothioneins (MTs) are also induced in response to heavy metal stress (Cobbett and Goldsbrough, 2002). It has been observed that, during heavy metal exposure, activity and accumulation of different enzymes were altered (Prasad, 1995). Cd, for example, inhibits the activity of enzymes involved in carbon and nitrogen assimilation and nitrogen mobilization (DalCorso *et al*., 2008).

Understanding the role of proteins induced under different stress conditions may provide insight into multiple stress tolerance mechanisms.

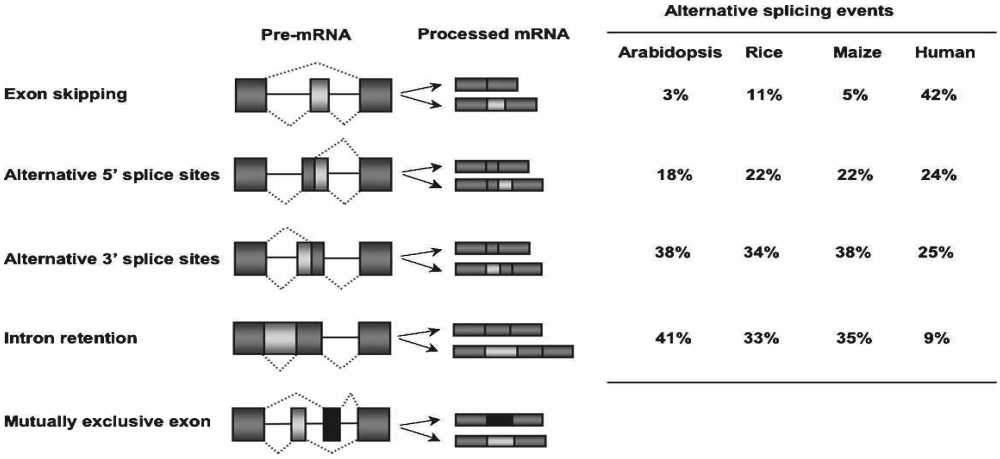
1.3 Post-transcriptional regulation of gene expression in plants during abiotic stress

It is becoming evident that the accumulation of proteins during stress exposure is also influenced by post-transcriptional regulatory mechanisms that can influence mRNA processing, stability and translation (Floris *et al*., 2009).

1.3.1 Splicing and alternative splicing

The splicing reaction processes eukaryotic mRNAs from their longer precursors and represents a unique means of gene regulation (Black, 2003). Alterations in splice site selection can affect the mRNA and protein products of a gene in different ways. Alternative splicing (AS) is an important mechanism that creates multiple mRNA transcripts (isoforms) from a single gene, in different

tissues or cells under certain circumstances. AS can affect mRNA stability and efficiency of translation as well as increase protein diversity (Stamm *et al*. 2005). AS has been shown to exist in nearly all metazoan organisms and was estimated to involve 30–70% of human genes (Xu *et al*., 2002). Indeed, the human genome is predicted to contain 32,000 genes (Lander *et al*. 2001), while the proteome have 90,000 proteins. This discrepancy could be explained by the fact that many human genes have been demonstrated to undergo AS (Barbazuk *et al*., 2008). In humans, altered expression of splicing variants has been correlated with numerous disease (Zhong *et al*., 2006). In plant, this phenomenon was not extensively studied because it was considered rare (Reddy, 2007). The first demonstration of AS in plants result from the characterization of the spinach and *Arabidopsis* rubisco activase (Werneke *et al*., 1989). Recent studies based on whole-genome data suggest that AS in plants occurs more frequently than originally expected (Barbazuk *et al*., 2008). Different types of AS have been observed (**Fig. 1.2**), including exon skipping, alternative 5' or 3' splice site, intron retention and mutually exclusive exon (Stamm *et al*., 2005; Reddy, 2007; Barbazuk *et al*., 2008).



Alternative splicing events			
Arabidopsis	Rice	Maize	Human
3%	11%	5%	42%
18%	22%	22%	24%
38%	34%	38%	25%
41%	33%	35%	9%

Fig. 1.2: Different types of alternative splicing in plants on the left, and the percent values of these events in four different organisms on the right. From Barbazuk *et al*., 2008.

As shown in **Fig. 1.2**, plants and animals use preferentially different AS mechanisms: in human the most abundant AS event is exon skipping (42%), while in plant is intron retention (41%). In fact, it has been demonstrated that intron retention is the common AS in *Arabidopsis* and rice (Kim *et al*., 2007; Ner-Gaon *et al*., 2004). These differences suggest that the mechanism of splice site selection may differ between plants and animals. AS events occur preferentially to mRNAs of

certain classes of genes commonly involved in signal transduction cascade and post-translational modifications (Ner-Gaon and Fluhr, 2006) and also in response to abiotic stress (Mazzucotelli *et al*., 2008). In *Arabidopsis*, cold and heat stresses regulate the alternative splicing of the mRNAs for many genes encoding serine/arginine-rich (SR) proteins, that have different functions under stress conditions (Floris *et al*., 2009).

1.3.2 RNA silencing

RNA silencing is a mechanism involved in gene expression regulation during plant development and in response to stress. This phenomenon is characterized by the presence of endogenous small RNA molecules (sRNAs), ranging from 20 to 25 nt in size and that inhibit gene expression in a sequence-specific manner, binding to mRNAs partially or fully complementary (Brown, 2002). sRNAs have been well characterized for their involvement in abiotic stress such as oxidative and salt stress. Upon induction, sRNAs can repress negative regulators of stress tolerance. Alternatively, they can be downregulated by the stress and thus allow the accumulation of positive regulators of stress tolerance (Floris *et al*., 2009). These two action mechanisms can be exemplified by miR399 and miR398. miR399 accumulates in response to inorganic phosphate (Pi) deficiency and negatively regulates *UBIQUITIN CONJUGATING ENZYME 24* (*UBC24*) expression. UBC24 is involved in the protein degradation pathway and regulates Pi transporters availability (Kraft *et al*., 2005). On the other hand, miR398 is downregulated in response to oxidative stress and this repression allows the accumulation of *CSD1* and *CSD2* (Cu/Zn-Superoxide dismutase) mRNAs, needed to reduce ROS production (Floris *et al*., 2009).

1.3.3 Translational regulation in plants

The regulation of mRNA translation allows the modulation of the level of protein synthesis. It has been demonstrated that many mRNAs undergo translational changes in response to environmental stimuli such as salt stress (Hua *et al*., 2001), heat stress (Horiguchi *et al*., 2000), drought (Wood *et al*., 2000) and pathogen infection (Bailey-Serres *et al*., 1999). Interestingly, the 5'UTR and 3'UTR of mRNAs seem to be involved in the translation control in response to environmental stresses (Floris *et al*., 2009).

1.4 Epigenetic control of development and stress response

Developmental and environmental stimuli can modulate gene expression through nucleosome histone modifications and DNA methylation. Most of these stress-induced epigenetic modifications are reset to the basal level once the stress is relieved, while some of them may be stable, that is, may be carried forward as "stress memory". Retention of stress memory for short periods is well known in plants, as evident from acclimation mechanisms, and it depend on the half-life of stress-induced proteins, RNAs and metabolites. Retention of stress memory for longer periods, instead, involves reprogramming in the plant phenology and morphology, but also epigenetic processes, such as stable histone modifications and DNA methylation (Boyko and Kovalchuk, 2008; Chinnusamy and Zhu, 2009).

1.4.1 Histone modification

Histone deacetylation is a process by which the acetyl groups are removed from histone tails and is catalyzed by the enzymes histone deacetylases (HDACs). The histones in heterochromatin are generally unacetylated, while those in functional domains are acetylated, an indication that this kind of modification is linked to the DNA packaging (Brown, 2002). In the nucleus, DNA is attached to a variety of protein that are not directly involved in genome expression and must be removed in order to allow the RNA polymerase and other expression proteins to gain access to the genes (Luger *et al*., 1997). Histone acetylation, catalyzed by the enzyme histone acetyltransferase (HATs), refers to the attachment of acetyl groups to lysine aminoacids in the N-terminal regions of each of the core molecules. This process reduces the affinity of the histones for DNA allowing the docking of enzymatic transcription complex (Brown, 2002). HATs are involved in the activation of stress responsive genes. On the other hand, environmental stimuli can repress the target genes through reduction in histone acetylation levels. The histone deacetylases (HDACs), namely HDA6 and HDA19, mediate histone deacetylation in response to biotic and abiotic stresses in *Arabidopsis* (Chinnusamy and Zhu, 2009).

1.4.2 DNA methylation

DNA methylation has evolved from an immune function in bacteria to a regulator of gene expression in higher eukaryotes (Singh *et al*., 2008). In eukaryotes, cytosine in DNA molecules

can be modified to 5-methylcytosine by the addition of a methyl group by enzymes called DNA methyltransferases. Cytosine undergo methylation processes for the silencing of gene expression in a wide number of eukaryotic organisms (Goll and Bestor, 2005). Both plants and animals use DNA methylation to protect their genome against transposons and to regulate specific endogenous genes (Goll and Bestor, 2005). Plants use DNA methylation for genomic imprinting and to modulate the expression of repeated gene families. In mammals, DNA methylation controls genomic imprinting, X-chromosome inactivation and the silencing of tumour-suppressor genes (Chan *et al*., 2005). This process has been observed mainly at tandem or inverted repeats of genomes (Law and Jacobsen, 2009). It has been proposed that tandem repeats attract the small interfering RNAs (siRNA)-making complex. siRNAs that are produced from these sequences recruit the methylation machinery, leading to the silencing of the gene containing the repeat (Robinson, 2006). In general, there are three different methylation systems that maintain cytosine methylation in three different sequence contexts: CG, CHG [where H may be A, T or C] and CHH. CG methylation is maintained by MET1 (DNA METHYLTRANSERASE 1) and VIM1 (VARIANT IN METHYLATION 1). CHG methylation is controlled by the plant-specific CMT3 (CHROMOMETHYLASE) and KRYPTONITE (SUVH4), while CHH is controlled by DRM2 (DOMAINS REARRANGED METHYLTRANSFERASE 2) that acts in a pathway called RNA-directed DNA methylation (RdDM) (Law and Jacobsen, 2009). While in mammals methylation is mainly restricted to CG dinucleotides, in plants there are high levels of asymmetric methylation (CHH) and also abundant CHG methylation (Cao and Jacobsen, 2002) (**Fig. 1.3**).

The *MET1* class of genes is most similar to mammalian Dnmt1 in both function and sequence. The *met1* mutants lack the majority of CG methylation (Cao and Jacobsen, 2002). These mutant are viable, but present several developmental abnormalities that increase progressively when the mutants are inbred (Cao and Jacobsen, 2002).

The *CMT3* gene family is plant-specific and encodes methyltransferase proteins containing a chromo domain. The *cmt3* mutants show wild-type morphology but a total loss of CHG methylation and a reduction of asymmetric methylation at some *loci* (Goll and Bestor, 2005). They also show the expression reactivation of endogenous retrotrasposon elements (Lindroth *et al*., 2001).

The third class of genes, composed of *DRM1* and *DRM2*, encode proteins containing catalytic domains similar to those of the mammalian Dnmt3 methyltransferase. It has been reported that DRM genes are important for the initial establishment of methylation in all contexts and for *de novo* methylation (Cao and Jacobsen, 2002).

The *FWA* transcription factor is an example of a gene regulated by DNA methylation. *FWA* is

expressed only in the endosperm and is silenced in all other tissues by the methylation of tandem repeats in its promoter region (Chan *et al.*, 2005). In plants, the DNA methylation of promoter regions usually causes transcription repression, but methylation in coding regions does not generally influence gene expression. *SUPERMAN* (*SUP*) and *AGAMOUS* (*AG*) genes are exceptions to this rule; in fact, DNA methylation in their transcribed regions seems to cause transcriptional shut-down (Chan *et al.*, 2005). It has been reported that *drm1drm2* double mutants lack *de novo* methylation of the *FWA* and the *SUP loci.* It has also been showed that DRMs act redundantly with CMT3, so that only in *cmt3drm1drm2* triple mutants all asymmetric methylation is lost (Cao and Jacobsen, 2002).

Therefore, although DRM2 and CMT3 act redundantly to maintain non-CG methylation, DRM2 functions alone during *de novo* methylation (Cao and Jacobsen, 2002).

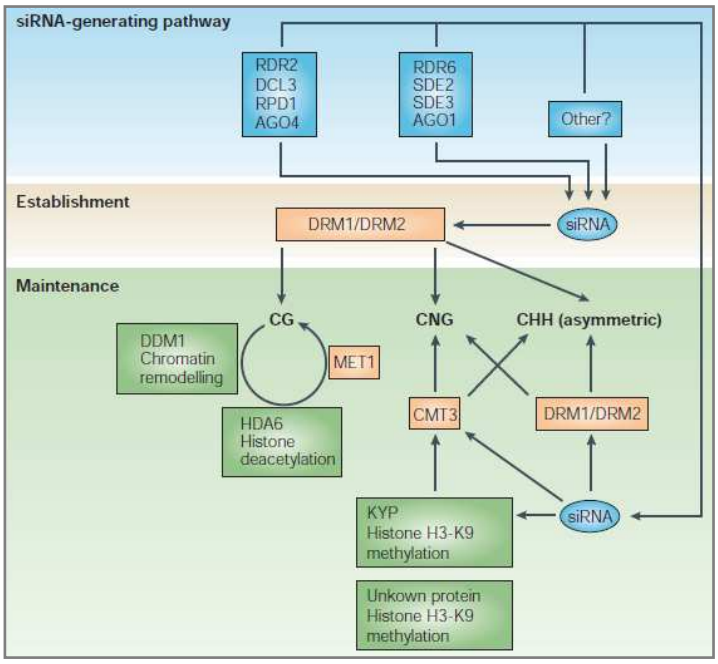


Fig. 1.3: A scheme showing the establishment and maintenance of DNA methylation in *A. thaliana* in different sequence contexts. DDM1, DECREASE IN DNA METHYLATION 1; AGO1, ARGONAUTE 1; AGO4, ARGONAUTE 4; DCL3, DICER-LIKE 3; RDR2, RNA-DEPENDENT RNA POLYMERASE 2; RDR6, RNA DEPENDENT RNA POLYMERASE 6; RPD1, RNA POLYMERASE D1; SDE3, SILENCING DEFECTIVE 3; SGS3, SUPPRESSOR OF GENE SILENCING 3. From Chan *et al.*, 2005.

Stresses can induce changes in gene expression through DNA hypomethylation or hypermethylation. It has been demonstrated that cold stress induced the expression of *ZmMI1* in maize roots, and this was correlated with a reduction in methylation in the DNA of the

nucleosome core (Choi and Sano, 2007). Microarray data showed that expression of many genes involved in RdDM pathways in *Arabidopsis* are influenced by abiotic stresses and ABA (Chinnusamy and Zhu, 2009).

1.5 Signal transduction

Cell signalling is one aspect of the complex system of intracellular communication that control basic cellular activities and cell-environment interactions. Signal transduction refers to any process by which a cell responds to one kind of signal, which is used to modify physiology, morphology and development of the entire plant. Signals can vary over the time in quantity and quality. Moreover, they can interact with each other in cooperative and synergic way to cause the final response, a phenomenon called "cross-talk" (Priest *et al*., 2009). Plant response depends on stage of development, environmental conditions and internal clock. The integration of different kind of information over the signal transduction will cause the plant final response (Chinnusamy *et al*., 2003).

Generally, a signal transduction pathway starts with receptors activation, which undergo a conformational modification after the binding with a ligand. Most of the receptors are transmembrane proteins, while some of them act in the citosol. After ligand binding, some receptors are able to directly regulate gene expression, while most of them need a wide range of intracellular mediators, which activate each other in a chain reaction that amplify the starting signal (Alberts *et al*., 2002). Signal transduction pathways, involved in modifications of gene expression after stress exposure, are not yet well understood. However, it is likely that hormones including ABA, ethylene and jasmonic acid (JA), and second messengers including calcium (Ca^{2+}) are involved (Alberts *et al*., 2002). The study of genome sequences of the two model plants, *Arabidopsis* and rice, has allowed to emphasize the limiting aspects of the linear pathways for environmental signalling cascades. In fact, signalling transduction pathways have often been described using a linear model, that begins with the signal perception and ends with the alteration of gene expression and, consequently, with the physiological responses (Shinozaki and Dennis, 2003). The genomics approach has shown the complex nature of gene families that encode signalling molecules and TFs (**Fig. 1.4**). In the *Arabidopsis* genome there are more than 1800 genes encoding TFs, more than 600 genes encoding protein kinases and more than 600 genes that encode F-box proteins. These gene families allow the control of signal cascades by redundant factors and supply complexity and flexibility in plant responses to external stimuli

(Shinozaki and Dennis, 2003). Although most studies focused on the transcriptional changes of gene expression, it is likely the involvement of post-transcriptional regulation mechanisms, that increase specific mRNA levels and transduction, modify protein activity or all of them (Floris *et al*., 2009). In fact, modification of proteins by phosphorylation, dephosphorylation, farnesylation and ubiquitination are important events in signalling pathways.

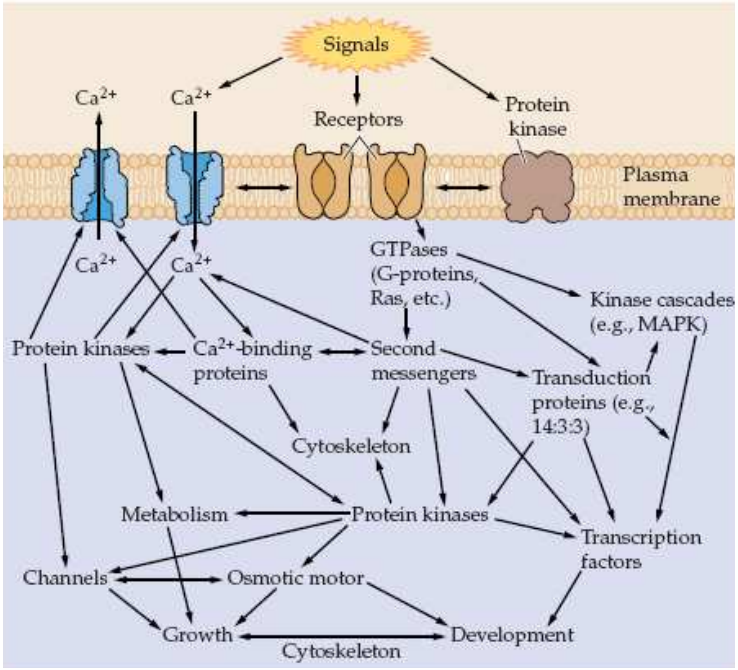


Fig. 1.4: Schematic representation of an intracellular signalling pathway. An extracellular signal molecule binds to a receptor protein thereby activating an intracellular signalling pathway that is mediated by a series of signalling proteins. Finally, one or more of these signalling proteins interacts with a target protein to cause changes in cell behaviour. MAPK: mithogen-activetd protein kinase. From *Biochemistry & Molecular Biology of Plants*. Buchanan *et al*., 2003.

1.6 Transcription factors

TFs regulate gene expression downstream the signal transduction pathways. TFs are sequence-specific DNA-binding proteins that can either activate or repress gene expression, binding to short *cis*-regulatory elements called TF-binding sites, that lie on the promoter region of the target gene. They are responsible for the selectivity in gene regulation and are often express in a tissue, development or stimulus-specific manner (Zhang, 2003).

TFs are composed of at least four domains: the DNA binding domain, the nuclear localization signal (NLS), the transcription activation domain and the oligomerization site (Liu *et al*., 1999). These domains cooperate to regulate many physiological and biochemical processes, modulating the rate of transcription initiation of target genes (Du *et al*., 2009).

With the completion of the *Arabidopsis* genome sequence, the whole number of genes coding for TFs can be described (Riechmann and Ratcliffe, 2000). In addition, it is also possible to investigate the similarities and the differences in transcriptional regulators among the three eukaryotic kingdoms: plants, animals (*Caenorhabditis elegans* and *Drosophila melanogaster*) and fungi (*Saccharomyces cerevisiae*) (Riechmann *et al*., 2000). As shown in **Table 1.1**, in *D. melanogaster* genes coding for transcription regulators are approximately 4,5 % of the entire genome, and this has been proposed to be related to the regulatory complexity. Likewise, considering that in *Arabidopsis* approximately the 6% of the genome is represented by TFs genes, it can be assumed that the regulation of transcription in plants is as complex as that in *Drosophila* (Riechmann *et al*., 2000).

Organism	Total number of genes	Genes coding for transcriptional regulators	
		Total number	Percentage of total number of genes
A. thaliana	~26,000	1533	5.9
S. cerevisiae	~6,000	209	3.5
C. elegans	~19,000	669	3.5
D. melanogaster	~14,000	635	4.5

Table 1.1. Content of TFs genes in eukaryotic genomes. The number of genes is given as an approximation, because the number predicted at the time that a genome is sequenced is always an estimation that is refined over the time. From Riechmann *et al*., 2000.

Transcription factors are usually divided into families according to their DNA-binding domains, which are important determinants of TF binding specificity (**Table 1.2**).

Major families of *Arabidopsis* transcription factors.

Gene family	Estimated number of genes in the *Arabidopsis* genome*	Gene family functions†	Genetically characterized *Arabidopsis* factors	Predicted number of proteins‡		
				D. melanogaster	*C. elegans*	*S. cerevisiae*
MYB	180	Secondary metabolism, cellular morphogenesis, signal transduction in plant growth, abiotic and biotic stress responses, circadian rhythm, and dorsoventrality	*AtMYB2, ATR1, CCA1, CPC, GL1, LHY, WER*	35	16	19
AP2/EREBP	150	Flower development, cell proliferation, secondary metabolism, abiotic and biotic stress responses, ABA response, and ethylene response	*ABI4, ANT, AP2, CBF1-3/DREB1A-C, DREB2A, ERF1*	0	0	0
NAC	105	Development, pattern formation, and organ separation	*CUC2, NAP*	0	0	0
bHLH/MYC	100	Anthocyanin biosynthesis, light response, flower development and abiotic stress	*PIF3*	61	38	8
bZIP	100	Seed-storage gene expression, photomorphogenesis, leaf development, flower development defense response, ABA response, and gibberellin biosynthesis	*ABI5, HY5, PAN*	24	18	15
HB	90	Development (leaf, root, internode, and ovule), stem cell identity, cell differentiation, growth responses, anthocyanin accumulation, and cell death	*ANL2, ATHB-2, BEL1, GL2, KNAT1, REV, STM, WUS*	113	88	10
Z-C_2H_2	85	Flower development, flowering time, seed development, and root nodule development	*FIS2, SUP*	352	138	47
MADS	80	Flower development, fruit development, flowering time, and root development	*AG, AGL15, ANR1, AP1 AP3, CAL, FLC, FUL, PI, SEP1, SEP2, SEP3, SHP1, SHP2, SOC1, SVP*	2	2	4
WRKY	75	Defense response		0	0	0
ARF-Aux/IAA	42	Auxin responses, development, and floral meristem patterning	AXR2, *AXR3, ETT, MP, NPH4, SHY2*	0	0	0
Dof	41	Seed germination, endosperm-specific expression, and carbon metabolism	*DAG1*	0	0	0

Table 1.2: Major families of *Arabidopsis* transcription factors. From Riechmann and Ratcliffe, 2000.

In plants, many biological processes are regulated at the transcriptional level and probably, the evolution of typical morphological features during species domestication has been associated with changes in TFs or their regulation (Zhang, 2003). Therefore, understanding TFs function is an important step towards studying plant development and evolution. Through the *Arabidopsis* and rice genome sequencing, it became possible to study the function of TFs on a genome-wide scale and to compare similarities and differences between the TFs of monocots and dicots plants (Qu and Zhu, 2006).

1.6.1 MYB transcription factors superfamily

The "*MYB*" is an acronym derived from "myeloblastosis", an old-fashioned name for a type of leukemia (Lipsick, 1996). The first *MYB* gene identified was the *v-MYB* gene of avian myeloblastosis virus (AMV) (Klempnauer *et al*., 1982). Subsequently, three v-*MYB*-related genes (c-*MYB*, A-*MYB*, and B-*MYB*) were found in many vertebrates, and are thought to be involved in the regulation of cell proliferation, differentiation, and apoptosis (Weston, 1998). Homologous genes were also identified in insects and fungi (Lipsick, 1996).

MYB TFs superfamily is the largest TFs family in plant (Stracke *et al*., 2001). MYB proteins are characterized by a MYB domain that generally consists of up to three imperfect repeats, each forming a helix-helix-turn-helix structure of about 52 aminoacids (Jin and Martin, 1999; Stracke *et al*., 2001). Three regularly spaced tryptophan residues are characteristic of a MYB repeat and they form a tryptophan cluster in the three-dimensional structure (Jin and Martin, 1999). MYB proteins can be classified into three groups depending on the number of adjacent repeats in the binding domain (Stracke *et al*., 2001): R1R2R3-MYB, with three adjacent repeats; R2R3-MYB, with two adjacent repeats; and the MYB1-related proteins, which usually contain a single MYB repeat or an atypical repeat (**Fig. 1.5**). R2R3-MYB proteins represent the largest MYB gene family in plants (Stracke *et al*., 2001).

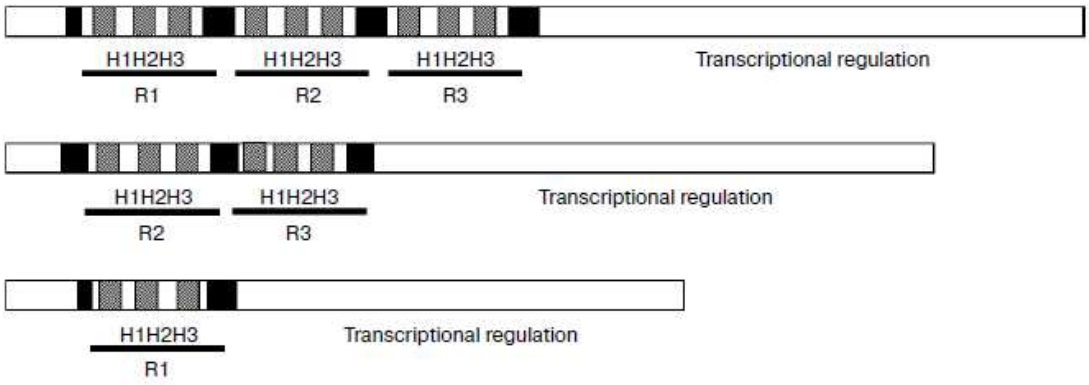


Fig. 1.5: Schematic representation of MYB protein domains. Shaded boxes represent the MYB domain; R1, R2 and R3 are repeats of the DNA binding domain; black bars indicate the three α-helices, with the second and the third ones that form a helix-turn-helix structure when bound to DNA sequence. From Du *et al*., 2009.

The fact that *MYB* genes exist widely in eukaryotes suggests that these genes may be very ancient during the evolutionary course (Yanhui *et al*., 2006). However, plant *MYB* genes are more complex compared with those of mammalians (Qu and Zhu, 2006).

In plant, the first MYB TF studied was C1 from maize, involved in antocyanin biosynthesis and

with significant structural homology to the vertebrate cellular proto-oncogene c-MYB (Martin and Paz-Ares, 1997). Since then, several *MYB* genes have been isolated. Some of the MYB proteins show a specific expression pattern. For example, AtMYB7 and AtMYB44 are present in several tissues, whereas the expression of AtMYB46 is only detected in siliques and AtMYB21 in flower buds (Shin *et al*., 2002; Du *et al*., 2009). AtMYB26 and AtMYB103, instead, are expressed in anthers (Du *et al*., 2009). In the context of the ovule, *MYB98* is expressed exclusively in the synergid cells (Kasahara *et al*., 2005) Moreover, it has been demonstrated that AtMYB17 is highly expressed only in flowers and siliques (Zhang *et al*., 2009).

1.6.2 Functions of MYB transcription factors

In higher plants, *MYB* genes play a wide number of functions, including the control of secondary metabolism, cell shape and organ development, and signal transduction pathways responding to different stimuli (Martin and Paz-Ares, 1997). Several MYB TFs are involved in the phenylpropanoid biosynthetic pathway. Phenylpropanoid metabolism is one of the three main types of secondary metabolism starting from phenylalanine. This pathway is divided into two parts: the biosynthesis of flavonoids and of lignin. The first is now completely elucidated. Many R2R3-MYB TFs from several plants, such as *Arabidopsis*, tobacco and maize, are involved in the regulation of different branches of flavonoid metabolism. For example, the seed-specific MYB factor TRASPARENT TESTA 2 (TT2) is involved in the accumulation of proanthocyanidin in developing seeds (Nesi *et al*., 2001). Moreover, AtMYB12 has been shown to be an activator of flavonoid biosynthesis (Du *et al*., 2009). Another example is represented by *PAP1* gene from *Arabidopsis* that encodes an R2R3-MYB which is able to alter lignin biosynthesis when over-expressed in *Arabidopsis* plants (Du *et al*., 2009). *AtMYB58* and *AtMYB63* genes are known to regulate secondary cell wall formation, activating lignin biosynthetic pathway (Zhou *et al*., 2009). It has also be suggested an involvement of MYB proteins in the regulation of secondary xylem formation in pine (Patzlaff *et al*., 2003). It has been also demonstrated that *MYB59* and *MYB48* genes were up-regulated in the *Arabidopsis* xylem (Oh *et al*., 2003).

Many MYB TFs play a role in the development of plant organs. For example, AtMYB33 and AtMYB65 facilitate anther development (Millar and Gubler, 2005). Noteworthy, neither of the two single mutants displayed a phenotype, whereas the *myb33myb65* double mutant was male sterile, implying that MYB33 and MYB65 are functionally redundant (Millar and Gubler, 2005). Another gene, *AtMYB103,* is involved in anther development, controlling tapetum development, callose dissolution and exine formation (Zhang *et al*., 2007). Moreover, in *Arabidopsis*, stamen

maturation requires the expression of *MYB108* gene that acts together with *MYB24* in a jasmonate-mediated pathway (Mandaokar and Browse, 2009). *MYB99*, *MYB101* and *MYB105* are expressed during pollen development (Alves-Ferreira *et al*., 2007). MYB TFs are also involved in determination of the trichomes fate. The *GLABRA1* (*GL1*) gene, for example, encodes a MYB protein essential for trichomes formation. The *WEREWOLF* (*WER*) gene, instead, encodes a MYB factor required for the specification of the non-hair cell type, and it is expressed in the non-hair cells of root (Du *et al*., 2009).

R2R3-MYB factors are also involved in the signal transduction pathways of ABA, salicylic acid (SA), giberellic acid (GA) and JA (Du *et al*., 2009). The phytohormone ABA plays a key role in plant adaptation to environmental stresses, such as drought and high salinity, and is produced under water stress conditions (Rock, 2000). The expression of *AtMYB2* has been shown to be induced by water deficit and ABA treatment (Abe *et al*., 2003). Further studies demonstrate that most of the drought-inducible genes are also induced by ABA. In fact, the expression of *AtMYB44* is induced by dehydration, salt treatment, and low temperatures and the transcript accumulation is also increased after ABA, methyl jasmonate, or ethylene treatment (Jung *et al*., 2008). MYB96 also mediates ABA signalling during drought stress response in *Arabidopsis* (Seo *et al*., 2009).

Recently, it has been demonstrated that GA promotes the production of jasmonate tha,t in turn, induces the expression of *MYB21*, *MYB24*, and *MYB57* to promote stamen filament development in *Arabidopsis* (Cheng *et al*., 2009).

A role for a MYB protein in cell cycle regulation has been recently proposed by Mu and colleagues. They demonstrated that *MYB59* affects cell growth by influencing DNA replication and cell division (Mu *et al*., 2009).

In eukaryotes, gene expression is often regulated by multi-protein complexes. Is has been shown that a functional relationship between MYB factors and helix-loop-helix (bHLH) proteins exists (Du *et al*., 2009). An example is the regulation of epidermal cell differentiation and cell patterning in root hair and trichomes development. GL1 MYB protein, for instance, interacts with GL3 and EGL3 bHLH factors controlling trichome formation in *Arabidopsis* (Zhang *et al*., 2003).

1.7 Aim of the work

Various abiotic stresses result in both general and specific effects on plant growth and development. Perception of signal and transduction mechanisms have not been completely identified, and there is little information about cross-talk between different stress signal transduction pathways in plants (Chinnusamy *et al*., 2003). The identification and characterization of TFs may help to get insight into these signal mechanisms, given that they are the principal effectors of gene expression modulation (Zhang, 2003).

In a previous work (Fusco *et al*., 2005) Cd-regulated genes were identified in *B. juncea*, a plant able to tolerate and accumulate high levels of Cd and other heavy metals in its own tissues (Chaney *et al*., 1997). In particular, plants were treated, in hydroponic solution, for 0, 6, 24 h and 6 weeks with 10 μM $Cd(NO_3)_2$, and numerous differentially expressed genes were found (Fusco *et al*., 2005).

This PhD work has been focused on the characterization of one of these genes and, in particular *BjCdR12*, that is induced after a Cd 6-hour-treatment and so is probably involved in the heavy metals tolerance mechanisms in *B. juncea*. Since *B. juncea* is an allotetraploid species and consequently difficult to study, the attention has been turned on *MYB59* gene of *Arabidopsis thaliana*, homolog to *BjCdR12* (90% homology). *MYB59* presents three splicing variant, called *MYB59.1*, *MYB59.2* and *MYB59.3*. The main objective of this work was the characterization of *MYB59*. Firstly, it has been confirmed the involvement of the gene in Cd stress in *Arabidopsis*. Then, gene expression analysis following abiotic stress exposure was conducted to clarify if this gene is involved in general stress responses and the transcriptional level of the three splicing forms in different plant organs was determined. Furthermore, a study of the effect of the three variant overexpression and the lack of gene expression (obtained through a gene silencing strategy and knock-out mutant plant analysis) was conducted in *Arabidopsis*. At the same time, the analysis of the promoter region was carried out to follow splicing variants localization in different plant organs and tissues. Moreover, the analysis of the methylation pattern of a direct repeat close to TATA box and the sequence of the first intron was conducted in WT plants. To verify the correspondence between mRNAs and proteins and to identify which of the three splicing variants actually occurs, two protein fusion strategy was carried out, using a FLAGtag and HaloTag®. It has been reported that *MYB59* gene undergoes a conserved alternative splicing such as its homolog *MYB48* in *Arabidopsis* (Li *et al*., 2006). Thus, a cross between homozygous single mutant *myb59* and *myb48* was carried out to better understand the role of the two gene as TFs.

Chapter 2

MATERIALS AND METHODS

2.1 Plant material and growth conditions

In this experimental work different *Arabidopsis thaliana* lines were used:

- ecotype Columbia (Col-0) wild-type;
- ecotype Columbia (Col-0) GK-627C09 GABI-Kat mutant line for *MYB59* gene;
- ecotype Landsberg (Ler) GT_5_9575 mutant line for *MYB48* gene.

To identify mutant lines the web sites www.tair.com and http://signal.salk.edu/cgi-bin/tdnaexpress were consulted. Mutant lines were ordered by Nottingham *Arabidopsis* Stock Centre (NASC), available on the web site http://*Arabidopsis*.org.uk/.
Seeds were incubated for three days at 4 °C in the dark, to break seed dormancy, and then transferred in the greenhouse in soil or in hydroponic Hoagland nutrient solution (Hoagland and Arnon, 1938). Plants were grown for different periods of time as indicated for each experiment.
For the *in vitro* experiments, seeds were sterilized with ethanol for 2 min, with sodium hypochlorite 10% and Triton X-100 0.03% for 15 min and rinsed three times with sterile water. Seeds were finally sown on Petri dishes on liquid MS soaked 3MM Whatman paper (Whatman, Maidstone, United Kingdom) or on solid MS (Murashighe and Skoog, 1962) medium (4.4 g/l MS and 30 g/l sucrose).

2.2 Plant sample and treatments

After sterilization, seeds of *Arabidopsis* WT were plated on solid MS medium. Four-week-old plants were grown in hydroponic culture for further two weeks. Different plant tissues were collected: rosette leaves, cauline leaves, stem, closed flowers, open flowers and roots. Samples were frozen in liquid nitrogen and kept at -80 °C.
For cadmium treatment, *Arabidopsis* plants were grown *in vitro* in MS medium for four weeks, grown in hydroponic culture for two weeks and then treated for 0.5 , 2 , 6 and 24 h with 10 μM

$Cd(NO_3)_2$. Untreated plants were used as controls. The plant materials (leaves and roots) were quickly frozen in liquid nitrogen and stored at –80 °C for subsequent RNA extraction.

For stress treatment, *Arabidopsis* plants were grown *in vitro* in MS medium for four weeks and then transferred on Petri dishes on Whatman paper soaked with MS liquid medium (half strength). Stress treatments involved leaves and roots being collected 5 h after thermal treatments (4 °C and 42 °C) and water stress and after the addition of 250 mM NaCl, abscisic acid (ABA), indoleacetic acid (IAA), giberellin (GA_3) and kinetin (kin) to 0.1 mM final concentration. Leaves and roots samples collected were frozen in liquid nitrogen and kept at -80 °C. All samples collected were analyzed by Real Time PCR (see below, in this section).

2.3 RNA extraction and cDNA synthesis

Total RNA was extracted from different plant tissues, using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. First-strand cDNA was synthetized using the Superscript II Rnase H-Reverse Transciptase Kit (Invitrogen) according to the manufacturer's instructions, and then used as the templates for Real-Time RT-PCR analysis.

2.4 Real-Time RT-PCR analysis

Real-Time RT-PCR analysis were performed in triplicate by using the ABI PRISM® 7000 Sequence Detection System (Applied Biosystems, Foster City, CA, USA) with the Platinum® SYBR® Green qPCR SuperMix-UDG (Invitrogen). One pair of specific primers for each of the three splicing variant of *MYB59* gene were used: for *MYB59.1*, MYB1F and MYB1R; for *MYB59.2*, MYB2F and MYB2R; for *MYB59.3*, MYB3F and MYB3R (**Table 2.1**). Gene encoding *β-actin* was used as endogenous standard using the primers: Act1 and Act2 (**Table 2.1**). PCR amplification was performed for 40 cycles, each with 95 °C for 15 s, 59 °C for 30 s and 72 °C for 30 s. The data were organized according to the $2^{-\Delta\Delta CT}$ method for relative gene expression analysis (Livak and Schmittgen, 2001). Real Time PCR analysis was also conducted on *Rd29a* and *PP2CA* gene, using the primers Rd29F, Rd29R and PP2CAF, PP2CAR (**Table 2.1**) on cDNA of WT and mutant plants exposed to drought and ABA stress.

2.5 Mutant plants analysis

Arabidopsis genomic DNA was isolated from fresh material using extraction buffer (0.1 M Tris-HCl (pH 7.5), 0.1 M NaCl, 0.05 M EDTA (pH 8.0), 2% SDS) followed by isopropanol precipitation. T-DNA insertion site or Ds element position were confirmed by a PCR reaction using a combination of insertion-specific and gene-specific primers and then sequencing the fragments obtained at the BMR Genomics sequencing service (www.bmr-genomics.it/). For *MYB59* gene, T-DNA-specific primers were LBb1 and RBb1. For *MYB48* gene, Ds element-specific primers were Ds3-4 and Ds5-1 (**Table 2.1**). To determine knock-out of gene expression, a PCR reaction on cDNA from homozygous mutant plants was performed using gene-specific primers. The two single mutant lines were crossed to obtain the double mutant *myb59myb48.* WT and double mutant plants were compared for developmental differences, growing seedlings vertically *in vitro* for 3 weeks on MS medium with and without sucrose (15 g/l).

2.6 Ectopic expression of *MYB59* splicing variants in plant

Full-length cDNA of the three splicing variants (*MYB59.1*, *MYB59.2* and *MYB59.3*) were amplified using the Platinum® *Pfx* DNA polymerase (Invitrogen) with one pair of specific primers of *MYB59* gene: MYBFL-*Xba*I and MYBFL-*Bam*HI for *MYB59.1* and *MYB59.2*, MYB3-*Xba*I and MYBFL-*Bam*HI for *MYB59.3* (**Table 2.1**). The amplified fragments were cloned into pMD1 expression vector under the control of CaMV-35S promoter sequence. The transformation of *Arabidopsis* WT plants was carried out through floral dip method (Clough and Bent, 1998). Transformed plants were selected *in vitro* for their resistance to kanamicin. In addition, the integration of the transgene in the plant genome was confirmed by a PCR reaction using a combination of promoter-specific and gene-specific primers. Transformed plants will called in this work 35S::*MYB59.1*, 35S::*MYB59.2*, 35S::*MYB59.3.* Through Real Time PCR the lines showing the higher expression level were selected.

2.6.1 Leaf area measure and root comparison

To study the effect of *MYB59* gene expression on plant vegetative growth, seeds of *Arabidopsis* WT, plants overexpressing the three *MYB59* variants and mutant plants were sowed on sterile soil and maintained in controlled conditions (24 °C, 16 h light/8 h dark). Rosetta leaves were photographed every three days starting from the four-leaf stage until the emission of floral stem.

The rosetta leaves areas were measured using the IMAGE ProPlus program. Three independent repeats of this analysis were conducted using about twenty plants for each line. Roots of the same lines were compared, growing seedlings vertically *in vitro* for 3 weeks.

2.7 Gene silencing analysis

In order to cause gene expression decrease, the destination vector pK7GWIWG2(II) (Invitrogen) was used. A 552 bp region at the 3' end of *MYB59* (common to the three splicing variants) was amplified with the Platinum® *Pfx* DNA polymerase (Invitrogen) using the following primers: MYB59SF and MYB59SR to obtain the sense fragment, MYB59ASF and MYB59ASR to obtain the antisense fragment (**Table 2.1**). The pENTR™ Directional TOPO® Cloning Kit (Invitrogen) was used, according to the manufacturer's instructions. The recombination reaction was performed using Gateway® LR Clonase™ II Enzyme Mix Kit (Invitrogen). The final construct was used for *Arabidopsis* WT transformation through floral dipping. Transformed plants were selected *in vitro* for their resistance to kanamicin. In addition, the integration of the transgene in the plant genome was confirmed by a PCR reaction using a combination of gene-specific primers. Transformed plants will called in his work RNAi. Through Real Time PCR the lines showing the lower expression level were selected.

2.7.1 Cadmium content quantification

Arabidopsis plants were grown *in vitro* in MS medium for four weeks, grown in hydroponic culture for two weeks and then treated for two weeks with 10 μM $CdSO_4$. Cadmium content was quantified in leaves and roots.

2.8 Promoter-reporter gene fusion

To analyzed the promoter region of the three *MYB59* splicing variants, the sequence of about 2.0 Kbp upstream each of three ATGs was considered and a Gateway® strategy was used. The Platinum® *Pfx* DNA polymerase (Invitrogen) was used for all amplifications steps.

For the amplification of *MYB59.1* promoter, further steps were needed, because the region upstream the starting codon ATG contains two other ATGs that could cause problems during the

expression of the fusion protein. So, the conversion of these two ATGs in TAG (stop codon) through mutagenesis was performed. Mutations were inserted through a set of three PCR reactions.

First ATG mutagenesis: for PCR1 the primer carrying the mutation Mut3F and FL*Sac*II were used. For PCR2 the primers FL and Mut3R carrying the mutation and complementary to Mut3F for 20 bp were used. For PCR3 the first and the second PCR products were used as template using the primers FL and FL*Sac*II (**Table 2.1**). The fragment obtained was cloned in pGEM® -T Easy vector (Promega, Madison, WI, USA) and sequenced.

Second ATG mutagenesis: for PCR1 the primer carrying the mutation Mut2F and FL*Sac*II were used. For PCR2 the primers FL and Mut2R carrying the mutation and complementary to Mut2F for 20 bp were used. For PCR3 the first and the second PCR products were used as template using the primers FL and FL*Sac*II (**Table 2.1**). The fragment obtained was cloned in pGEM® -T Easy vector (Promega) and sequenced.

Then the promoter region with two mutations was amplified using a set of three PCR reactions. For PCR1 the primers pTOPO and pREV were used. For PCR2 the primers FL/FL*Sac*II were used and for PCR3 the first and the second PCR products were used as template using the primers pTOPO and FL*Sac*II (**Table 2.1**).

The promoter region of *MYB59.2* (2.0 Kbp) was amplified from *Arabidopsis* genomic DNA using the following primers: Topopr1 and MYBpr2 (**Table 2.1**). The pENTR™ Directional TOPO® Cloning Kit (Invitrogen) was used, according to the manufacturer's instructions. The recombination reaction was assembled between the entry clone pENTR™/D-TOPO and the destination vector pKGWFS7 using Gateway® LR Clonase™ II Enzyme Mix Kit (Invitrogen). The final construct was used for *Arabidopsis* transformation through the floral dip protocol. Transformed plants were selected *in vitro* for their resistance to kanamicin. In addition, the integration of the transgene in the plant genome was confirmed by a PCR reaction using a combination of promoter-specific and gene-specific primers.

The promoter region of *MYB59.3* (2.1 Kbp) was cloned in pCAMBIA 1381 (CAMBIA, Canberra, Australia). Seeds of these *Arabidopsis* transformed were kindly provided by Dr. Hongya Gu (Peking-Yale Joint Center for Plant Molecular Genetics and Agro-Biotechnology, National Laboratory of Protein Engineering and Plant Genetic Engineering, Peking University, Beijing, China). Transgenic plants used for promoter analysis will called in this work *pMYB59.1*, *pMYB59.2*, *pMYB59.3*.

2.8.1 *GUS* gene histochemical assay

For detection of GUS expression, transgenic plant material was submerged in 25 mg/ml 5-bromo-4-chloro-3-indoyl glucuronide (X-gluc) in 50 mM phosphate buffer (pH 7.0) in 10% Triton X-100 (v/v) and incubated overnight at 37 °C (Jefferson *et al*., 1987). Chlorophyll was removed by incubating the tissues for several hours in ethanol-acetic acid 3:1. The tissues were examined using a Leica MZ16 F stereomicroscope (Leica Microsystem GmbH, Wetzlar, Germany).

2.9 Analysis of sequence-specific DNA methylation

To determinate the methylation pattern of two specific sequences the bisulfite method was used. Sodium bisulfite is able to cause the target DNA conversion of unmethylated cytosine residues into uracil. PCR amplification of bisulfite-treated DNA results in conversion of uracil to thymine. The methylation pattern of genomic DNA sequences of *Arabidopsis* WT leaves and anthers was established by bisulfite method. Genomic DNA was isolated using a DNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) and subjected to treatment using an Epitect® bisulfite kit (Qiagen) according to the manufacturer's instructions.

After treatment and clean-up of bisulfite converted DNA, PCR reactions was used to analyzed two specific *MYB59* sequences, using the Platinum® *Pfx* DNA polymerase (Invitrogen). A region of 250 bp containing the TATA box element and a direct repeat, and the region of 276 bp that include the first intron sequence were amplified, using respectively the primers PromCH3F, PromCH3R and PromCH3F2, GenCH3R (**Table 2.1**). The two fragments, obtained from leaves and anthers DNA, were cloned in pGEM® -T Easy vector (Promega) and sequenced. At least three plasmids for each the two regions and for leaves and anthers were sequenced.

2.10 Protein expression analysis

To find out if and which proteins are produced in plant from the *MYB59* three splicing variants, two protein fusion strategies were used. The Platinum® *Pfx* DNA polymerase (Invitrogen) was used for all amplifications steps.

2.10.1 FLAGtag strategy

In the first strategy, the FLAGtag peptide sequence was used to create a fusion protein for the expression in *E. coli* and in plant. This tag is a 8-aminoacid peptide aspartate-rich (Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys).

The three ORFs of *MYB59* splicing variants were specifically amplified, fused to the FLAGtag, using different set of primers: ORF1f*Xho*I and ORFrFLAG*Bam*HI for the amplification of *MYB59.1* ORF (**Table 2.1**). For the amplification of *MYB59.2* and *MYB59.3* ORFs, mutagenesis passages were needed, to eliminate the splicing site that can interfere with the expression of a determinate form. In particular for *MYB59.2*, splicing site of *MYB59.3* variant was mutagenized using the following primers: SPL2-1f and SPL2-1r, SPL2-2f and ORFr. The fragment resulting was amplified using ORF2f*Xho*I and ORFrFLAG*Bam*HI. For *MYB59.3*, splicing site of *MYB59.2* variant and the *MYB59.1* start codon (ATG) were mutagenized using the following primers: ORF3f*Xho*I and SPL3-1r, SPL3-2f and ORFr, ORF3f*Xho*I and ATG3-3r, ATG3-4f and ORFrF, ORF3f*Xho*I and ORFrFLAG*Bam*HI (**Table 2.1**). The amplified fragments were cloned into pET15b expression vector (Novagen Inc., Madison, WI, USA) and *E.coli* BL21 (DE3) cells were transformed. The expression of the three proteins was induced in *E. coli* at 37 °C and 21 °C with 1 mM IPTG for 1, 3, 6 and 16 h in LB medium (Sambrook *et al.*, 1989). The best results have been obtained at 37 °C after 1 h IPTG treatment.

The promoter region (about 2.0 Kbp) and the genomic sequence full-length were amplified using the following primers: PROMf*Xho*I and PROMr*Bam*HI, GENf*Bam*HI and ORFrFLAG*Bam*HI (**Table 2.1**). The fragments obtained were cloned in the pPCV812ΔNot-Pily expression vector and the resulting construct was used to transform *myb59* mutant plants through floral dip method. Transformed plants were selected *in vitro* for their resistance to hygromicin. In addition, the integration of the transgene in the plant genome was confirmed by a PCR reaction using a combination of promoter-specific and gene-specific primers.

2.10.2 HaloTag® strategy

In the second strategy the HaloTag® (Promega) sequence was fused to the *MYB59* full-length genomic sequence and cloned into an expression vector under the control of CaMV-35S promoter sequence. A linker sequence was introduced between the *MYB59* sequence and the HaloTag sequence. The HaloTag® (Promega) sequence was amplified from the pHT2 vector using the primers HALO*Kpn*I and HALO*Xho*I (**Table 2.1**). The *MYB59* genomic sequence was amplified using the following primers: MYB59forTOPO, MYB59rev+linker, MYB59rev+linker1, LINKER1

and LINKER2+siti (**Table 2.1**). The fragments obtained were cloned in the pLEELA expression vector kindly provided by Dr. Csaba Koncz (Max Planck Institute for Plant Breeding Research, Cologne, Germany). The resulting construct was used to transform *myb59* mutant plants through floral dip method. Transformed plants were selected in vivo for their resistance to the BASTA herbicide. In addition, the integration of the transgene in the plant genome was confirmed by a PCR reaction using a combination of promoter-specific and gene-specific primers.

2.10.3 Protein extraction

For total protein extraction, bacterial suspension (OD_{600}=0.4) was centrifuge at 8,000 g 5 min at 4 °C and the pellet was resuspended in SDS-loading buffer (6 M Urea, 50 mM Tris-Cl (pH 6.8), 100 mM dithiothreitol, 2% SDS, 10% glycerol, 0.1% bromophenol blue).

Arabidopsis samples were collected and homogenized in solubilisation buffer (100 mM Tris (pH 8), 50 mM EDTA (pH 8), 0.25 mM NaCl, 1 mM DTT and 0.7% SDS). The homogenate was heated at 65°C for 10 min and centrifuged at 16,000 g 10 min at RT to remove cellular debris. Proteins were precipitated with ice-cold acetone for 30 min on ice and resuspended in SDS-loading buffer.

For nuclei-rich extraction, *Arabidopsis* leaves (20 g) were collected and homogenized in HB buffer (25 mM PIPES (pH 7.0), 10 mM NaCl, 5 mM EDTA (pH 8.0), 250 mM sucrose, 0.1% Triton X-100, 0.2 mM PMSF, 20 mM β–mercaptoethanol). The homogenate was filtered through 120 µm and 60 µm nylon filters and through 50 µm CellTrics® filters (Partec GmbH, Münster, Germany) and then centrifuged at 4500 g 20 min at 4 °C. The fractions was resuspended in NRB buffer (50 mM HEPES pH 7.6, 110 mM KCl, 5 mM $MgCl_2$, 50% glycerol, 1 mM DTT, protease inhibitors cocktail) and centrifuged at 2000 g 10 min at 4 °C. After rinsed three times with HB buffer, the nuclear preparation was resuspended in NRB buffer. Nuclear rich preparation was precipitated with ice-cold acetone for 30 min on ice and resuspended in SDS-loading buffer (6 M Urea, 50 mM Tris-Cl (pH 6.8), 100 mM dithiothreitol, 2% SDS, 10% glycerol, 0.1% bromophenol blue).

2.10.4 Protein quantification and SDS-PAGE

To quantified proteins extracted, the DC protein assay (Bio-Rad Laboratories, Hercules, CA, USA) was conducted, using bovine serum albumin (BSA) as standard protein, according to the manufacturer's instructions. Proteins extracted were separated on 12% acrylamide Tris-Glycine SDS-PAGE following standard protocols (Sambrook, *et al*. 1989).

To obtain a density gradient gel a gradientator was used (Acrylamide: running 10-16%, stacking 4%. Sucrose: running 5-17.5%, stacking: no sucrose. Running buffer: lower (+) 0.2 M Tris-HCl pH 8.9, and upper (-) 0.1 M Tris-HCl pH 8.9, 0.1 M Tricine pH 8.9, 0.1% SDS, 1 mM EDTA pH 8.0).

2.10.5 Western Blot analysis

The proteins were transferred to a PVDF membrane at 20 V o/n. The membrane was blocked with 5% non-fat milk in TTBS (50 mM Tris Base pH 7.6, 150 mM NaCl, 0.05% Tween 20) at room temperature for 3 h, washed twice in TTBS for 5 min each and incubated with the specific antibody for 1 h. Then the membrane was washed in TTBS six times for 5 minutes each and the detection was carried out using Amersham ECL Plus™ Western Blotting Detection Reagents (GE Healthcare, Waukesha, WI, USA) as substrate of peroxidase. For the FLAGtag strategy, the ANTI-FLAG M2® monoclonal antibody-peroxidase conjugate (Sigma-Aldrich, Saint Louis, MO, USA) titered at 1:1000 at room temperature was used. For the HaloTag strategy, the Anti-HaloTag® pAb (Promega) titered at 1:1000 and the secondary antibody ECL™ Peroxidase labelled anti-rabbit (GE Healthcare) titered 1: 5000 at room temperature will be used**.**

Table 2.1: Forward and reverse PCR primers. The underlined nucleotides correspond to different restriction sites used for different cloning strategies. The underlined nucleotides CACC were used for the cloning in the pENTR™/D-TOPO.

MYB59F	AACATGGTAGTGGTCGCCTT
MYB59R	GGAAACTACTACAGTCGTCG
MYB1RT-F	ATAGGTATAGGTTTGTTTTGGAA
MYB1RT-R	AAACCTACAACCAAAACCAGGT
MYB2RT-F	GAAACATAAGAATAGGTTTAAACA
MYB3RT-F	AGCGAAAGTTTCAGGTTTAAAC
MYBRT-R	TGGAGTCATCTTACCACGTTT
Rd29F	GGAGCTTTAAGAATATGAGAA
Rd29R	CACAAACAAGGAATTATACCAT
PP2CAF	ATGTGCCTTCGAGGTGCTG
PP2CAR	TCGGAGCTCTGTCTTGCTAG
Act 1	GAACTACGAGCTACCTGATG
Act 2	CTTCCATTCCGATGAGCGAT
LBb1	GCGTGGACCGCTTGCTGCAACTC
RBb1	TCAGTGACAACGTCGAGCAC
Ds3-4	CCGTCCCGCAAGTTAAATAT
Ds5-1	ACGGTCGGGAAACTAGCTCT
MYBFL-*XBa*I	GCTCTAGATACCGTAAAGGACCGTGGAC
MYBFL-*Bam*HI	CGGGATCCACAGTGGGTGGTGATTTTTGA
MYB3-*XBa*I	GCTCTAGATTCTATTGCAGAGAGAAAGAGA
MYB59SF	CACCATGTCTCCTACTTCCTCATCT
MYB59SR	GGAAATAGAGTTCTGACTTGTA
MYB59ASF	ATGTCTCCTACTTCCTCATCT
MYB59ASR	CACCGGAAATAGAGTTCTGACTTGTA
Mut3F	GAGAGTAGAAACTTGTGCAAGAA
FL*Sac*II	CCCCGCGGTGGCGTGAAGCTCAAGGACTAA
FL	TTCTATTGCAGAGAGAAAGAGAG
Mut3R	ATTCTTCTTGCACAAGTTTCTACT
Mut2F	GTTCGGAGATCGAAGTAGGGA
Mut2R	ACTTTCGCTACAAAATCCCTACT
pTOPO	CACCCTAGTCTTAACCTTGGTCTGG
pREV	TCTTTCTCTCTGCAATAGAAAATT
Topopr1	CACCCTAGTCTTAACCTTGGTCTGG
MYBpr2	TTGA GACCAGGATGCAGGTAA
ORF1f*Xho*I	CCCTCGAGATGACTCCACAAGAAGAGCGTTTAGTCC
ORFrFLAG*Bam*HI	CGGGATCCCTACTTGTCGTCATCGTCCTTGTAGTCAAGGCGACCACTACCATGTT
SPL2-1f	TACCGTAAAGGACCGTGGAC
SPL2-1r	CTCCACCTTCAAATCTGAAACT
SPL2-2f	TTTCAGATTTGAAGGTGGAGG
ORFr	AAGGCGACCACTACCATGTT
ORF2f*Xho*I	CCCTCGAGATGGGATTTTGTAGCGAAAGTTTCAGAT
ORF3f*Xho*I	CCCTCGAGATGAAACTTGTGCAAGAAGAATACCGT

SPL3-1r	TGTTTAAACCGGAAACTTTCGC
SPL3-2f	AGTTTCCGGTTTAAACAGAACA
ATG3-3r	TTCTTGTGGAGTGATCTTACCA
ATG3-4f	GTAAGATCACTCCACAAGAAGA
PROMf*Xho*I	CCCTCGAGCTAGTCTTAACCTTGGTCTGG
PROMr*Bam*HI	CGGGATCCAGCTGGTTTATGGGAACCTAT
GENf*Bam*HI	CGGGATCCGAGAAGTAAAATTTTCTATTGCAG
HALO*Kpn*I	CGGGTACCATGGGATCCGAAATCGGTAC
HALO*Xho*I	CCCTCGAGTTAGCCGGCCAGCCCGG
MYB59forTOPO	CACCATGAAACTTGTGCAAGAAGAATA
MYB59rev+linker	CGCCACCACTAAGGCGACCACTACCATGTT
MYB59rev+linker1	CCGCTCCCTCCGCCACCACTAAGGCGAC
LINKER1	ACCCCCTCCGCCGCTCCCTCCGCCACC
LINKER2+siti	CCCTCGAGCGGGTACCACCCCCTCCGCCGCTCC
PromCH3F	TGAGTGTTATATCGCGTAGACA
PromCH3R	CCGAACAAGTGGACAAAGTTG
PromCH3F2	TTGTTCGGAGATCGAAGATGG
GenCH3R	CCATTTGGCGTGAAGCTCAA

Chapter 3

RESULTS

3.1 Alignment and sequence homologies

Through the BLAST algorithm (BLASTN), the homology (E value 2e-20) between *BjCdR12* fragment isolated from *B. juncea* and *MYB59* gene sequence from *Arabidopsis* was found. **Fig. 3.1** shows the sequence of *BjCdR12* fragment isolated from *B. juncea* (Fusco *et al*., 2005).

5'AATGTTTGTTCCAGATCGATCAATCTCTCTCCATATGTCATCCATAGAGTAGTATCCAT
CATCGCGTCTTGATTCATTTTCCCAT 3'

Fig. 3.1: Nucleotide sequence of the *BjCdR12* fragment isolated from *B. juncea* treated with Cd.

3.2 *MYB59* transcription factor

The *MYB59* gene is localized on the *A. thaliana* chromosome 5, at the locus At5g59780 and encodes for a MYB TFs. It was found to undergo a conserved alternative splicing that results in four different spliced transcripts in *Arabidopsis*, producing either MYB-related or R2R3-MYB proteins (Li *et al*., 2006). In this thesis, only three of these transcripts (At5g59780.1, At5g59780.2, At5g59780.3) were considered, being the fourth splicing form very low abundant.
Information about the three splicing form sequences, called in this work *MYB59.1*, *MYB59.2* and *MYB59.3*, is available on the TAIR website: http://www.arabidopsis.org/.
In **Fig. 3.2 a)** the gene structure is shown. *MYB59.1* transcript is the longest, with the first intron unspliced, and encodes a MYB-related protein with a single MYB domain repeat; *MYB59.2* transcript, with the first intron spliced, encodes a MYB-related protein with an only one complete MYB domain repeat; *MYB59.3* transcript, chosen as model ("canonical") sequence, with the first intron spliced, encodes a typical R2R3-MYB protein (Romero *et al*., 1998; Stracke *et al*., 2001). The second intron in all of the three transcripts is removed. In **Fig. 3.2 b)** information regarding the three splicing forms, such as the length of transcripts, proteins and protein molecular weights (MW) is reported.

a

b

Gene name	Gene code	mRNA length (bp)	ORF length (bp)	Protein length (aa)	MW (Da)
MYB59.1	AT5G59780.1	1128	513	170	19481.5
MYB59.2	AT5G59780.2	1035	645	214	24519.3
MYB59.3	AT5G59780.3	1010	708	235	27236.5

Fig. 3.2: Schematic representation of the structure (**a**) and the information (**b**) about the three splicing variants of *MYB59*. **a**: Exons are shown as boxes and introns as lines. The putative open reading frames (ORFs) are indicated by black boxes. The start codon (ATG) and the stop codon (TAG) of each splicing variant are also indicated. **b**: Information regarding the thee splicing variants of *MYB59*.

Choosing *MYB59.3* as gene model, that is the longer ORF, the alternative splicing event forming *MYB59.1* can be referred to the intron retention mechanism, that is the most common type of AS in plants, with a frequency of the 41%, whereas an alternative 5' splice site mechanism may be responsible for *MYB59.2* formation (Barbazuk *et al*., 2008).

In the alignment of the three *MYB59* different transcript sequences, the two alternative splicing sites can be noticed: regions of 87 bp and 110 bp are deleted in *MYB59.2* and *MYB59.3* respectively (**Fig. 3.3**). This alignment was carried out using Clustal W program offered by EMBL-EBI Institute (European Molecular Biology Laboratory) available on the website http://www.ebi.ac.uk/clustalw/, with default parameters.

```
AT5G59780.1       GAGAAGTAAAATTTTCTATTGCAGAGAGAAAGAGAGTTAGAGAAAGAGAGAGAGAGAGAG 60
AT5G59780.2       -------AAAATTTTCTATTGCAGAGAGAAAGAGAGTTAGAGAAAGAGAGAGAGAGAGAG 53
AT5G59780.3       -----------------------------------TTAGAGAAAGAGAGAGAGAGAGAG 24
                                                     ************************

AT5G59780.1       AGAGAGAGAGAGATGAAACTTGTGCAAGAAGAATACCGTAAAGGACCGTGGACAGAACAG 120
AT5G59780.2       AGAGAGAGAGAGATGAAACTTGTGCAAGAAGAATACCGTAAAGGACCGTGGACAGAACAG 113
AT5G59780.3       AGAGAGAGAGAGATGAAACTTGTGCAAGAAGAATACCGTAAAGGACCGTGGACAGAACAG 84
                  ************************************************************

AT5G59780.1       GAGGACATCCTCTTGGTCAACTTTGTCCACTTGTTCGGAGATCGAAGATGGGATTTTGTA 180
AT5G59780.2       GAGGACATCCTCTTGGTCAACTTTGTCCACTTGTTCGGAGATCGAAGATGGGATTTTGTA 173
AT5G59780.3       GAGGACATCCTCTTGGTCAACTTTGTCCACTTGTTCGGAGATCGAAGATGGGATTTTGTA 144
                  ************************************************************
```

```
AT5G59780.1      GCGAAAGTTTCAGGTTTGAAGGTGGAGGGAGAAACATAAGAATAGGTATAGGTTTGTTTT 240
AT5G59780.2      GCGAAAGTTTCAGGTTTGAAGGTGGAGGGAGAAACATAAGAATAGGT------------ 220
AT5G59780.3      GCGAAAGTTTCAGGT---------------------------------------------- 161
                 ***************

AT5G59780.1      GGAAAAATGGGCGGTCTTTGGAAAGGACCTTCCATTTAAAGAAATGACCTGGTTTTGGTT 300
AT5G59780.2      ------------------------------------------------------------
AT5G59780.3      ------------------------------------------------------------

AT5G59780.1      GTAGGTTTAAACAGAACAGGAAAGAGTTGCAGGTTAAGGTGGGTTAATTACCTGCATCCT 360
AT5G59780.2      ------TTAAACAGAACAGGAAAGAGTTGCAGGTTAAGGTGGGTTAATTACCTGCATCCT 274
AT5G59780.3      ------TTAAACAGAACAGGAAAGAGTTGCAGGTTAAGGTGGGTTAATTACCTGCATCCT 213
                       **************************************************

AT5G59780.1      GGTCTCAAACGTGGTAAGATGACTCCACAAGAAGAGCGTTTAGTCCTTGAGCTTCACGCC 420
AT5G59780.2      GGTCTCAAACGTGGTAAGATGACTCCACAAGAAGAGCGTTTAGTCCTTGAGCTTCACGCC 334
AT5G59780.3      GGTCTCAAACGTGGTAAGATGACTCCACAAGAAGAGCGTTTAGTCCTTGAGCTTCACGCC 273
                 ************************************************************

AT5G59780.1      AAATGGGGAAACAGGTGGTCAAAAATTGCCCGGAAATTACCGGGGAGAACAGATAATGAG 480
AT5G59780.2      AAATGGGGAAACAGGTGGTCAAAAATTGCCCGGAAATTACCGGGGAGAACAGATAATGAG 394
AT5G59780.3      AAATGGGGAAACAGGTGGTCAAAAATTGCCCGGAAATTACCGGGGAGAACAGATAATGAG 333
                 ************************************************************

AT5G59780.1      ATAAAGAACTACTGGAGGACTCATATGAGGAAGAAGGCTCAAGAGAAGAAGCGACCTATG 540
AT5G59780.2      ATAAAGAACTACTGGAGGACTCATATGAGGAAGAAGGCTCAAGAGAAGAAGCGACCTATG 454
AT5G59780.3      ATAAAGAACTACTGGAGGACTCATATGAGGAAGAAGGCTCAAGAGAAGAAGCGACCTATG 393
                 ************************************************************

AT5G59780.1      TCTCCTACTTCCTCATCTTCAAACTGTTGCTCATCATCTATGACCACTACTACTAGTCAA 600
AT5G59780.2      TCTCCTACTTCCTCATCTTCAAACTGTTGCTCATCATCTATGACCACTACTACTAGTCAA 514
AT5G59780.3      TCTCCTACTTCCTCATCTTCAAACTGTTGCTCATCATCTATGACCACTACTACTAGTCAA 453
                 ************************************************************

AT5G59780.1      GACACTGGAGGCTCCAACGGGAAAATGAATCAAGAATGCGAAGACGGGTACTACTCCATG 660
AT5G59780.2      GACACTGGAGGCTCCAACGGGAAAATGAATCAAGAATGCGAAGACGGGTACTACTCCATG 574
AT5G59780.3      GACACTGGAGGCTCCAACGGGAAAATGAATCAAGAATGCGAAGACGGGTACTACTCCATG 513
                 ************************************************************

AT5G59780.1      GATGACATATGGAGAGAGATTGATCAGTCTGGAGCAAACGTTATTAAACCGGTAAAAGAC 720
AT5G59780.2      GATGACATATGGAGAGAGATTGATCAGTCTGGAGCAAACGTTATTAAACCGGTAAAAGAC 634
AT5G59780.3      GATGACATATGGAGAGAGATTGATCAGTCTGGAGCAAACGTTATTAAACCGGTAAAAGAC 573
                 ************************************************************

AT5G59780.1      AACTACTACTCAGAGCAAAGCTGTTACTTGAATTTCCCTCCTCTGGCTTCTCCAACATGG 780
AT5G59780.2      AACTACTACTCAGAGCAAAGCTGTTACTTGAATTTCCCTCCTCTGGCTTCTCCAACATGG 694
AT5G59780.3      AACTACTACTCAGAGCAAAGCTGTTACTTGAATTTCCCTCCTCTGGCTTCTCCAACATGG 633
                 ************************************************************

AT5G59780.1      GAAAGTTCCTTGGAATCTATATGGAACATGGATGCAGATGAAAGTAAGATGTCTTCTTTT 840
AT5G59780.2      GAAAGTTCCTTGGAATCTATATGGAACATGGATGCAGATGAAAGTAAGATGTCTTCTTTT 754
AT5G59780.3      GAAAGTTCCTTGGAATCTATATGGAACATGGATGCAGATGAAAGTAAGATGTCTTCTTTT 693
                 ************************************************************

AT5G59780.1      GCTATTGATCAGTTTCCTCTAAGTTTTGAACATGGTAGTGGTCGCCTTTAGTCTAGGATT 900
AT5G59780.2      GCTATTGATCAGTTTCCTCTAAGTTTTGAACATGGTAGTGGTCGCCTTTAGTCTAGGATT 814
AT5G59780.3      GCTATTGATCAGTTTCCTCTAAGTTTTGAACATGGTAGTGGTCGCCTTTAGTCTAGGATT 753
                 ************************************************************

AT5G59780.1      TGATTCATTTGGAATGTTTATATGTGCAGCATATATATGTTATCAAACGACGACTGTAGT 960
AT5G59780.2      TGATTCATTTGGAATGTTTATATGTGCAGCATATATATGTTATCAAACGACGACTGTAGT 874
AT5G59780.3      TGATTCATTTGGAATGTTTATATGTGCAGCATATATATGTTATCAAACGACGACTGTAGT 813
                 ************************************************************
```

```
AT5G59780.1       AGTTTCCTATGACTTACATCAAAAATCACCACCCACTGTACTAATCTCATAAGTAGTCAT 1020
AT5G59780.2       AGTTTCCTATGACTTACATCAAAAATCACCACCCACTGTACTAATCTCATAAGTAGTCAT 934
AT5G59780.3       AGTTTCCTATGACTTACATCAAAAATCACCACCCACTGTACTAATCTCATAAGTAGTCAT 873
                  ************************************************************

AT5G59780.1       CATCTTATGCCTTTGTTTAGTTTGTAGAGTGAGTGAAAAGATGTGTAATACAAGTCAGAA 1080
AT5G59780.2       CATCTTATGCCTTTGTTTAGTTTGTAGAGTGAGTGAAAAGATGTGTAATACAAGTCAGAA 994
AT5G59780.3       CATCTTATGCCTTTGTTTAGTTTGTAGAGTGAGTGAAAAGATGTGTAATACAAGTCAGAA 933
                  ************************************************************

AT5G59780.1       CTCTATTTCCAAAATAAATAGACTTTTGAAGTTTCTGTG 1119
AT5G59780.2       CTCTATTTCCAAAATAAATAGACTTTTGAAGTTTCTGTG 1033
AT5G59780.3       CTCTATTTCCAAAATAAATAGACTTTTGAAGTTTCTGTG 972
                  *************************************
```

Fig. 3.3: Alignment of the three splicing variant sequences of *MYB59* carried out using Clustal W. The two alternative splicing sites can be noticed.

Using the *Swiss-Prot/TrEMBL* database (on the website http://www.expasy.org/) the three putative amino acid sequences were found (entry name Q4JL86_ARATH, Q4JL85_ARATH and Q9FN86_ARATH respectively for MYB59.1, MYB59.2 and MYB59.3 protein). The alignment of these sequences is showed in **Fig. 3.4**. It has been demonstrated that MYB59.2 and MYB59.3 proteins are localized in the nucleus, whereas MYB59.1 is only partially nuclear (Li *et al*., 2006). In fact, two basic regions, predicted to be Nuclear Localization Signals (NLSs, KRGK and RKKAQEKKR), are present in the R3 domain of MYB59.2 and MYB59.3, whereas in MYB59.1 protein only the second can be found (Li *et al*., 2006) (**Fig. 3.4**).

```
                        <------------ R2 ------------>
MYB59.1           ------------------------------------------------------------
MYB59.2           MGFCSESFR----FEG-----------GGR-----NIRIGLNRTGKSCRLRWVNYLHPG 39
MYB59.3           MKLVQEEYRKGPWTEQEDILLVNFVHLFGDRRWDFVAKVSGLNRTGKSCRLRWVNYLHPG 60

                        <------------ R3 ------------>
MYB59.1           -----MTPQEERLVLELHAKWGNRWSKIARKLPGRTDNEIKNYWRTHMRKKAQEKKRPMS 55
MYB59.2           LKRGKMTPQEERLVLELHAKWGNRWSKIARKLPGRTDNEIKNYWRTHMRKKAQEKKRPMS 99
MYB59.3           LKRGKMTPQEERLVLELHAKWGNRWSKIARKLPGRTDNEIKNYWRTHMRKKAQEKKRPMS 120
                    ************************************************************

MYB59.1           PTSSSSNCCSSSMTTTTSQDTGGSNGKMNQECEDGYYSMDDIWREIDQSGANVIKPVKDN 115
MYB59.2           PTSSSSNCCSSSMTTTTSQDTGGSNGKMNQECEDGYYSMDDIWREIDQSGANVIKPVKDN 159
MYB59.3           PTSSSSNCCSSSMTTTTSQDTGGSNGKMNQECEDGYYSMDDIWREIDQSGANVIKPVKDN 180
                  ************************************************************

MYB59.1           YYSEQSCYLNFPPLASPTWESSLESIWNMDADESKMSSFAIDQFPLSFEHGSGRL 170
MYB59.2           YYSEQSCYLNFPPLASPTWESSLESIWNMDADESKMSSFAIDQFPLSFEHGSGRL 214
MYB59.3           YYSEQSCYLNFPPLASPTWESSLESIWNMDADESKMSSFAIDQFPLSFEHGSGRL 235
                  ********************************************************
```

Fig. 3.4: Aligment of the three splicing variant aminoacid sequences. The DNA binding domains are in colour. Conserved tryptophan residues in the MYB domains are undersigned. Note that only MYB59.3 protein shows the three tryptophan residues, typical of the MYB domain, only in the R2 domain repeat. R2 and R3 domain repeats are indicated. The regions corresponding to the predicted NLSs are shown in the boxes.

3.3 *MYB59* gene expression analysis

3.3.1 Gene expression analysis on Cd-treated plants

In a previous work, cDNA-AFLP analysis on *B. juncea* showed that the *BjCd26* expression level increased after a 6-hour-Cd treatment (Fusco *et al.*, 2005). To confirm putative involvement of *MYB59* in Cd stress response in *Arabidopsis,* such as its homolog *BjCdR12* in *B. juncea*, the expression level of *AtMYB59* was measured in WT plants treated for 0.5, 2, 6 and 24 h with 10 µM $Cd(NO_3)_2$ by means of Real-Time PCR (**Fig. 3.5**). For this analysis the primers Act1 and Act 2 for *β-actin* gene, MYB59F and MYB59R for *MYB59* gene were used. The latter are designed on the 3'UTR, that is a region common to all the three splicing variants.

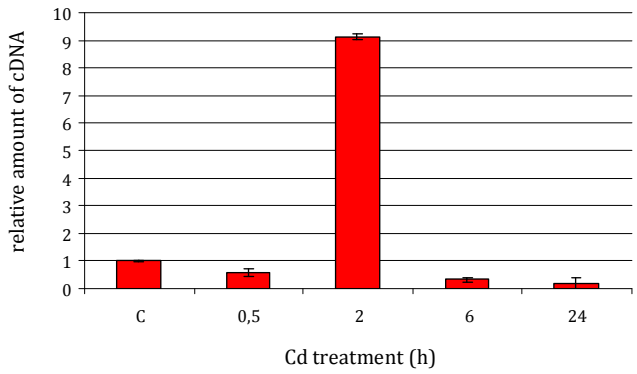


Fig. 3.5: Graphic representation of the transcript quantification by Real-Time PCR carried out on leaves of *Arabidopsis* plants treated with Cd. In the histogram the value $2^{-\Delta\Delta CT}$ for each line was reported. The values were normalized to *β-actin* control. The error bars correspond to standard error (± SE) of three triplicates. PCR amplification was carried out for 40 cycles. **C**: non-treated control.

As shown in **Fig. 3.5**, Cd modulates *MYB59* transcription level in leaves of *Arabidopsis* plants. In particular, the expression profile of this gene does not change after 0.5 h of Cd treatment, it increases after a 2-hour-treatment, and returns to a basal level, similar to the untreated control after a Cd exposure of 6 and 24 h.

This result confirms that the expression of *MYB59* is modulated by Cd such as its homologous *BjCdR12* in *B. juncea*, suggesting that the two genes may have a conserved function in the heavy metal stress response.

3.3.2 Expression analysis of the three transcripts during different abiotic stresses

To clarify whether the expression pattern observed after Cd treatment is specific for Cd stress, or if the transcription of this gene is modulated also in response to other abiotic stresses, the transcription levels of the three splicing variants were measured in leaves and roots of *Arabidopsis* subjected, *in vitro*, to different stress and hormone treatments for 5 hours:

- Low and high temperature (4 °C and 42 °C);
- Water stress (without culture medium);
- Salt stress (250 mM NaCl);
- Hormone treatment (0.1 mM ABA, IAA, GA_3 and kinetin).

The expression pattern analysis of alternatively spliced transcripts was carried out by Real Time PCR (**Fig. 3.7 a** and **b**) using the primers Act1 and Act2 for *β-actin* gene, and *MYB59* splicing variant-specific primers: MYB1RT-F and MYB1RT-R for *MYB59.1*, MYB2RT-F and MYBR for *MYB59.2* and MYB3RT-F and MYBR for *MYB59.3*. These primers were designed on a unique region of each splicing variant. In particular, the primers for *MYB59.1* were designed on the first intron that is not spliced in this form; the primers for *MYB59.2* and *MYB59.3* were designed on the regions of the first intron splice site respectively (**Fig. 3.6**).

Fig. 3.6: Alternative spliced transcript-specific primers. The arrows indicate the primers for each splicing variant: in red for *MYB59.1* (sense and antisense), in orange for *MYB59.2* (sense), in blue for *MYB59.3* (sense) and in green the antisense primer common to *MYB59.2* and *MYB59.3*.

As shown in **Fig. 3.7 a)** and **b)**, various stress conditions alter differently the transcription level of the three splicing variants of *MYB59*. The *MYB59.1* transcription level was found to be not greatly affected by any treatments. The expression pattern of *MYB59.2*, instead, was found to be modulated by cold, drought and ABA treatment both in leaves and roots. In fact, *MYB59.2* transcription increased of than 4-fold in leaves and about 7-fold in roots after a low temperature exposure in comparison to the untreated control, about 3-fold after water deficit in both leaves and roots, and about 3-fold in leaves and 2-fold in roots after ABA treatment when compared with the untreated control. *MYB59.3* was expressed mainly after drought stress both in leaves and roots. It was observed an increase in expression level of 2-fold in leaves and 3,5-fold in roots, in respect to the untreated control, during water deficit conditions. These results suggest that the three splicing forms of *MYB59* may be involved in multiple signalling pathways, responding to different kind of stress.

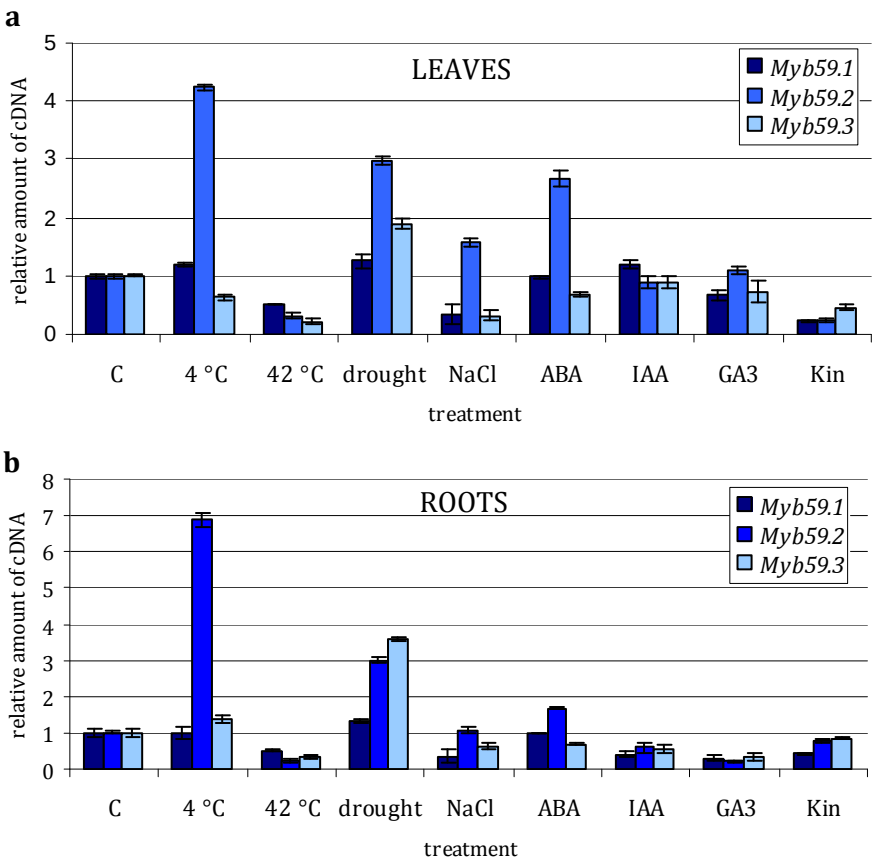


Fig. 3.7: Graphic representation of the transcript quantification by Real-Time PCR carried out on leaves (**a**) and roots (**b**) of *Arabidopsis* plants treated with different stress for 5 hours. In the histogram the value $2^{-\Delta\Delta CT}$ for each line was reported. The values were normalized to *β-actin* control. The error bars correspond to standard error (± SE) of three triplicates. PCR amplification was carried out for 40 cycles. **C**: non-treated control.

3.3.3 Expression analysis of the three transcripts in different plant organs

To gain insight into the expression pattern of the three *MYB59* splicing variants, the transcription profiles of these forms were examined, by Real Time PCR using the transcript-specific primers described above, in different plant organs: rosetta leaves, cauline leaves, stem, closed flowers, open flowers and roots (**Fig. 3.8 a, b, c**).

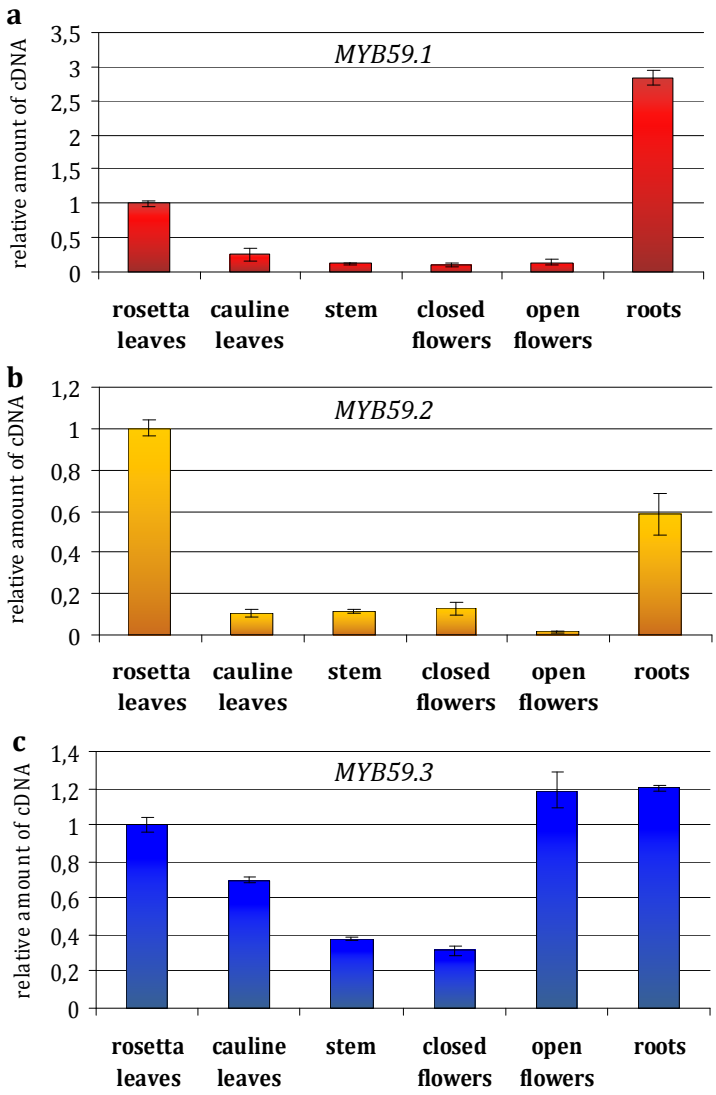


Fig. 3.8: Graphic representation of the transcript quantification by Real-Time PCR carried out on different organs of *Arabidopsis* plants. Expression level of **a***: MYB59.1;* **b**: *MYB59.2*; **c**: *MYB59.3*. In the histogram the value $2^{-\Delta\Delta CT}$ for each line was reported. The values were normalized to *β-actin* control. The error bars correspond to standard error (± SE) of three triplicates. PCR amplification was carried out for 40 cycles. Rosetta leaves were chosen as reference sample.

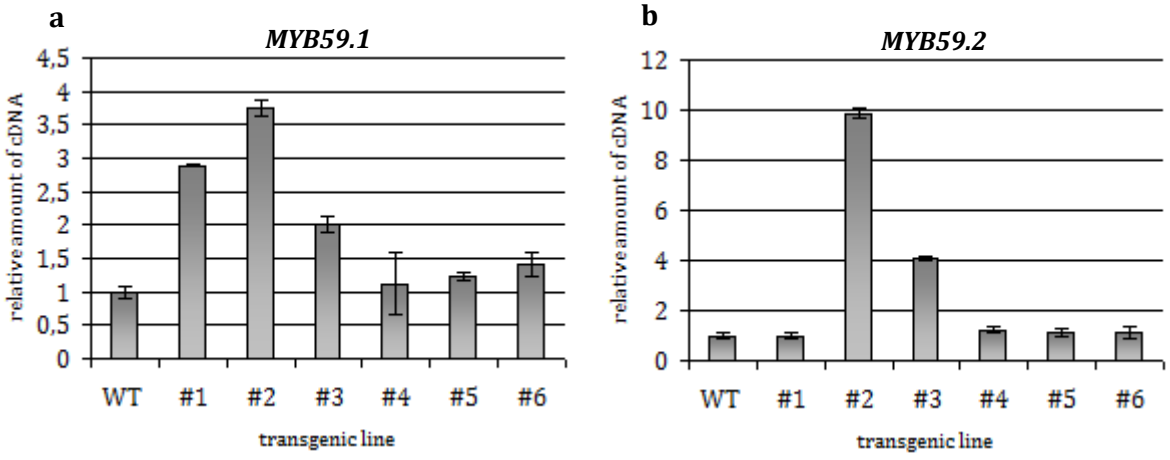
The expression levels in the different organs were normalized respect to rosetta leaves sample, chosen as reference. Interestingly, the results indicated that the three alternatively spliced

variants were expressed in typical organ-specific patterns. In fact, *MYB59.1* was expressed mainly in roots with a transcription level 3-fold higher than in rosetta leaves and with a much lower transcription in all the other organs considered (**Fig. 3.8 a**). The expression of *MYB59.2*, instead, was lower in all the samples in respect to rosetta leaves (**Fig. 3.8 b**), whereas *MYB59.3* was expressed at a little higher levels in open flowers and roots in comparison with rosetta leaves (**Fig. 3.8 c**).

3.4 Study of overexpression and lack of gene expression

3.4.1 Selection of transgenic plants

To further characterize the *MYB59* gene, the effects of over-expression and lack of expression were considered. Full-length cDNA of the three splicing variants were amplified using specific primers and cloned into the pMD1 vector under the control of the CaMV-35S promoter, in order to obtain the ectopic expression of the three transcripts in all plant organs. Transformed *Arabidopsis* plants were analyzed, through Real Time PCR, to select lines showing the higher expression level of each splicing variants (**Fig. 3.9 a**, **b** and **c**). Moreover, since at this time of the experimental work mutant lines were not available on the TAIR websites (www.tair.com) and http://signal.salk.edu/cgi-bin/tdnaexpress, a construct was preparing for studying the effect of shut-down expression of *AtMYB59*. By means of the Gateway® technology, a region of the last exon of *AtMYB59* was cloned in the pK7GWIWG2(II) vector, in both sense and in antisense, in order to cause gene silencing. Transformed *Arabidopsis* plants were analyzed, by Real Time PCR, to select lines showing the lower gene expression profile. RNAi line presenting 98% decrease in expression was considered for the further analysis (**Fig. 3.9 d**).



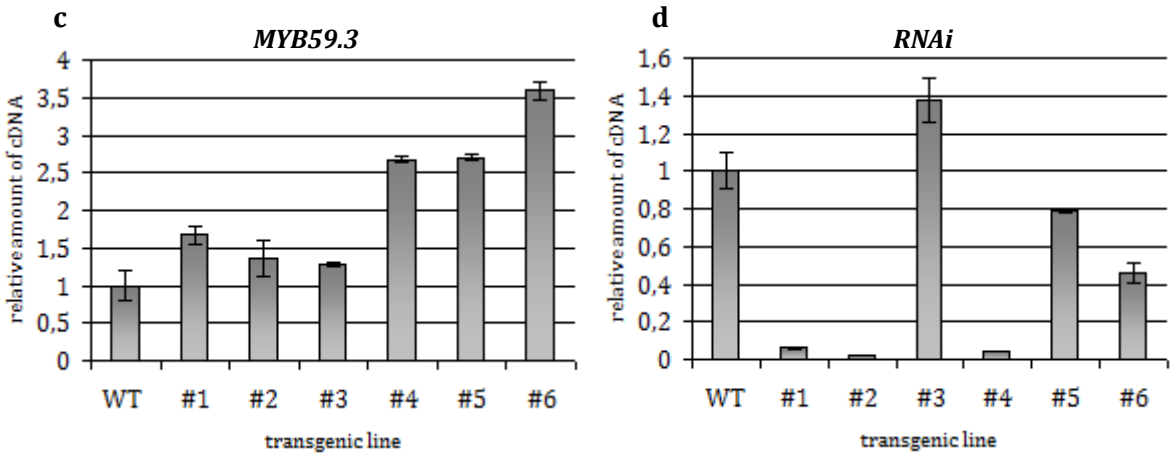


Fig. 3.9: Graphic representation of the transcript quantification by Real-Time PCR carried out on different transgenic lines, to determine the lines with the higher (**a**: for *MYB59.1*, **b**: for *MYB59.2* and **c**: for *MYB59.3*) and the lower (**d**: for RNAi) expression level in comparison with non transformed plant. In the histogram the value $2^{-\Delta\Delta CT}$ for each line was reported. The values were normalized to *β-actin* control. The error bars correspond to standard error (± SE) of three triplicates. PCR amplification was carried out for 40 cycles. WT plant for each transformed line was chosen as reference sample.

On the basis of these results, transgenic lines for the further analysis were selected. In particular, the line representative for the overexpression of *MYB59.1* was #2, of *MYB59.2* #2 and of *MYB59.3* #6. The line representative for the silencing of the gene was #2 (**Fig. 3.9 a**, **b**, **c** and **d**).

3.4.2 Analysis of T-DNA insertion mutant for *MYB59*

After testing the RNAi effect, one line carrying a T-DNA insertion in *AtMYB59* was identify by browsing the T-DNA Express database available on the website http://signal.salk.edu/cgi-bin/tdnaexpress. The mutant line GK-627C09 was analyzed. This line belongs to the GABI-Kat collection created in background *Arabidopsis* Columbia (Col-0). As described in Materials and Methods, the T-DNA insertion site was confirmed by PCR, using gene and T-DNA specific primers, followed by sequencing of the PCR fragment. The insertion of T-DNA is in the second intron of the gene (**Fig. 3.10 a**). To determine whether the insertion effectively interrupted the gene transcription, RT-PCR on cDNA from homozygous mutant plants was performed using gene-specific primers (**Fig. 3.10 b**).

Fig. 3.10: Schematic representation of the position of T-DNA in the *MYB59* gene and its effect on transcript accumulation. **a**: Exons (boxes), as well as intron sequences (lines) are depicted. The T-DNA insertion site and orientation are provided and drawn not in scale. **b**: Transcript levels were detected by RT-PCR that was carried out using gene- and *β-actin* specific primer (the latter used as control).

Since AtMYB59 is a transcription factor, it could be involved in abiotic stress response, regulating the expression of specific genes that enable plants to tolerate/acclimate to the stress condition. It is known that *Rd29a* (*RESPONSE TO DESICCATION 29A*) is induced very quickly by drought stress (Yamaguchi-Shinozaki and Shinozaki, 1993) and its mRNA level changes differently in response to dehydration, cold, salt stress and ABA exposure (Yamaguchi-Shinozaki and Shinozaki, 1994; Ingram and Bartels, 1996). *PP2CA* (*PROTEIN PHOSPHATASE 2CA*) gene is induced by water deficit, low temperature, salinity (Tähtiharju and Palva, 2001), and plays a key role in ABA signal transduction (Rodriguez, 1998). Recently, it has been demonstrated that a type 2C protein phosphatase is able to regulate ABA-activated protein kinases in *Arabidopsis* (Umezawa *et al*., 2009). Besides, it has been shown that AtPP2CA protein strongly regulates ABA signalling during germination (Yoshida *et al*., 2006). Considering these finding, and the fact that *MYB59.2* splicing form was shown to be involved in the response to cold, drought and ABA treatment and *MYB59.3* in drought stress (**Fig. 3.7 a** and **b**), the transcriptional levels of *Rd29a* and *PP2CA* genes were measured in WT and *myb59* plants after stress exposure. The measure of *Rd29a* and *PP2CA* transcription profile in plants that lacks *MYB59* transcription, could give an indication of the signalling pathways in which this TF plays a role.

As described in Materials and Methods, WT and mutant plants were therefore subjected to *in vitro* water stress (without culture medium) and ABA treatment for 5 h. By Real Time PCR, the transcription patterns of *Rd29a* (**Fig. 3.11 a**) and *PP2CA* (**Fig. 3.11 b**) were followed.

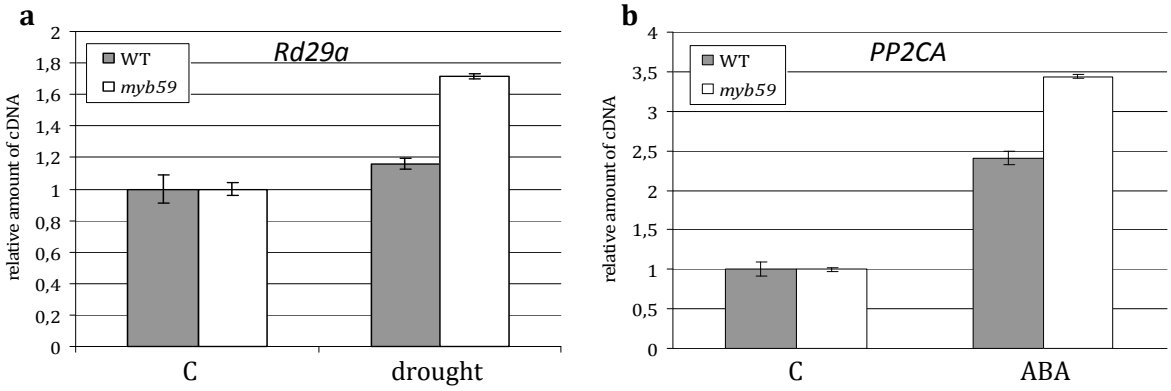
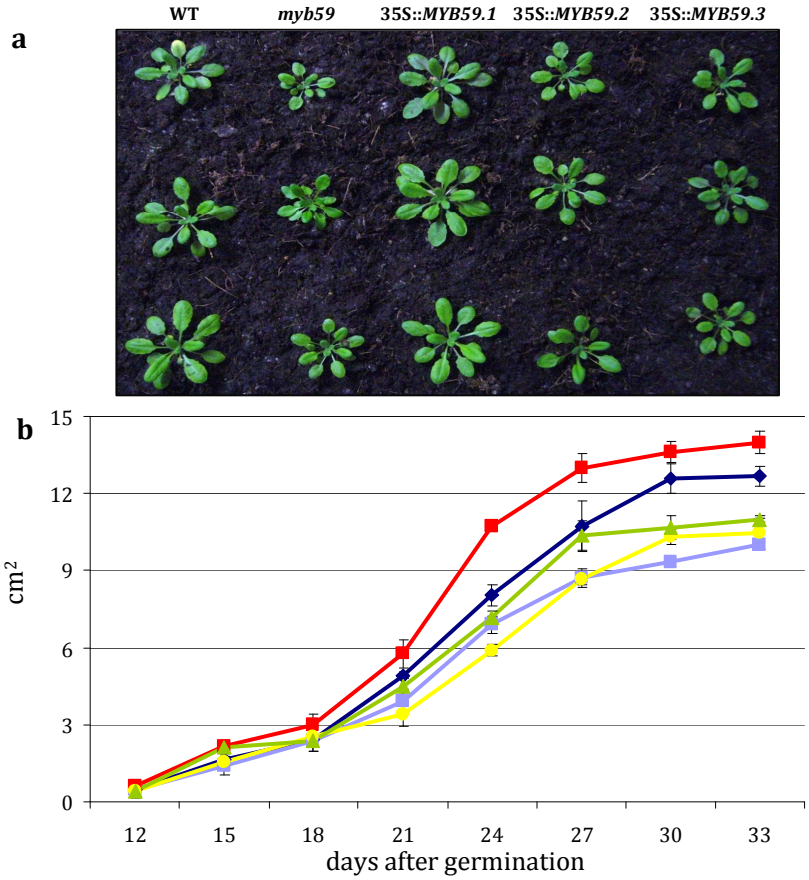


Fig. 3.11: Graphic representation of the transcript quantification by Real-Time PCR carried out on leaves of *Arabidopsis* WT and mutant plants treated with water (**a**) and ABA (**b**) stress for 5 hours. In the histogram the value $2^{-\Delta\Delta CT}$ for each line was reported. The values were normalized to *β-actin* control. The error bars correspond to standard error (± SE) of three triplicates. PCR amplification was carried out for 40 cycles. **C**: non-treated control.

As can be noticed in **Fig. 3.11**, the transcription level of both genes increased after stress treatments, but their expression was higher in mutant plants than in WT plants. The lack of *MYB59* expression, therefore, caused an increase of *Rd29a* and *PP2CA* transcription levels. This result may suggest that *MYB59* is really involved in the signalling transduction pathways that regulate stress response to drought and ABA treatment, even if its role in the mediation and coordination of stress signals is not completely clear.

3.5 Phenotypic analysis of *MYB59* expression

Comparison between WT, overexpressing and mutant plants showed that *MYB59* affects the vegetative growth. In fact, growth alteration was observed in plants grown on soil for four weeks under standard condition (24 °C, 16 h light/8 h dark) (**Fig. 3.12 a**). As described in Materials and Methods, to quantify the differences between the five lines considered, leaf area of *in vivo* plants was measured during plant growth from germination until the emission of stem, using the IMAGE ProPlus program (**Fig. 3.12 b**).



12	0,44 ±0,28	0,47 ± 0,11	0,64 ± 0,20	0,40 ± 0,07	0,41 ± 0,07
15	1,65 ± 0,30	1,39 ± 0,34	2,14 ± 0,15	1,54 ± 0,23	2,11 ± 0,24
18	2,42 ± 0,44	2,35 ± 0,15	3,00 ± 0,39	2,53 ± 0,40	2,37 ± 0,44
21	4,90 ± 0,13	3,91 ± 0,46	5,75 ± 0,55	3,41 ± 0,46	4,46 ± 0,31
24	8,03 ± 0,43	6,88 ± 0,32	10,69 ± 0,07	5,89 ± 0,23	7,17 ± 0,27
27	10,73 ± 0,94	8,72 ± 0,36	12,98 ± 0,56	8,67 ± 0,22	10,33 ± 0,58
30	12,57 ± 0,59	9,32 ± 0,14	13,59 ± 0,41	10,29 ± 0,31	10,66 ± 0,48
33	12,65 ± 0,40	9,98 ± 0,13	13,97 ± 0,44	10,45 ± 0,60	11,00 ± 0,11

Fig. 3.12: Comparison between WT, overexpressing and mutant plants. **a**: Plants were grown on soil for 4 weeks. **b**: Graphic representation and data of rosetta leaf areas for each lines. The error bars represent standard deviation (± SD) of three triplicates.

As shown in **Fig. 3.12 a)**, mutant plants presented a reduced leaf area compared with the control, whereas plants overexpressing *MYB59.1* showed rosetta leaves bigger than WT. Plants overexpressing *MYB59.2* and *MYB59.3* had sizes in-between WT and mutant plants. In **Fig. 3.12 b)**, a graphical representation of leaf area measures was reported. As shown, 33 days after germination, mutant plants had rosetta leaves smaller of about 20% than that in WT plants. Interestingly, plants expressing ectopically *MYB59.1*, showed differences in leaf area in respect to the other lines considered already after 12 days from germination and these differences were maintained during growth. Roots of the same lines were compared, growing seedlings vertically *in vitro* for 3 weeks, but the observed differences were not statistically significant (**Fig. 3.13**).

Fig. 3.13: Comparison between root growth of WT, overexpressing and mutant plants. Plants were grown vertically *in vitro* for 3 weeks.

3.6 Effect of *MYB59* expression on Cd transport

To further investigate the role of *MYB59* in response to Cd-induced stress, WT, overexpressing and RNAi plants were analyzed for their Cd content after a 2-week-treatment with 10 µM $CdSO_4$. Cd content values in leaves and roots is represented in **Fig. 3.14**. T-student test was applied to verify the statistical significance of measured values.

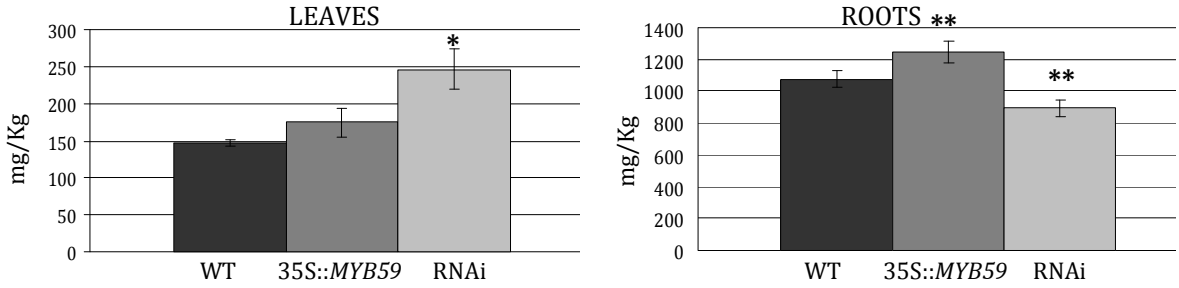


Fig. 3.14: Graphical representation of the values obtained from the quantification of Cd in leaves (on the left) and in roots (on the right) of WT, overexpressing and RNAi plants treated with $CdSO_4$ for 2 weeks. 35S::*MYB59* correspond to the lines over-expressing the three splicing variants, that has been grouped in a single histogram, due to their similar Cd content. The error bars correspond to standard deviation (± SD) of three independent repeats. * indicates a significant ($P<0.05$) difference among transcript levels in Cd-treated plants respect to control plants. ** indicates that the difference is highly significant ($P<0.01$).

As noticed in **Fig. 3.14**, compared lines showed a different accumulation of Cd in leaves and roots. In particular, while in control plants Cd content in shoots reached 150 mg/Kg, in over-expressing plants the accumulation of this metal was about 170 mg/Kg, and in plants with the suppression of *MYB59* expression (RNAi) the content was further increased (about 250 mg/Kg). Through t-student test, it was determined that the difference in Cd content in leaves of RNAi plants was statistically significant respect to WT. Considering roots, the accumulation of Cd in

35S::*MYB59* plants was greater in comparison to control, whereas in RNAi plants was lower, and both the values resulted statistically highly significant from the t student test. Summarizing, in the aerial part of the plant, the drastic decrease of MYB59 protein resulted in a higher accumulation of the metal, whereas in roots the over-expression of all the three splicing variants resulted in a greater Cd content respect to control plants. This result provides an indication of the involvement of MYB59 transcription factor in heavy metal root-to-shoot transport. It can be hypothesized that *MYB59* may be a negative regulator (direct or indirect) of heavy metal transporters. So, when the gene is absent Cd can be transported to leaves where it accumulates, whereas when the gene is overexpressed, due to a possibly diminished root-to-shoot transport, this metal accumulates principally in roots.

3.7 Promoter region analysis

To localized the expression of the three alternatively spliced variants, three constructs containing the region upstream (about 2.0 Kbp) the ORF of each splicing form were prepared and used for *Arabidopsis* transformation. For the analysis of the promoter region regulating *MYB59.1*, mutagenesis of the start codon, ATG, of *MYB59.2* and *MYB59.3*, both upstream the starting codon of this variant, was performed. This step was necessary because the ATGs of *MYB59.3* and *MYB59.1* are in frame and this region contains the entire sequence of the first intron. So, these two ATGs were converted in the TAG stop codons, through a set of three PCRs performed using primers carrying the point mutation, to assure the expression of the correct fusion protein (**Fig. 3.15**).

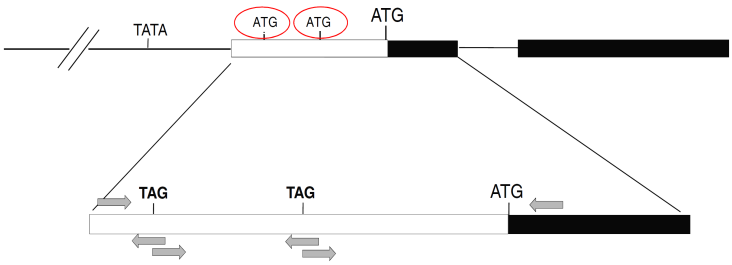


Fig. 3.15: Schematic representation of the technique used to change the two ATGs in TAGs. The arrows indicate the primers used in the set of the three PCRs necessary to introduce the mutations desired.

Through the Gateway® technology, the region upstream *MYB59.1* and *MYB59.2* ORFs were cloned in the pKGWFS7 vector containing the GUS gene coding sequence. *Arabidopsis*

transformation was confirmed through PCRs using promoter-specific and gene-specific primers. Seeds of *Arabidopsis* transformed with the promoter region of *MYB59.3* were kindly provided by Dr. Hongya Gu (Peking-Yale Joint Center for Plant Molecular Genetics and Agro-Biotechnology, National Laboratory of Protein Engineering and Plant Genetic Engineering, Peking University, Beijing, China). In **Fig. 3.16** a scheme of the three promoter-GUS fusion constructs is depicted.

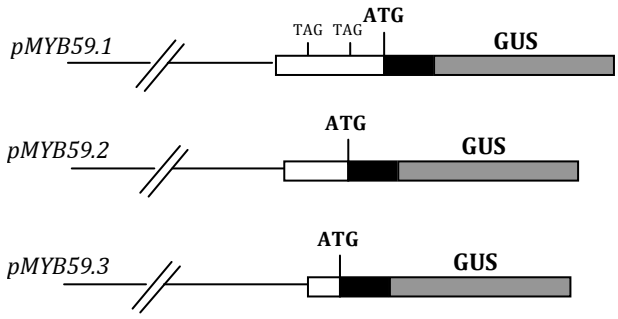
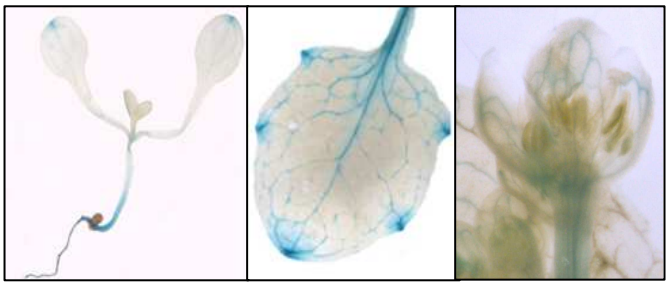


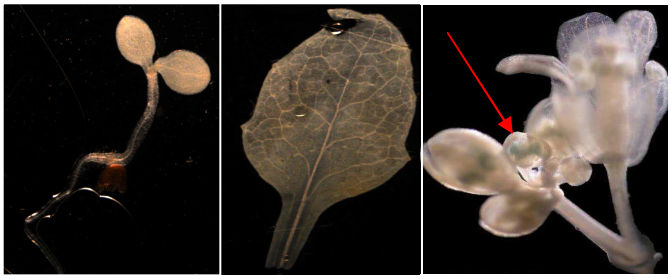
Fig 3.16: Schematic representation of the three promoter-GUS fusion constructs. A short sequence immediately downstream of the starting ATG (in bold) were kept for each construct, to maintain the complete starting context. The scheme is drawn not in scale.

Histochemical assays of GUS activity in plants transformed with the three different constructs, grown on soil under standard conditions, were performed as described by Jefferson *et al*. (1987). The GUS assay allowed to localize the expression of the three splicing variants in different plant organs. *MYB59.1* expression was found mainly in leaf veins and hydathods both apical and lateral (**Fig. 3.17 a**). The expression of *MYB59.2* was found only in anthers of closed flowers (**Fig. 3.17 b**), whereas *MYB59.3* expression was detected in vegetative tissues and sepals (**Fig. 3.17 c**). It can be hypothesized that the three alternatively spliced variants, having different localizations, may also play different roles in plant.

a



b



c

Fig. 3.17: Expression localization of the three splicing variants of *MYB59* using histochemical GUS assay, confirmed by observing at least three different transgenic lines for each construct. The analysis was carried on seedling, rosetta leaves and inflorescence for each construct. **a:** *pMYB59.1*, **b**: *pMYB59.2*, **c**: *pMYB59.3*.

3.8 Analysis of a direct repeat sequence-specific DNA methylation

Analysis of the methylation pattern of a specific sequence may give an indication on its expression. In fact, the methylation mechanism is used to protect the genome against transposons and to modulate the expression of endogenous genes (Goll and Bestor, 2005).

Because methylation occurs mainly in repeat regions (Law and Jacobsen, 2009), a repeat sequence in the promoter region of *MYB59*, just upstream the *MYB59.3* ATG starting codon, found through the *Tandem Repeats Finder* program (Benson, 1999; website http://tandem.bu.edu/trf/trf.html) was considered. Such region of 250 bp, amplified using the primers PromCH3F and PromCH3R , also contains the TATA box element. Moreover, a region of 276 bp, including the first intron sequence, was analyzed using the primers PromCH3F2 and GenCH3R (**Fig. 3.18**).

TGAGTGTTATATCGCGTAGACACAACACAGTATTCAAGTTGGCT**TATA**TAAATATTGGTAAGAGGAATGCCATAG
GTTCCCATAAACCAGCTGAGAAGTAAAATTTTCTATTGC**AGAGAGAAAGAGAGTTAGAGAAAGAGAGAGAGAGAG**
AGAGAGAGAGAGAGATGAAACTTGTGCAAGAAGAATACCGTAAAGGACCGTGGACAGAACAGGAGGACATCCTCT
TGGTCAACTTTGTCCACTTGTTCGGAGATCGAAGATGGGATTTTGTAGCGAAAGTTTCAGGTTTGAAGGTGGAGG
GAGAAACATAAGAATAGGTATAGGTTTGTTTTGGAAAAATGGGCGGTCTTTGGAAAGGACCTTCCATTTAAAGAA
ATGACCTGGTTTTGGTTGTAGGTTTAAACAGAACAGGAAAGAGTTGCAGGTTAAGGTGGGTTAATTACCTGCATC
CTGGTCTCAAACGTGGTAAGATGACTCCACAAGAAGAGCGTTTAGTCCTTGAGCTTCACGCCAAATGG

Fig. 3.18: Sequence that was considered to study the methylation pattern in leaves and anthers of *Arabidopsis* WT. The putative TATA box and the ATGs of the three splicing variants are indicated in red, whereas the primer sequences are underlined. The direct repeat is indicated in bold.

The methylation status of individual cytosine residues in these sequences was analyzed applying the bisulfite method. In fact, sodium bisulfite is able to convert cytosine residues to uracil, but does not affect 5-methylcytosine residues. PCR amplification of bisulfite-treated DNA results in conversion of uracil to thymine. Thus, the methylation pattern of these genomic DNA sequences of *Arabidopsis* WT in leaves and anthers was established by this method. After treatment and clean-up of converted DNA, PCRs using the Platinum® *Pfx* DNA polymerase were performed to analyzed the two specific *MYB59* sequences above. The two fragments, obtained from leaves and anthers sequence, were cloned in pGEM®-T Easy vector and sequenced. A minimum of three plasmids were extracted from different colonies and the inserts were sequenced for both leaves and anthers. The obtained sequence were aligned with the Clustal W program, to highlight differences in methylation pattern (**Fig. 3.19** and **3.20**).

```
sequence        ----------------------------------------------------TGAGTGTTA 9
leaves          ---NNNNNNCTCCCGGCCGCCATGGCGGCCGCGG-GAATTC------GATTTGAGTGTTA 50
anthers         GAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCACTAGTGATTTGAGTGTTA 60
                                                             *********

sequence        TATCGCGTAGACACAACACAGTATTCAAGTTGGCTTATATAAATATTGGTAAGAGGAATG 69
leaves          TATCGCGTAGACATAACACAGTATTCAAGTTGGCTTATATAAATATTGGTAAGAGGAATG 110
anthers         TATCGCGTAGACCCAACCCAATATTCAAATTAACTTATATAAATATTAATAA-AAAAATA 119
                ************  *** ** ******* **  **************  *** *  ***

sequence        CCATAGGTTCCCATAAACCAGCTGAGAAGTAAAATTTTCTATTGCAGAGAGAAAGAGAGT 129
leaves          CCATAGGTTCCCATAAACCAGCTGAGAAGTAAAATTTTCTATTGCAGAGAGAAAGAGAGT 170
anthers         CCATAAATTCCCATAAACCAACTAAAAAATAAAATTTTTTTTTCCAAAAAAAAAAAAAAT 179
                *****  ************** ** * ** ********* * ** ** * * *** * * *

sequence        TAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGAAACTTGTGCAAGAAGAATACC 189
leaves          TAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGA--TGAAACTTGTGCAAGAAGAATACC 228
anthers         TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA--TAAAACTTATACAAAAAAAATACC 237
                ** * *** * * * * * * * * * * * * *  * ****** * *** ** ******

sequence        GTAAAGGACCGTGGACAGAACAGGAGGACATCCTCTTGGTCAACTTTGTCCACTTGTTCG 249
leaves          GTAAAGGACCGTGGACAGAACAGGAGGACATCCTCTTGGTCAACTTTGTCCACTTGTTCG 288
anthers         ATAAAAAACCGTAAACAAAACAAAAAACATCCTCTTAATCAACTTTGTCCACTTGTTCG 297
                 ****  *****  *** ****  *  **********  ********************

sequence        G---------------------------------------------------------- 250
leaves          GAATCACTAGTGAATTCGCGGCCGCC-TGCAGGTCGACCATATGGGAGAGCTCCCAACGC 347
anthers         GAATC----GAATTCCCGCGGCCGCCATGGCGGCCGGGAGCATGCGANNN---------- 343
                *
```

Fig. 3.19: Alignment between the bisulfite-treated sequence containing the direct repeat from leaves and anthers of *Arabidopsis* WT and the same sequence such as reported in the public database (named "sequence" in the alignment). The ATG starting codons of *MYB59.3* and *MYB59.2* splicing variants are depicted in red. In anther sequence, the nucleotides that have been modified by bisulfite treatment are shown in yellow. The single nucleotide converted in leaf sequence is represent in blue.

As can be noticed by the alignment in **Fig. 3.19**, the majority of C residues in anther sequence were converted to T by bisulfite, whereas the same sequence in leaves maintained the pattern as in the sequence from the database. Thus, given that sodium bisulfite is able to modify

unmethylated C residues, then C nucleotides in the direct repeat from anther sequence may be not methylated. The result may signify that this repeat in the promoter region is highly methylated in leaves, whereas it is not in anthers.

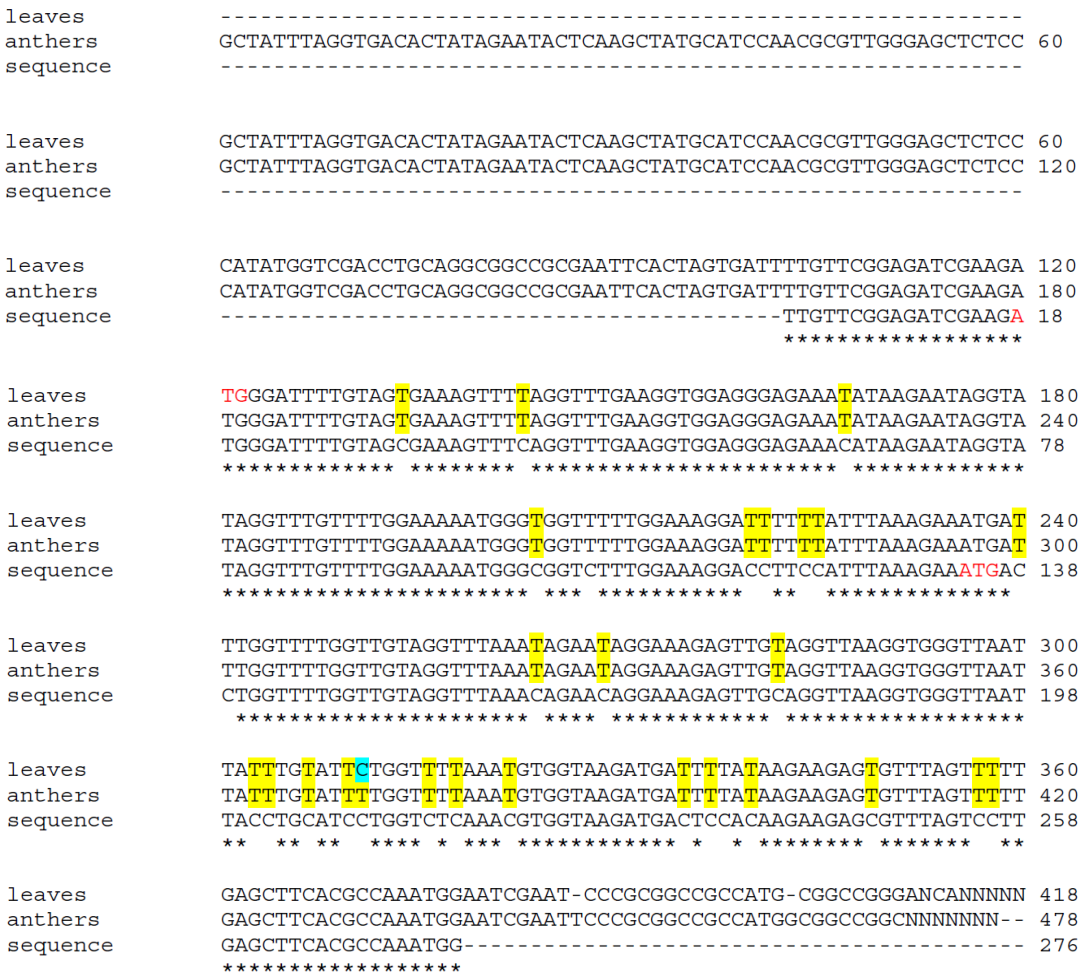


Fig. 3.20: Alignment between the bisulfite-treated sequence containing the first intron sequence from leaves and anthers of *Arabidopsis* WT and the same sequence such as reported in the public database (named "sequence" in the alignment). The ATG starting codons of *MYB59.2* and *MYB59.1* splicing variants are depicted in red. The nucleotides modified by bisulfite treatment are shown in yellow. In leaf sequence, the single nucleotide converted is represent in blue.

As supported by **Fig. 3.20,** in the sequence considered all C residues were modified by bisulfite in comparison with the sequence from the public database, but there was no differences in the methylation status between leaves and anthers (with the exception of one C residue). This result may suggest that this region, which contains the entire sequence of the first intron, does not undergo a different expression regulation in the two organs considered.

3.9 Protein expression analysis

To find out if a correspondence between mRNAs and proteins exists and to identify which of the three splicing variants is actually translated in plant, two protein fusion strategies were used using the FLAGtag and the HaloTag®.

In the first strategy, the FLAGtag peptide sequence was used to create a fusion protein for the expression in *E. coli* and plant. For the expression in bacteria, this tag was fused to the three DNA fragments carrying the ORFs of *MYB59* splicing variants, with mutagenesis steps for *MYB59.2* and *MYB59.3* ORFs, in order to eliminate the splicing recognition site that could interfere with the expression of a determinate form. The ORFs were cloned into the pET15b vector and their expression was induced in *E.coli* BL21 (DE3) cells at 37 °C with 1 mM IPTG for 1h (Sambrook *et al.*, 1989). After separation in SDS-PAGE, protein extracts were transferred to a PVDF membrane and incubated with anti-FLAG antibody-peroxidase conjugate. The result of Western blot analysis was shown in **Fig. 3.21**. As can be noticed in the figure, the three MYB proteins were expressed in *E.coli* also without IPTG induction and showed the expected MW.

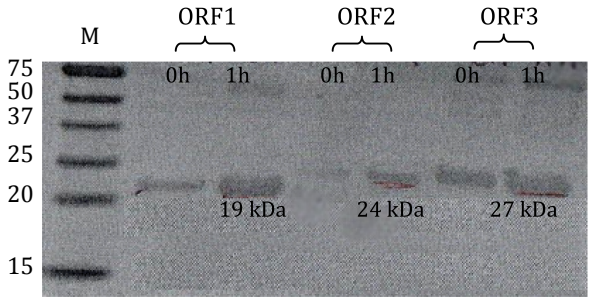


Fig. 3.21: Western blot analysis on protein extracts from *E.coli* after 1h induction with IPTG at 37 °C. **0h**: untreated control. The MW of the three proteins are indicated. M: marker. The MWs reported are in KDa.

For the expression in plant, a fusion protein was create joining the *MYB59* genomic sequence full-length and the FLAGtag. The chimeric protein was cloned downstream the native *MYB59* promoter region (about 2.0 Kbp) in the pPCV812ΔNot-Pily vector and the construct was used to transform *myb59* mutant plants. Proteins were extracted from leaves, flowers and roots of transgenic plants both under standard conditions and after a 1h-cold (4 °C) treatment. After separation in SDS-PAGE using a density gradient gel, proteins were transferred to a PVDF membrane and incubated with anti-FLAG antibody-peroxidase. In **Fig. 3.22 a)** the result of Western blot analysis is shown. In plants, aspecific signals arise in all samples and in both conditions. The signal of about 30 kDa, that can be attributed to the MYB59.3 protein, was also found in the WT lane, even if at very low level.

Considered that MYB59 proteins are nuclear localized (Li *et al.*, 2006), a nuclei-rich extraction

from *Arabidopsis* leaves was carried out. Unfortunately, also in this case only the presence of unspecific signal was found (**Fig. 3.22 b**). Since the results were not reliable, this strategy was abandoned.

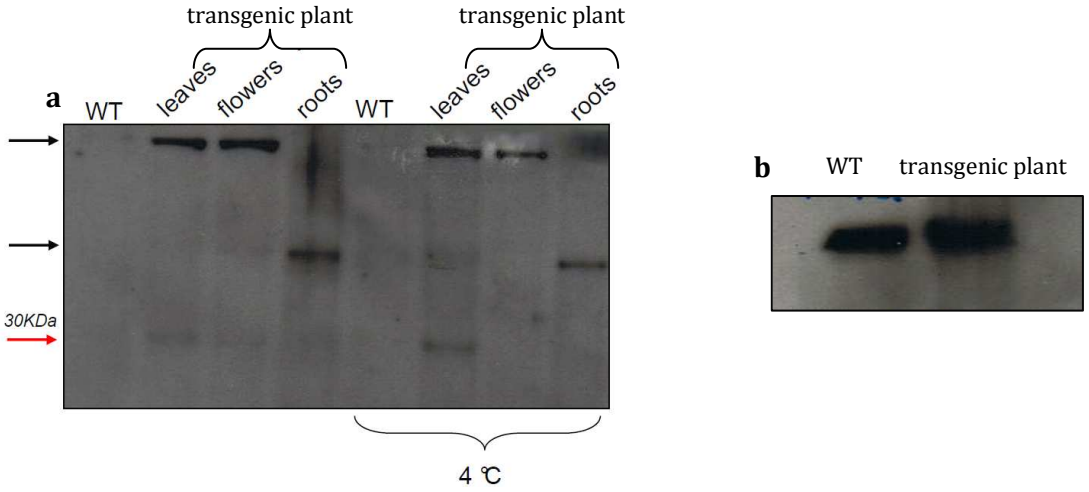


Fig. 3.22: Western blot analysis on protein extracts from WT and leaves, flowers and roots of transgenic plants under standard condition and after cold exposure (**a**) and on nuclei-rich extracts from leaves of WT and transgenic plants (**b**). The arrows indicates the position of aspecific signals.

The second strategy concerned the preparation of a fusion protein between the *MYB59* full-length genomic sequence and the HaloTag® sequence, under the control of CaMV-35S promoter. The HaloTag® is a 33kDa monomeric protein, isolated from *Rhodococcus rhodochrous*, and thus, not endogenous to mammalians, plants or *E. coli* cells (Urh *et al*., 2008). Therefore, the use of a fusion protein with the HaloTag® should avoid background signal aspecificity. This strategy was chosen because, using this single genetic construct, several applications can be carried out, including specific labelling, imaging and immobilization of protein *in vivo* and *in vitro* (Urh *et al*., 2008), also complexed with DNA. The resulting construct was used to transform *myb59* mutant plants. Unfortunately, also in this case, Western blot analysis on protein extracts from WT and transgenic plants had a negative result, since only an aspecific signal was detected in both samples after a long exposure (30 min)(**Fig. 3.23**).

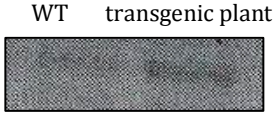


Fig. 3.23: Western blot analysis on protein extracts from WT and transgenic plants.

Since both strategy considered were not effective, the antibody anti-MYB59 is under preparation.

3.10 The homolog *AtMYB48* gene

3.10.1 Comparison between *MYB59* and *MYB48* genes

The *MYB48* gene is localized on the *A. thaliana* chromosome 3 (At3g46130) and encodes for a MYB TF. It has been showed that *AtMYB48* is phylogenetically closed to *AtMYB59* (Romero *et al*., 1998). They share 74.2% nucleotide sequence identity in their coding region and probably they are the result of a relatively recent duplication event (Li *et al*., 2006). In fact, it has been demonstrated that also *MYB48* gene undergoes an alternative splicing event that results in the creation of four variants. Moreover, this conserved alternative splicing pattern was found in two genes homologous to *MYB59* and *MYB48* in rice (Li *et al*., 2006). In **Fig. 3.24** the structures of these splicing variants are depicted in comparison with those of *MYB59*.

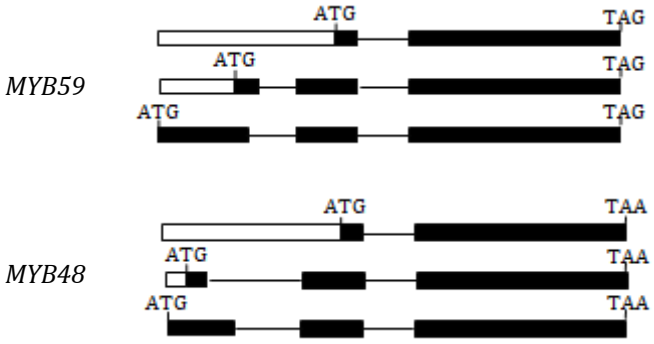


Fig. 3.24: Schematic representation of the two genes structures. Exons are shown as boxes and introns as lines. The putative open reading frames (ORFs) are indicated by black boxes. The start codon (ATG) and the stop codon (TAG and TAA) of each splicing variant in the two genes are also indicated.

It has been demonstrated that also the different splicing variants of *MYB48*, like those of *MYB59*, were expressed in different organs (Li *et al*., 2006). It has also been proposed that *MYB59* and *MYB48* are involved in the transcriptional regulation of secondary xylem formation, since they were up-regulated in *Arabidopsis* xylem. Furthermore, two MYB binding *cis*-element were found in the promoter region of *MYB48*, suggesting that the expression of this gene could be controlled by other MYB TFs (Oh *et al*., 2003). In **Fig. 3.25**, the aminoacid sequences of the two representative proteins, MYB59.3 and MYB48.3, were aligned to show the identical and conserved aminoacidic residues in the two MYB proteins (**Fig. 3.25**). It has been demonstrated that MYB48 proteins share similar localization patterns with those of MYB59, that is, the second and the third protein were localized in the nucleus, whereas the first protein was found only partially in the nucleus (Li *et al*., 2006). This is consistent with their role as TFs in mediating signalling transduction pathways. The R2R3-type MYB factors were grouped into 22 subgroups on the basis of the conserved aminoacid sequence motifs in the MYB domain C-terminal (Kranz *et al*., 1998; Stracke *et al*., 2001). It has been reported that *AtMYB59* and *AtMYB48* could not be classified in any subgroup, due to the fact that their sequence motif do not be conserved (Kranz

et al., 1998; Stracke *et al*., 2001). As shown in **Fig. 3.25**, the two repeats (R2 and R3) are identical in MYB59 and MYB48, except for one aminoacid in the R2 domain repeat (Valine in MYB59.3 and Isoleucine in MYB48.3), which remain a non polar residue.

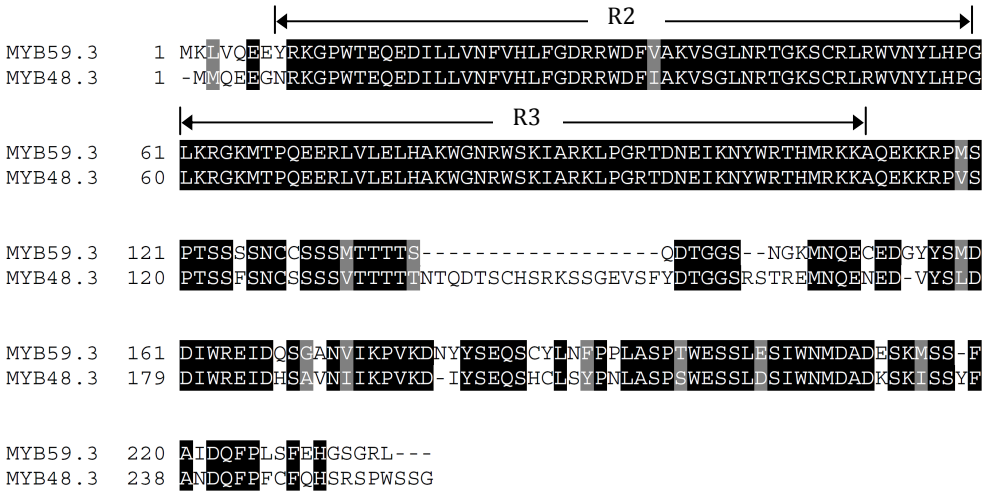


Fig. 3.25: Sequence alignment of the proteins derived from the third splicing variant of the two genes. BOXSHADE 3.21 program (http://www.ch.embnet.org/software/BOX_form.html) was used. Identical aminoacids are showed in black, while conserved are in grey. R2 and R3 domain repeats are indicated.

3.10.2 Analysis of T-DNA insertion mutant for *MYB48*

T-DNA insertion mutant for *MYB48* was identify by searching on the T-DNA Express database available on website http://signal.salk.edu/cgi-bin/tdnaexpress. The mutant line GT_5_9575 in Landsberg (Ler) background was analyzed. This line presents the insertion of the Dissociator (Ds) transposable element. Ds insertion site was confirmed using a combination of gene and Ds-specific primers (**Fig. 3.26 a**). To determine if the insertion was able to knock-out the gene, a RT-PCR reaction on cDNA from homozygous mutant plants was performed using gene-specific primers (**Fig. 3.26 b**).

Fig. 3.26: Schematic representation of the position of Ds element in the *MYB48* gene and its effect on transcript accumulation. **a**: Exons (boxes), as well as intron sequences (lines) are depicted. The Ds insertion sites and orientation are provided and drawn not in scale. **b**: Transcript levels were detected by PCR that was carried out using gene- and *β-actin* specific primer (the latter used as control).

Mutant plants did not show any phenotypic differences respect to WT (Ler) plants when grown on soil under standard conditions (**Fig. 3.27**).

Fig. 3.27: Comparison between WT and *myb48* mutant plants grown on soil under standard conditions for two weeks.

Since *myb48* plants had a WT phenotype, *myb59* and *myb48* plants were crossed, to obtain the double mutant, that may give an indication of the role of the two TFs in plant development. Since the cross was between two different *Arabidopsis* ecotypes, which showed some phenotypic differences, during the selection of double mutant, plants that were mutant for both loci (genes) and plants WT for both genes were selected. The latter were chosen as control WT. To find out possible phenotypic differences, WT and double mutant plants were grown vertically *in vitro* for three weeks on MS medium with and without sucrose. As shown in **Fig. 3.28 a)** and **b)**, in both conditions no differences in root growth were observed, whereas on MS medium without sucrose double mutant plants were smaller than WT.

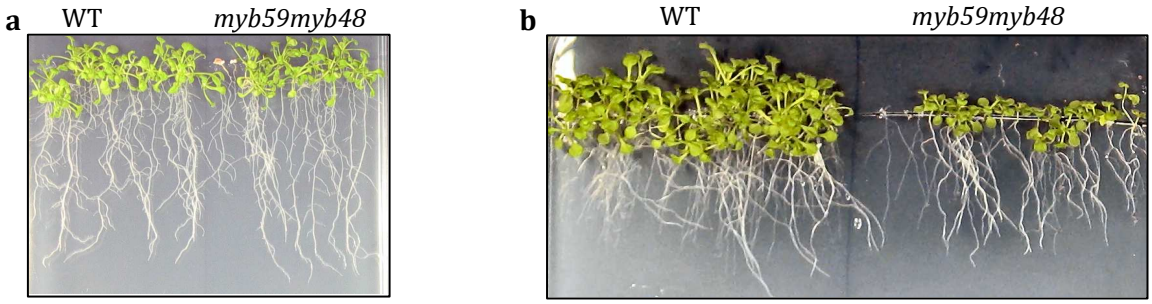


Fig. 3.28: Comparison between WT and *myb59myb48* double mutant plants grown on MS medium with (**a**) and without (**b**) sucrose for three weeks.

Chapter 4

DISCUSSION

Due to their sessile nature, plants are affected by various environmental stresses including extremes of temperature, water deficit, high salinity and presence of heavy metals. Tolerance or susceptibility to abiotic stresses is a very complex phenomenon depending on plant as well as on stress features (Chinnusamy *et al*., 2003). In fact, plant response depends on species, developmental stage and organ or tissue affected, but also on stress duration and intensity (Buchanan *et al*., 2003). Since general and specific tolerance mechanisms exist, it is logical to think that plants have multiple stress perception and signal transduction pathways, which may cross-talk (Priest *et al*., 2009). TFs are proteins acting downstream the signalling cascades and are responsible for the regulation of gene expression (Zhang, 2003). So, characterizing TFs putatively involved in stress response, may help to gain insight into the complex mechanisms that plants activate in response to external stimuli. Between TF families that are involved in stress response, MYB superfamily is the largest in plants (Stracke *et al*., 2001), with 198 genes identified (Yanhui *et al*., 2006). Member of this family are characterized by a MYB domain of about 52 aminoacids (Jin and Martin, 1999; Stracke *et al*., 2001) and are classified into three groups depending on the number of adjacent repeats in the binding domain (Stracke *et al*., 2001): R1R2R3-MYB, R2R3-MYB and the MYB1-related proteins. R2R3-MYB proteins represent the largest MYB group in plants (Stracke *et al*., 2001). In higher plants, *MYB* genes perform different functions, including the control of secondary metabolism, cell shape and organ development, and signal transduction pathways responding to different stimuli (Martin and Paz-Ares, 1997).

The *MYB59* gene has been classified as R2R3-MYB TF (Kranz *et al*., 1998; Romero *et al*., 1998; Stracke *et al*., 2001). Recently, it has been demonstrated that this gene undergoes a conserved alternative splicing that results in the formation of four transcripts, that differ only in their MYB domain. In fact, *MYB59.1* and *MYB59.2* transcripts encode MYB-related proteins with a single MYB domain repeat, whereas *MYB59.3* transcript, chosen as model sequence, encodes a typical R2R3-MYB protein. The fourth transcript has a very low abundance and unlikely encodes a protein (Li *et al*., 2006). These different MYB proteins, encoded by *MYB59* gene, were shown to be nuclear (MYB59.2 and MYB59.3) or only partially nuclear (MYB59.1) and may have binding

affinities to different targets (Li *et al*., 2006).

Alternative splicing is a very complex mechanism that produces multiple forms of mRNA from a unique transcript and seems to be a regulatory process, that contribute to the increase of biological complexity. Typically, AS determines the differential inclusion of coding and non-coding sequence in a transcript. The resulting variants may differ in their stability or display specific chemical and biological activity (Lareau *et al*., 2004). It has been estimated that in plant between 20-60% of genes are alternatively spliced (Ner-Gaon *et al*., 2007) and in many cases alternative isoforms differ by only small alterations of functional elements or domains (Lareau *et al*., 2004). Between different types of AS observed (**Fig. 1.2**), two of them could be involved in the formation of *MYB59* splicing forms: choosing *MYB59.3* as gene model, *MYB59.1* transcript could be formed by the intron retention mechanism, whereas the alternative 5' splice site mechanism could be responsible for *MYB59.2* formation. These events can be observed by the alignment between the three *MYB59* transcripts (**Fig. 3.2** and **3.3**), where it is clear that in *MYB59.1* the first intron is not removed, and the 5' splice site of the first intron of *MYB59.2* is different from that in *MYB59.3*. It has been reported that intron retention is the most common type of AS in plants (Ner-Gaon *et al*., 2004; Barbazuk *et al*., 2008) and it has been proposed that the position of retained introns may have relevance on the transcript stability (Ner-Gaon *et al*., 2004). In the case of *MYB59*, the alternatively spliced intron can be considered as part of 5'UTR in *MYB59.1* transcript, that is the most abundant form (Li *et al*., 2006), and as part of the coding region in *MYB59.2* and *MYB59.3* transcripts.

Moreover, for the formation of the three MYB59 proteins, three different ATG starting codons must be recognized by the ribosomes. According to the first-AUG rule, the first AUG codon of a mRNA is the exclusive site of the translation initiation even if a second AUG is positioned just a few bases downstream (Kozak, 1999). However, two other mechanisms, context-dependent leaky scanning and reinitiation, can bypass this rule. A context-dependent leaky scanning mechanism enables the production of different proteins from a single mRNA. In fact, it has been demonstrated that sometimes ribosomes can bypass the first AUG codon, depending on the sequence context around it, and initiate at the second or, rarely, the third AUG. Reinitiation mechanism involves the translation of upstream ORFs (uORF) before the downstream (Kozak, 1999; Kozak, 2002). The starting AUG of *MYB59.1* and *MYB59.2* transcripts are located downstream the first AUG, and thus, different translation mechanisms probably act to translate the three proteins. In particular, it has been proposed that when the first AUG resides in a weak context (e.g. lacking conserved residues in defined positions) the scanning continues and initiates further downstream (Kozak, 2002). *MYB59.3* starting codon is not in a conserved context and therefore, it can be hypothesized that a context-dependent leaky scanning

mechanism may be adopted for *MYB59.2* translation. The reinitiation mechanisms could be responsible for the *MYB59.1* translation, since a putative uORF is present upstream its AUG starting codon (Li *et al*., 2006). However, examples of mRNA initiating from three sites, by leaky scanning mechanism have been proposed (Kozak, 2002).

In this work it has been shown that the three alternatively spliced variants of *MYB59* are expressed in different organ-specific patterns (**Fig. 3.8**), and this result may confirm the hypothesis that different functions may be played by the three different forms. Respect to the expression in rosetta leaves, that was chosen as reference sample, *MYB59.1* was expressed mainly in roots (**Fig. 3.8 a**), the expression of *MYB59.2* was low in all the samples (**Fig. 3.8 b**), whereas *MYB59.3* was expressed at a little higher levels in open flowers and roots (**Fig. 3.8 c**). Different localization of gene splicing variants has also been shown in other plant species. For example, *rbohB* (*respiratory burst oxidase homolog*) gene in maize underwent AS by intron retention event, and the two variants were differentially expressed in various tissues and at different developmental stages (Lin *et al*., 2009).

Our results indicated that *MYB59* gene participates in Cd stress response. In fact, like its homolog *BjCdR12* in *B. juncea* (Fusco *et al*., 2005), is involved in the early signal transduction cascade due to Cd exposure, suggesting that this function is conserved in the two plant species. Its transcription level changed in *Arabidopsis* WT plants treated with $Cd(NO_3)_2$ and, in particular, increased after a 2-hour-treatment (**Fig. 3.5**). Moreover, by the quantification of Cd content measured in leaves and roots of WT, overexpressing and RNAi plants, treated with $CdSO_4$, it was emerged that *MYB59* could be involved in the metal root-to-shoot transport. In fact, as shown in **Fig. 3.14**, while in leaves of overexpressing plants Cd content was lower than in WT, in roots the effect of *MYB59* overexpression was opposite. No differences in Cd accumulation between lines that overexpress the three splicing variants were observed. This may mean that the three variants have the same role in Cd stress response. In RNAi plants, where *MYB59* expression is knocked-down (98% expression reduction, **Fig. 3.9 d**), Cd accumulated at higher levels in leaves, while, in roots, the value measured was very low in respect to WT (**Fig. 3.14**). These results suggest that *MYB59* is probably not involved in Cd uptake, but its expression may impair the metal transport from roots to shoots, maybe by regulating negatively heavy metal transporters or by inducing compartmentalization mechanisms in roots. In other words, when plants lack *MYB59* expression, Cd can be conveyed in leaves, whereas in overexpressing plants the repression of the root-to-shoot transport causes the accumulation of Cd in roots. It should be interesting to verify how the transcription level of genes encoding metal transporters changes in overexpressing and RNAi plants, to better comprehend the role of *MYB59* in heavy metal transport. Recently, it has been reported that a bZIP TF isolated from *B. juncea* treated with Cd,

is involved in the long-distance transport of Cd from root to shoot in *Arabidopis* (Farinati *et al*., 2010).

Our data indicate that *MYB59* gene plays a role also in other abiotic stress response. As shown in **Fig. 3.7**, the transcription level of the three splicing variants of *MYB59* changed differently under stress conditions, suggesting a role of the three splicing forms in multiple signalling pathways, activated in response to different stresses. In particular, *MYB59.2* expression was induced after cold, drought and ABA treatment, whereas *MYB59.3* was highly expressed after drought stress. It has been reported that there is a connection between AS and environmental conditions (Mastrangelo *et al*., 2005), even if the mechanisms by which stresses influence plant splicing are still unknown (Ner-Gaon *et al*., 2004). An example is represented by the *Bronze2* (*Br2*) locus in maize, that encodes a glutathione-S-transferase induced by Cd and other stresses. It was observed that Cd exposure affected the splicing pattern, leading to an increase of the unspliced transcript, whereas other stress conditions did not influence the splicing event (Marrs and Walbot, 1997). Moreover, although several studies have reported that the splicing efficiency decreased after a high temperature treatment, because of a less stable interaction between mRNA and splicing factors, it has recently been demonstrated that cold stress induces intron retention in two *e-cor* genes in durum wheat, but in the *Arabidopsis* orthologous genes this effect was not observed or not associated with stress response (Mastrangelo *et al*., 2005). Moreover, the expression of the major splicing form of the maize *rbohB* was activated under several stress stimuli, such as heat and salt treatment (Lin *et al*., 2009). The rice *DREB2B* gene presents two forms that are produced by AS. It has been demonstrated that the ratio between the two variants changed under stress conditions. The *DREB2B2* form, that lack an exon, was strongly induced by low temperature exposure (Matsukura *et al*., 2010). Our results show that the expression of *MYB59.1* transcript, that have a retained intron, was not affected by the stress treatments, whereas the expression of *MYB59.2* and *MYB59.3* transcripts were modulated by different stress conditions. We did not observe the cold-dependent intron retention, but an alteration of spliced transcripts following stress exposure. This fact could suggest that some stresses, such as cold, dehydration and ABA, can affect the spliceosome complex, that, in turn, regulates the abundance of the different transcripts. Stress conditions alter splicing mechanism leading to an increase of spliced transcripts, suggesting that these two forms of *MYB59* are important for the activation of plant stress response. This feature of *MYB59* may be used in a biotechnology approach to improve plant tolerance to abiotic stresses. In fact, one of the strategies that can be adopted to engineer stress tolerance in plant is represented by the use of transcriptional activators and repressors of stress-responsive genes (Hazen *et al*., 2003). It has been demonstrated that the overexpression of some drought-responsive TFs increased drought tolerance (Umezawa *et al*.,

2006). For example, an increase in drought and salt tolerance has been observed in plants overexpressing *MYB15*, that is induced by cold and salt stresses (Ding *et al*., 2009). Recently, it has been reported that the overexpression of *AtMYB44*, which expression was activated under various abiotic stress conditions, is able to confer salt and dehydration tolerance in transgenic plants, through a reduced expression of *PP2Cs* genes, which have been described as negative regulators of ABA signalling (Kuhn *et al*., 2006; Jung *et al*., 2008). It has also been reported that a null mutation in *AtMYB60*, which is involved in the regulation of stomata movement and is negatively regulated under drought conditions, can minimize wilting in water deficit situations (Cominelli *et al*., 2005), while the overexpression of *AtMYB61* has been proposed as a drought-resistance strategy for plant growth in arid region (Liang *et al*., 2005). It has also demonstrated that *OsMYB4* overexpressing plants showed a significant increase in cold and freezing tolerance in *Arabidopsis* (Vannini *et al*., 2004). In the case of *MYB59* the potential of the three variants could be evaluated.

To characterized this gene, a knock-out mutant line for *MYB59* was also considered. The involvement of *MYB59* in the abiotic stress response, was confirmed by analyzing the expression patterns of two genes known to be induced by abiotic stress: *Rd29a* and *PP2CA*. The transcription levels of these two genes were measured in WT and mutant plants after drought and ABA treatment. We observed that the transcription level of *Rd29a* and *PP2CA* was higher in mutant plants than in WT in water deficit conditions and after ABA treatment respectively (**Fig. 3.11**). This result confirms that *MYB59* participates to the abiotic stress response and give an indication in which signalling pathways this TF is involved, but its effective role remains unclear. As shown in **Fig. 3.7**, the expression of *MYB59.2* and *MYB59.3* variants was induced after cold, drought and ABA stress. Considering that the transcription level of the *Rd29a* and *PP2CA* genes increased in mutant plants (**Fig. 3.11**), it could be hypothesized that *MYB59* may act in the control of the expression of negative regulators, that participate in signalling pathways responding to stress conditions. As *AtMYB44* knock-out mutant that exhibited an increase of the salt-induced expression of PP2C-encoding genes (Jung *et al*., 2008), *MYB59* could play a similar role. However, further analysis are needed to clarify how this gene and, in particular, the splicing variants control plant response to stress.

Moreover, we found that *MYB59* gene may play a role in vegetative growth. In fact, plants showing ectopic expression of *MYB59.1* had bigger rosetta leaves than control plants, whereas mutant line had a reduced leaf area (**Fig. 3.12**). It has been demonstrated that the *AtMYB59* gene is expressed during the S and S to G2 phases in *Arabidopsis* suspension and interferes with the cell cycle progression in yeast and plant cells (Cominelli and Tonelli, 2009; Mu *et al*., 2009). In particular, it was able to suppress cell proliferation in yeast cells: *MYB59* overexpressing cells

were longer than the control. However, in plant, the cell size and structure were not changed in overexpressing plants in respect to WT. Whether the leaf phenotype is due to a higher cell dimension or a higher cell number must be investigated with a more detailed microscopic analysis. Moreover, Mu *et al*. (2009) showed that differences in root length exist between WT, overexpressing and mutant plants, even if they did not observe relevant differences in root cell dimension in the plants analyzed. We did not observe any important differences in root length comparing these plants (**Fig. 3.13**). It is not surprising that an alteration of a TF expression leads to an evident phenotype, since they are the main responsible for the modulation of gene expression and often involved in developmental processes. It has been demonstrated, for example, that the overexpression of *AtMYB41*, another R2R3-MYB protein, caused a dwarf phenotype due to the alterations of cell expansion (Cominelli *et al*., 2008; Lippold *et al*., 2009). Also the ectopic expression of *AtMYB44* caused development changes, such as growth and flowering time retardation (Jung *et al*., 2008).

To investigate the expression localization of the three splicing variants, we characterized the promoter region, considering the sequences upstream each of the three ATG starting codons and preparing a fusion construct with the GUS reporter gene (**Fig. 3.16**). Data deriving from GUS assay revealed that the three alternative spliced forms localized in different plant organs. The analysis was conducted on seedling, rosetta leaves and flowers. In particular, *MYB59.1* expression was found mainly in the leaf veins and hydathods both apical and lateral (**Fig. 3.17 a**). The expression of *MYB59.2* was found only in anthers of closed flowers (**Fig. 3.17 b**), whereas *MYB59.3* expression was detected in all vegetative tissues and sepals (**Fig. 3.17 c**). Probably, since the three alternatively spliced variants have different localizations, they may also play different roles in plant development. This result could be confirmed by the fact that the three MYB59 proteins differ in the number of repeats in the MYB binding domain.

The methylation mechanism is used to protect the genome against transposons and to modulate the expression of endogenous genes (Goll and Bestor, 2005). The analysis of the methylation pattern of a specific sequence may give an indication about its expression. Since repeated regions are often subjected to methylation on C residues (Law and Jacobsen, 2009), we analyzed a direct repeat found in the promoter region of *MYB59*, close to the TATA box element. This repeated sequence, of about 50 bp, could be an important point of gene regulation. We also analyzed the region including the first intron sequence, since *MYB59* gene has three different splicing variants that are expressed in an-organ specific manner and, thus, intron sequence may contain regulatory regions important for the gene expression. In fact, it has been reported that the methylation in a putative *cis*-element in pMADS3 intron 2 is involved in the transcriptional activation of the gene in *Petunia* (Shibuya *et al*., 2009). Moreover, it has been showed that its

homologous gene in *Arabidopsis*, *AG* (*AGAMOUS*), contains an intron with a regulatory sequence that is crucial for its expression in stamen and carpel (Busch *et al*., 1999; Deyholos and Sieburth, 2000). Regarding the first sequence considered, we found that the majority of C residues in anthers were converted in T by the bisulfite reagent, whereas in leaves not (**Fig. 3.19**). Thus, given that sodium bisulfite is able to modify unmethylated C residues, C nucleotides in the direct repeat in anther sequence appear not to be methylated. In the second sequence considered, we found that all of the C residues were modified, but no differences in the methylation status between leaves and anthers were observed (**Fig. 3.20**). It is known that there is an interconnection between methylation and transcription (Zilberman, 2008). It has been reported that DNA methylation of promoter regions usually leads to an inhibition of transcriptional initiation, whereas the effects of methylation within coding regions often do not influence gene expression (Chan *et al*., 2005). Genes methylated in the promoter region are more likely to be expressed at low levels (Zhang *et al*., 2006; Gehring and Henikoff, 2007). Probably the best study is represented by the *FWA*, a gene encoding a homeodomain protein expressed only in the endosperm. It contains two tandem direct repeats within its promoter that are highly methylated at CG residues. The removal of methylation is required for the transcriptional activation of *FWA* in the endosperm (Kinoshita *et al*., 2004; Gehring and Henikoff, 2007). Since the methylation in the two sequences considered for *AtMYB59* is asymmetrical (non-CG type), it will also be analyzed in mutant plants lacking the expression of the enzymes responsible for this kind of methylation, and, in particular, in the *cmt3* (N16392), *cmt3drm1drm2* (N16384; Henderson and Jacobsen, 2008) and *drm1drm2* (N6366; Cao and Jacobsen, 2002) mutants. It has been shown that the *cmt3* plants show a reduction of asymmetric methylation at some *loci* (Goll and Bestor, 2005), whereas *DRM* genes are important for the initial establishment of methylation in all contexts and for *de novo* methylation (Cao and Jacobsen, 2002). It has also been demonstrated that DRMs act redundantly with CMT3: in the *cmt3drm1drm2* triple mutants all asymmetric methylation is lost (Cao and Jacobsen, 2002). The analysis of the methylation profile of the two regions considered in mutant plants could be important to understand if these sequences are involved in the control of *MYB59* expression. In fact, repeated DNA sequences constitute a large proportion of eukaryotic genomes and seem to be involved in the regulation of heterochromatin formation, gene expression and epigenetic processes (Ugarkovic, 2005). Generally formed by duplication events, these sequences were classified in two categories: dispersed repeats, such as retrotrasposons, and tandem repeats, such as simple sequence repeats (Richard *et al*., 2008). The latters are defined by the length of the repeated unit (n): microsatellite clusters are characterized by a repeat unit usually between 1 and 13 bp, while minisatellites repeat units vary from 14 to 500 bp (Vergnaud and Denoeud, 2000). In the

promoter region of *MYB59* we found a dinucleotide (GA) microsatellite. How this sequence could affect gene expression must be evaluated, since the function of microsatellites is still unknown. However, it has been recently reported that a 23-bp repeat motif in the promoter region of *MYB10* in red apple caused the TF autoregulation and thus, accumulation of anthocyanins (Espley *et al*., 2009). It has also been demonstrated that the two tandem repeats in *FWA* promoter are able to attract the siRNA-making complex, that, in turn, stimulates methylation by RNA-mediated DNA methylation process (RdDM) (Chan *et al*., 2006; Robinson, 2006) and thus, silencing of gene containing the repeat.

Unfortunately we did not obtain any indications of the proteins encoded by the three splicing forms of *MYB59*, since the two strategies adopted did not lead to clear results. However, we are going to prepare a MYB59-specific antibody that could give an indication on whether and which of the three splicing variants is actually translated in plant.

The R2R3-type MYB factors were grouped into 22 subgroups on the basis of the conserved aminoacid sequence motifs in the MYB domain C-terminal, but *MYB59* gene, together with *MYB48*, does not belong to any subgroup (Kranz *et al*., 1998; Stracke *et al*., 2001). These two genes are phylogenetically related (Romero *et al*., 1998) and the fact that both genes undergo a similar AS event may confirm that they are homolog (Li *et al*., 2006) (**Fig. 3.24**). Moreover, the aminoacid sequences of the two representative proteins, MYB59.3 and MYB48.3, were identical except for one aminoacid (Valine in MYB59.3 and Isoleucine in MYB48.3), that maintains an hydrophobic side chain (**Fig. 3.25**). It has been demonstrated that also the different splicing variants of *MYB48* were expressed in different organs and that MYB48 proteins share similar nuclear localization patterns (Li *et al*., 2006). The fact that both genes have been found to be involved in abiotic stress response (Li *et al*., 2006) and plant secondary growth (Oh *et al*., 2003) may confirm their role in the environmental and developmental transcriptional control. We analyzed a Ds insertional knock-out mutant for *MYB48* in the *Arabidopsis* Landsberg (Ler) background. Since mutant plants did not show any phenotypic differences respect to WT (Ler) plants when grown on soil under standard conditions (**Fig. 3.27**), we crossed *myb59* and *myb48* single mutants. The resulting double mutant may give an indication about the role of the two genes in plant development and may help to clarify if and which functions, played by the two TFs, are redundant. In fact, as shown in **Fig. 3.28**, the double mutant was smaller than WT when plants were grown *in vitro* on MS medium without sucrose, even if no differences in roots growth were observed. Since *myb59* plants were smaller than WT (**Fig. 3.12**), *myb48* plants showed a WT phenotype (**Fig. 3.27**) and double mutant plants had a reduced growth in respect to WT (**Fig. 3.28**), we could think that *MYB59* is able to complement *MYB48* gene in the control of vegetative growth. An example of function redundancy is represented by *MYB5* and *MYB23* in

regulating trichome and seed coat development. The *MYB5* gene is expressed in trichomes and seeds and the lack of its expression lead minimal changes in trichome morphology. The *MYB23* gene is expressed in developing trichomes and regulates trichome branching, and *myb23* mutant produced increased numbers of small trichomes and two-branched trichomes. A *myb5myb23* double mutant developed more small rosette trichomes and two-branched trichomes than the single mutants (Li *et al*., 2009). However, further analysis will be carried out on *myb59myb48* plants to better understand the role of the two homologous genes in plant development and signalling transduction pathways activated in response to abiotic stress. In fact, the results obtained are preliminary and must be confirmed, comparing WT, single and double mutant plants in the same growth conditions.

Chapter 5

CONCLUDING REMARKS

The characterization of a TF may be very important to understand the mechanisms that underlie plant responses to environmental and developmental stimuli, since TFs are the proteins responsible for the regulation of gene expression. In this PhD work, the attention has been addressed to the characterization of the *MYB59* gene in *A. thaliana*, that presents three splicing variants produced by an alternative splicing event (Li *et al*., 2006). In particular, we showed that the three alternatively spliced variants of *MYB59* are expressed according to different organ-specific patterns, and probably play different functions in plant.

This gene has been considered since the expression of its homolog *BjCdR12* in *B. juncea* was induced by Cd (Fusco *et al*., 2005). We confirmed an involvement of *MYB59* in Cd stress and, in particular, our results indicate that this gene participates in the early stress response, since its expression increased after a 2 h Cd exposure. Its implication in the heavy metal stress response and, a possible role in the metal root-to-shoot transport has also been confirmed by the comparison between WT, overexpressing and RNAi plants regarding Cd content in leaves and roots.

MYB59 factor plays also a role in other abiotic stress response. We found that even though *MYB59.1* transcription was not influenced by stress treatments, the expression of *MYB59.2* increased after ABA, cold and drought exposure, and *MYB59.3* after drought treatment. This result confirms that the three variants play different roles in plant.

We demonstrated that *MYB59* may be involved in the control of vegetative growth. In fact, the comparison between WT, overexpressing and mutant plants showed that plants expressing *MYB59.1* ectopically had a leaf area higher than control plants. Conversely, mutant plants showed to be smaller. The overexpression of *MYB59.2* and *MYB59.3* also induced a decrease of leaf area in respect to WT plants.

The analysis of the promoter region was carried out considering three regions of about 2.0 Kbp upstream ATGs for each splicing form. GUS assay allowed to localized the expression of the three variants in different plant organs and tissues. The expression of *MYB59.1* was found mainly in leaf veins; the expression of *MYB59.2* was mainly detected in immature anthers, whereas the

expression of *MYB59.3* was localized in most vegetative tissues, sepals, but not in anthers. This is another confirmation that the three splicing variants, having different localizations, may also play different roles in plant.

The analysis of the methylation pattern of a direct repeat close to the TATA box and the region containing the entire sequence of the first intron underlines that differences in the direct repeat methylation profile exist between leaves and anthers, although how this characteristic could influence gene expression is still unknown.

It has not been possible to study the MYB59 proteins since the two strategy applied did not lead to any clear results. The antibody against MYB59 could give us information about which protein is actually synthesised in plant.

MYB59 and *MYB48* are homologous genes and undergo a similar alternative splicing mechanism. To understand whether the two genes are functionally redundant, we crossed *myb59* and *myb48* plants and observed the double mutant phenotype. It was found that, when grown on MS medium without sucrose, *myb59myb48* plants were significantly smaller than WT. This result may suggest that these two genes have an overlapping role in the control of the rosetta dimension. Further analysis are needed for the comprehension of the role of the two genes in plant development and response to abiotic stress.

Chapter 6

REFERENCES

- **Abe H., Urao T., Ito T., Seki M., Shinozaki K. and Yamaguchi-Shinozaki, K.** (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcription activators in abscisic acid signalling. *Plant Cell*, **15**: 63-78.

- **Alberts B., Johnson A., Lewis J., Raff M., Keith R. and Peter W.** (2002) Molecular biology of the cell. *New York and London: Garland Science*.

- **Alves-Ferreira M., Wellmer F., Banhara A., Kumar V., Riechmann J. L. and Meyerowitz E. M.** (2007) Global expression profiling applied to the analysis of *Arabidopsis* stamen development. *Plant Physiology*, **145**: 747–762.

- **Bailey-Serres J.** (1999) Selective translation of cytoplasmic mRNAs in plants. *Trends in Plant Science*, **4**: 142-148.

- **Barbazuk W. B., Fu Y. and McGinnis K. M.** (2008) Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Research*, **18**: 1381-1392.

- **Bartels D. and Sunkar R.** (2005) Drought and salt tolerance in plants. *Critical Reviews in Plant Sciences*, **24**: 23–58.

- **Benson G.** (1999) Tandem repeats finder: a program to analyze DNA sequenze. *Nucleic Acids Research*, **27**: 573-580.

- **Black D. L.** (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, **72**: 291-336.

- **Boyko A. and Kovalchuk I.** (2008) Epigenetic control of plant stress response. *Environmental and molecular mutagenesis*. **49**: 61-72.

- **Brown T. A.** (2002) Genomes. *Second Edition, BIOS Scientific Publishers Ltd.*

- **Buchanan B. B., Gruissem W. and Jones R. L.** (2003) Biochemistry and molecular biology of plants*. American Society of Plant Physiologists*.

- **Busch M. A., Bomblies K. and Weigel D.** (1999) Activation of a floral homeotic gene in *Arabidopsis*. *Science*, **285**: 585-587.

- **Cao X. and Jacobsen S. E.** (2002) Role of the *Arabidopsis DRM* methyltransferases in *de novo* DNA methylation and gene silencing. *Current Biology*, **12**: 1138–1144.

- **Chakrabortee S., Boschetti C., Walton L. J., Sarkar S., Rubinsztein D. C. and Tunnacliffe A.** (2007) Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proceeding of National Academy of Science of the United States of America*, **104**: 18073-18078.

- **Chan S. W.-L., Henderson I. R. and Jacobsen S. E.** (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Reviews. Genetics*, **6**: 351-360.

- **Chan S. W.-L., Zhang X., Bernatavichute Y. V. and Jacobsen S. E.** (2006) Two-step recruitment of RNA-directed DNA methylation to tandem repeats. *PLoS Biology*, **4**: 1923-1933.

- **Chaney R. L., Malik M., Li Y. M., Browson S. L. Brewer E. P., Angle J. S. and Baker J. M. A.** (1997) Phytoremediation of soil metal. *Current Opinion in Biotechnology*, **8**: 279-284.

- **Cheng H., Song S., Xiao L., Soo H. M., Cheng Z., Xie D. and Peng J.** (2009) Gibberellin acts through jasmonate to control the expression of MYB21, MYB24, and MYB57 to promote stamen filament growth in *Arabidopsis*. *Plos Genetics*, **5**: 1-13.

- **Chinnusamy V. and Zhu J-K.** (2009) Epigenetic regulation of stress responses in plants. *Current Opinion in Plant Biology*, **12**: 133- 139.

- **Chinnusamy V., Shumaker K. and Zhu J. K.** (2003) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *Journal of Experimental Botany*, **55**: 225–236.

- **Chinnusamy V., Zhu J. and Zhu J-K.** (2007) Cold stress regulation of gene expression in plants. *Trends in Plant Science*, **12**: 444- 451.

- **Choi C. S. and Sano H.** (2007) Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. *Molecular Genetics and Genomics*, **277**: 589-600.

- **Clough S. J. and Bent A. F.** (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal,* **16**: 735-743.

- **Cobbett C. S. and Goldsbrough P.** (2002). Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. *Annual Review of Plant Physiology and Plant Molecular Biology*, **53**: *159–182.*

- **Cominelli E. and Tonelli C.** (2009) A new role for plant R2R3-MYB transcription factors in cell cycle regulation. *Cell* Research, **19**: 1231-1232.

- **Cominelli E., Galbiati M., Vavasseur A., Conti L., Sala T., Vuylsteke M., Leonhardt N., Dellaporta S. L., Tonelli C.** (2005) A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Current Biology*, **15**: 1196-2000.

- **Cominelli E., Sala T., Calvi D., Gusmaroli G. and Tonelli C.** (2008) Over-expression of the *Arabidopsis AtMYB41* gene alters cell expansion and leaf surface permeability. *The Plant Journal*, **53**: 53–64.

- **DalCorso G., Farinati S., Maistri S. and Furini A.** (2008) How plants cope with Cadmium: staking all on metabolism and gene expression. *Journal of Integrative Plant Biology*, **50**: 1268–1280.

- **Delseny M., Bies-Etheve N., Carles C., Hull G., Vicient C., Raynal M., Grellet F. and Aspart L.** (2001) Late Embryogenesis Abundant (LEA) protein gene regulation during *Arabidopsis* seed maturation. *Journal of Plant Physiology*, **158**: 419–427.

- **Deyholos M. K. and Sieburth L. E.** (2000) Separable whorl-specific expression and negative regulation by enhancer elements within the *AGAMOUS* second intron. *The Plant Cell*, **12**: 1799-1810.

- **Ding Z., Li S., An X., Liu X., Qin H. and Wang D.** (2009) Transgenic expression of *MYB15* confers enhanced sensitivity to abscisic acid and improved drought tolerance in *Arabidopsis thaliana. Journal of Genetics and Genomics*, **36**: 17-29.

- **Du H., Zhang L., Liu L., Tang X., Yang W., Wu Y., Huang Y. and Tang Y.** (2009) Biochemical and molecular characterization of plant MYB transcription factor family. *Biochemistry (Moscow)*, **74**:1-11.

- **Espley R. V., Brendolise C., Chagnè D., Kutty-Amma S., Green S., Volz R., Putterill J., Schouten H. J., Gardiner S. E., Hellens R. P. and Allan A. C.** (2009) Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. *The Plant Cell*, **21**: 168-183.

- **Farinati S., DalCorso G., Varotto S. and Furini A.** (2010) The *Brassica juncea* BjCdR15, an ortholog of *Arabidopsis* TGA3, is a regulator of cadmium uptake, transport and accumulation in shoots and confers cadmium tolerance in transgenic plants. *New Phytologist*, **185**: 964–978.

- **Feder M. E. and Hofmann G. E.** (1999) Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annual Review of Physiology*, **61**: 243-282.

- **Ferguson I. B.** (2004) The plant response: stress in the daily environment. *Journal of Zhejiang University. Science*, **5**: 129-132.

- **Floris M., Mahgoub H., Lanet E., Robaglia C. and Menand B.** (2009) Post-transcriptional regulation of gene expression in plants during abiotic stress. *International Journal of Molecular Sciences*, **10**: 3168-3185.

- **Fusco N., Micheletto L., Dal Corso G., Borgato L. and Furini A.** (2005) Identification of cadmium-regulated genes by cDNA-AFLP in the heavy metal accumulator *Brassica juncea* L. *Journal of Experimental Botany*, **56**: 3017-3027.

- **Gehring M. and Henikoff S.** (2007) DNA methylation dynamics in plant genomes. *Biochimica et Biophysica Acta*, **1769**: 276-286.

- **Goll M. G. and Bestor T. H.** (2005) Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry*, **74**: 481–514.

- **Goyal A., Walton L. J. and Tunnacliffe A.** (2005) LEA proteins prevent protein aggregation due to water stress. *The Biochemical Journal*, **388**: 151-157.

- **Hazen S. P., Wu Y. and Kreps J. A.** (2003) Gene expression profiling of plant responses to abiotic stress. *Functional & integrative genomics*, **3**: 105-111.

- **Henderson I. A. and Jacobsen S. E.** (2008) Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes & Development*, **22**: 1597-1606.

- **Hoagland D.R. and Arnon D. I.** (1938) The water culture method for growing plants without soil. *University of California College Agriculture Experimental Station Circular. Berkeley, CA*, 347-353.

- **Horiguchi G., Fuse T., Kawakami N. and Kodama H.** (2000) Temperature-dependent translational regulation of the ER omega-3 fatty acid desaturase gene in wheat root tips. *Plant Journal*, **24**: 805-813.

- **Hua X. J., Van de Cotte B., Van Montagu M. and Verbruggen N.** (2001) The 5' untranslated region of the At-P5R gene is involved in both transcriptional and post-transcriptional regulation. *Plant Journal*, **26**: 157-169.

- **Ingram J. and Bartels D.** (1996) The molecular basis of dehydration tolerance in plants. *Annual Review in Plant Physiology and Plant Molecular Biology*, **47**: 377–403.

- **Jefferson R. A., Kavanagh T. A. and Bevan M. W.** (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO Journal*, **6**: 3901-7.

- **Jin H. and Martin C.** (1999) Multifunctionality and diversity within the plant *MYB*-gene family. *Plant Molecular Biology*, **41**: 577-585.

- **Jung C., Seo J. S., Han S. W., Koo Y. J., Kim C. H., Song S. I., Nahm B. H., Choi Y. D. and Cheong J.-J.** (2008) Overexpression of *AtMYB44* enhances stomatal closure to confer abiotic stress tolerance in transgenic *Arabidopsis*. *Plant Physiology*, **146**: 623–635.

- **Kasahara R., Portereiko M. F., Sandaklie-Nikolova L., Rabiger D. S. and Drews G. N.** (2005) *MYB*98 is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. *Plant Cell*, **17**: 2981-2992.

- **Kawasaki S., Borchert C., Deyholos M., Wang H., Brazille S., Kawai K., Galbraith D. and Bohnert H. .J** (2001) Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell*, **13**: 889–905.

- **Kim E., Magen A. and Ast G.** (2007) Different levels of alternative splicing among eukaryotes. *Nucleic Acids Research*, **35**: 125–131.

- **Kinoshita T., Miura A., Choi Y., Kinoshita Y., Cao X., Jacobsen S. E., Fischer R. L. and Kakutani T.** (2004) One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science*, **303**: 521–523.

- **Klempnauer K.-H., Gonda T. J. and Bishop J. M.** (1982) Nucleotide sequence of the retroviral leukemia gene *v-myb* and its cellular progenitor *c-myb:* The architecture of a transduced oncogene. *Cell*, **31**: 453-463.

- **Kozak M.** (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene*, **234**: 187-208.

- **Kozak M.** (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, **299**: 1-34.

- **Kraft E., Stone S. L., Ma L., Su N., Gao Y., Lau O. S., Deng X. W. and Callis J.** (2005) Genome analysis and functional characterization of the E2 and RING-type E3 ligase ubiquitination enzymes of Arabidopsis. *Plant Physiology*, **139**: 1597-1611.

- **Kranz H. D., Denekamp M., Greco R., Jin H., Leyva A., Meissner R. C., Petroni K., Urzainqui A., Bevan M., Martin C., Smeekens S., Tonelli C., Paz-Ares J. and Weisshaar B.** (1998) Towards functional characterization of the members of the *R2R3-MYB* gene family from *Arabidopsis thaliana*. *The Plant Journal*, **16**: 263-276.

- **Kuhn J. M., Boisson-Dernier A., Dizon M. B., Maktabi M. H. and Schroeder J. I.** (2006) The protein phosphatase AtPP2CA negatively regulates ABA signal transduction in Arabidopsis. *Plant Physiology*, **140**: 127-139.

- **Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., *et al.*** (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**: 860–921.

- **Lareau L. F., Green R. E., Bhatnagar R. S. and Brenner S. E.** (2004) The evolving roles of alternative splicing. *Current Opinion in Structural Biology*, **14**: 273–282.

- **Law J. A. and Jacobsen S. E.** (2009) Dynamic DNA methylation. *Science*, **323**:1568-1569.

- **Li J., Li X., Guo L., Lu F., Feng X., He K., Wei L., Chen Z., Qu L.-J. and Gu H.** (2006) A subgroup of *MYB* transcription factor genes undergoes highly conserved alternative splicing in *Arabidopsis* and rice. *Journal of Experimental Botany*, **57**: 1263-1273.

- **Li J., Yang X., Wang Y., Li X., Gao Z., Pei M., Chen Z., Qu L.-J. and Gu H.** (2006) Two groups of MYB transcription factors share a motif which enhances trans-activation activity. *Biochemical and Biophysical Research Communications*, **341**: 1155-1163.

- **Li S. F., Milliken O. N., Pham H., Seyit R., Napoli R., Preston J., Koltunow A. M. and Parisha R. W.** (2009) The *Arabidopsis* MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *The Plant Cell*, **21**: 72–89.

- **Liang Y., Dubos C., Dodd I. C., Holroyd G. H., Hetherington A. M. and Campbell M. M.** (2005) *At*MYB61, an R2R3-MYB transcription factor controlling stomatal aperture in *Arabidopsis thaliana. Current Biology*, **15**: 1201-1206.

- **Lin F., Zhang Y. and Jiang M.-Y.** (2009) Alternative splicing and differential expression of two transcripts of nicotine adenine dinucleotide phosphate oxidase B gene from *Zea mays*. *Journal of Integrative Plant Biology*, **51**: 287–298.

- **Lindroth A. M., Cao X., Jackson J. P., Zilberman D., McCallum C. M., Herikoff S. and Jacobsen S. E.** (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science*, **292**:2077–80.

- **Lippold F., Sanchez D. H., Musialak M., Schlereth A., Scheible W.-R., Hincha D. K. and Udvardi M. K.** (2009) AtMYB41 regulates transcriptional and metabolic responses to osmotic stress in *Arabidopsis*. *Plant Physiology*, **149**: 1761-1772.

- **Lipsick J. S.** (1996) One billion years of MYB. *Oncogene* **13**: 223-235.

- **Liu L., White M. J. and MacRae T.** (1999) Transcription factors and their genes in higher plants, functional domains, evolution and regulation. *European Journal of Biochemistry*, **262**: 247-257.

- **Livak K. J. and Schmittgen T. D.** (2001) Analysis of relative gene expression data using real time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*, **25**: 402-408.

- **Luger K., Mäder A. W., Richmond R. K., Sargent D. F. and Richmond T. J.** (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**: 251-260.

- **Mandaokar A. and Browse J.** (2009) MYB108 acts together with MYB24 to regulate jasmonate-mediated stamen maturation in *Arabidopsis*. *Plant Physiology*, **142**: 851-862.

- **Marrs K. A. and Walbot V.** (1997) Expression and RNA splicing of the maize glutathione-S-transferase Bronze2 gene is regulated by cadmium and other stresses. *Plant Physiology*, **113**: 93-102.

- **Martin C. and Paz-Ares J.** (1997) MYB transcription factors in plants. *Trends in Genetics*, **13**: 67-73.

- **Mastrangelo A. M., Belloni S., Barilli S., Ruperti B., Di Fonzo N., Stanca A. M. and Cattivelli L.** (2005) Low temperature promotes intron retention in two *e-cor* genes of durum wheat. *Planta*, **221**: 705-715.

- **Matsukura S., Mizoi J., Yoshida T., Todaka D., Ito Y., Maruyama K., Shinozaki K. and Yamaguchi-Shinozaki K**. (2010) Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. *Molecular Genetics and Genomics*, **283**: 185-96.

- **Mazzucotelli E., Mastrangelo A. M., Crosatti C., Guerra D., Stanca A. M. and Cattivelli L.** (2008) Abiotic stress response in plants: when post-transcriptional and post-translational regulations control transcription. *Plant Science*, **174**: 420–431.

- **Millar A. A. and Gubler F.** (2005) The *Arabidopsis* GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *The Plant Cell*, **17**: 705-721.

- **Miura K., Jin J. B., Lee J., Yoo C. Y., Stirm V., Miura T., Ashworth E. N., Bressan R.A., Yun D.-J. and Hasegawaa P. M.** (2007) SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1A expression and freezing tolerance in Arabidopsis. *The Plant Cell*, **19**: 1403–1414.

- **Moore J. P., Vicrè-Gibouin M., Farrant J. M. and Driouich A.** (2008) Adaptations of higher plant cell walls to water loss: drought vs desiccation. *Physiologia Plantarum*, **134**: 237–245.

- **Mu R.-L., Cao Y.-R., Liu Y.-F., Lei G., Zou H.-F., Liao Y., Wang H.-W., Zhang W.-K., Ma B., Du J.-Z., Yuan M., Zhang J.-S. and Chen S.-Y.** (2009) An R2R3-type transcription factor gene *AtMYB59* regulates root growth and cell cycle progression in *Arabidopsis. Cell Research*, **19**: 1291-1304.

- **Ner-Gaon H. and Fluhr R.** (2006) Whole-genome microarray in *Arabidopsis* facilitates global analysis of retained introns. *DNA Researh.*, **13**: 111–121.

- **Ner-Gaon H., Halachmi R., Savaldi-Goldstein S., Rubin E., Ophir R. and Fluhr R.** (2004) Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*, *Plant J*ournal, **39**: 877–885.

- **Ner-Gaon H., Leviatan N., Rubin E. and Fluhr R.** (2007) Comparative cross-species alternative splicing in plants. *Plant Physiol*ogy, **144**: 1632–1641.

- **Nesi N., Jond C., Debeaujon I., Caboche M. and 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. *The Plant Cell*, **13**: 2099-2114.

- **Oh S., Park S. and Han K.-H.** (2003) Transcriptional regulation of secondary growth in *Arabidopsis thaliana*. *Journal of Experimental Botany*, **54**: 2709-2722.

- **Patzlaff A., McInnis S., Courtenay A., Surman C., Newman L. J., Smith C., Bevan M. W., Mansfield S., Whetten R. W., Sederoff R. R. and Campbell M.M.** (2003) Characterization of a pine MYB that regulates lignification. *The Plant Journal*. **36**: 743-754.

- **Prasad M. N. V.** (1995) Cadmium toxicity and tolerance in vascular plants. *Environmental and Experimental Botany*, **35**: 525-545.

- **Priest H. D.**, **Filichkin S. A. and Mockler T. C.** (2009) Cis-regulatory elements in plant cell signalling. *Current Opinion in Plant Biology*, **12**: 643-9.

- **Puhakainen T., Hess M. W., Mäkelä P., Svensson J., Heino P. and Palva E. T.** (2004) Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. *Plant Molecular Biology*, **54**: 743–753.

- **Qu L. J. and Zhu Y. X.** (2006) Transcription factor families in Arabidopsis: major progress and outstanding issues for future research. *Current Opinion in Plant Biology*, **9**: 544-549.

- **Ramanjulu S. and Bartels D.** (2002) Drought- and desiccation-induced modulation of gene expression in plants. *Plant, Cell and Environment*, **25**: 141-151.

- **Reddy A. S. N.** (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. *The Annual Review of Plant Biology*, **58**: 267-294.

- **Richard G.-F. Kerrest A. and Dujon B.** (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiology and Molecular Biology Reviews*, **72**: 686-727.

- **Riechmann J. L. and Ratcliffe O. J.** (2000) A genomic perspective on plant transcription factors. *Current Opinion in Plant Biology*, **3**: 423–434.

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

- **Robinson R.** (2006) siRNA and DNA methylation do a two-step to silence tandem sequences. *PLoS Biology*, **4**: 1879-1880.

- **Rock C. D.** (2000) Pathways to abscisic acid-regulated gene expression. *New Phytologyst*, **148**: 357–396.

- **Rodriguez P. L.** (1998) Protein phosphatase 2C function in highr plants. *Plant Molecular Biology*, **38**: 919-927.

- **Romero I., Fuertes A., Benito M. J., Malpica J. M., Leyva A. and Paz-Ares J.** (1998) More than 80R2R3-MYB regulatory genes in the genome of Arabidopsis thaliana. *The Plant Journal*, **14**: 273–284.

- **Sambrook J., Fritsch E. F. and Maniatis T.** (1989) *Molecular cloning*. New York: Cold Spring Harbor Laboratory Press.

- **Seo P. J., Xiang F., Qiao M., Park J.-Y., Lee Y. N., Kim S.-G., Lee Y.-H., Park W. J. and Park C.-M.** (2009) The MYB96 transcription factor mediates abscisic acid signalling during drought stress response in Arabidopsis. *Plant Physiology*, **151**: 275–289.

- **Shibuya K., Fukushima S. and Takatsuji H.** (2009) RNA-directed DNA methylation induces transcriptional activation in plants. *Proceeding of National Academy of Science of the United States of America*, **106**: 1660-1665.

- **Shin B.**, **Choi G.**, **Yi H.**, **Yang S.**, **Cho I.**, **Kim J.**, **Lee S.**, **Paek N. C.**, **Kim J. H.**, **Song P. S. and Choi G.** (2002) AtMYB21, a gene encoding a flower-specific transcription factor, is regulated by COP1. *The Plant Journal*, **30**: 23-32.

- **Shinozaki K. and Dennis E. S.** (2003) Cell signalling and gene regulation: global analyses of signal transduction and gene expression profiles. *Current Opinion in Plant Biology*, **6**: 405–409.

- **Singh A., Zubko E. and Meyer P.** (2008) Cooperative activity of DNA methyltransferases for maintenance of symmetrical and non-symmetrical cytosine methylation in *Arabidopsis thaliana. The Plant Journal*, **56**: 814–823.

- **Stamm S., Ben-Arib S., Rafalskaa I., Tanga Y., Zhanga Z., Toiberb D., Thanarajc T.A. and Soreq H.** (2005) Function of alternative splicing. *Gene*, **344**: 1 –20.

- **Stracke R., Weber M. and Weisshaar B.** (2001) The R2R3-*MYB* gene family in *Arabidopsis thaliana. Current Opinion in Plant Biology*, **4**: 447-456.

- **Swindell W. R., Huebner1 M. and Weber A. P.** (2007) Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways**.** *BMC Genomics*, **8**:125-140.

- **Tähtiharju S. and Palva T.** (2001) Antisense inhibition of protein phosphatase 2C accelerates cold acclimation in *Arabidopsis thaliana*. *The Plant Journal*, **26**: 461–470.

- **Taiz L. and Zeiger E.** (2006) *Plant Physiology*, Fourth Edition, Sinauer Associates INC Publishers, Sunderland, Massachusetts.

- **Ugarkovic D.** (2005) Functional elements residing within satellite DNAs. *EMBO Reports*, **6**: 1035-1039.

- **Umezawa T., Fujita M., Fujita Y., Yamaguchi-Shinozaki K. and Shinozaki K.** (2006) Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. *Current Opinion in Biotechnology*, **17**: 113-122.

- **Umezawa T., Sugiyama N., Mizoguchi M., Hayashi S., Myouga F., Yamaguchi-Shinozaki K., Ishihama Y., Hirayama T. and Shinozaki K.** (2009) Type 2C protein phosphatises directly regulate abscisic acid-activated protein kinases in *Arabidopsis*. *Proceeding of National Academy of Science of the United States of America*, **106**: 17588-17593.

- **Urh M., Hartzell D., Mendez J., Klaubert D. H. and Wood K.** (2008) Methods for detection of protein-protein and protein-DNA interactions using HaloTag™. *Methods in Molecular Biology*, **421**: 191-209.

- **Vannini C., Locatelli F., Bracale M., Magnani E., Marsoni M., Osnato M., Mattana M., Baldoni E. and Coraggio I.** (2004) Overexpression of the rice *OsMYB4* gene increases chilling and freezing tolerance of *A. thaliana* plants. *Plant Journal*, **37**: 115-127.

- **Vergnaud G. and Denoeud F.** (2000). Minisatellites: mutability and genome architecture. *Genome Research*, **10**: 899-907.

- **Werneke J. M., Chatfield J. M. and Ogren W.L.** (1989). Alternative mRNA splicing generates the two polypeptides in spinach and *Arabidopsis*. *Plant Cell*, **1**:815–825.

- **Weston K.** (1998) Myb proteins in life, death and differentiation. *Current Opinion in Genetics & Development*, **8**: 76-81.

- **Wood A. J., Joel Duff R. and Oliver M. J.** (2000) The translational apparatus of *Tortula ruralis*: polysomal retention of transcripts encoding the ribosomal proteins RPS14, RPS16 and RPL23 in desiccated and rehydrated gametophytes. *Journal of Experimental Botany*, **51**: 1655-1662.

- **Xu Q., Modrek B. and Lee C.** (2002) Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Research*, **30**:3754-3766.

- **Yamaguchi-Shinozaki K. and Shinozaki K.** (1993) *Arabidopsis* DNA encoding two dessication-responsive rd29 genes. *Plant Physiology*, **101**: 1119-1120.

- **Yamaguchi-Shinozaki K. and Shinozaki K.** (1994) A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *Plant Cell*, **6**: 251–264.

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

- **Yoshida T.**, **Fujita Y.**, **Sayama H.**, **Kidokoro S.**, **Maruyama K.**, **Mizoi J.**, **Shinozaki K. and Yamaguchi-Shinozaki K.** (2010) AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signalling involved in drought stress tolerance and require ABA for full activation. *The Plant Journal*, **61**: 672-685.

- **Yoshida T., Nishimura N., Kitahata N., Kuromori T., Ito T., Asami T., Shinozaki K. and Hirayama T.** (2006) ABA-hypersensitive germination3 encodes a protein phosphatise 2C (AtPP2CA) that strongly regulates abscisic acid signalling during germination among *Arabidopsis* protein phosphatise 2Cs. *Plant Physiology*, **140**: 115-126.

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

- **Zhang J.** (2003) Overexpression analysis of plant transcription factors. *Current Opinion in Plant Biology*, **6**: 430-40.

- **Zhang X., Yazaki J., Sundaresan A., Cokus S., Chan S. W.-L., Chen H., Henderson I. R., Shinn P., Pellegrini M., Jacobsen S. E. and Ecker J. R.** (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis. Cell*, **126**: 1189-1201.

- **Zhang Y., Cao G., Qu L.-J. and Gu H.** (2009) Characterization of *Arabidopsis MYB* transcription factor gene *AtMYB17* and its possible regulation by LEAFY and AGL15. *Journal of Genetics and Genomics*, **36**: 99-107.

- **Zhang Z. B., Zhu J., Gao J. F., Wang C., Li H., Li H., Zhang H. Q., Zhang S., Wang D. M., Wang Q. X., Huang H., Xia H. J. and Yang Z. N.** (2007) Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. *The Plant Journal*, **52**: 528-538.

- **Zhong N., Kim C. Y., Rizzu P., Geula C., Porter D. R., Pothos E. N., Squitieri F., Heutink P. and Xu J.** (2006) DJ-1 transcriptionally up-regulates the human tyrosine hydroxylase by inhibiting the sumoylation of pyrimidine tract-binding protein-associated splicing factor. *The Journal of Biological Chemistry*, **281**: 20940-20948.

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

- **Zhu B., Chen T. H. H. and Li P. H.** (1995) Activation of two osmotin-like protein genes by abiotic stimuli and fungal pathogen in transgenic potato plants. *Plant Physiology*, **108**: 939-937.

- **Zilberman D.** (2008) The evolving functions of DNA methylation. *Current Opinion in Plant Biology*, **11**: 554-559.

PUBBLICATIONS

I) DalCorso G., Farinati S., Maistri S. and Furini A. (2008) How Plants Cope with Cadmium: Staking All on Metabolism and Gene Expression. *Journal of Integrative Plant Biology*, **50**: 1268–1280.

Authors' contributions

DalCorso G. drafted the paragraphs concerning the Cd uptake and transport and Cd toxicity in plants. Farinati S. and Maistri S. drafted the paragraphs concerning the plant responses to Cd. Furini A. drafted the introduction and the paragraph concerning the hyperaccumulator plants. All the authors read and approved the final manuscript.

II) Gazzani S., Li M., Maistri S., Scarponi E., Graziola M., Barbaro E., Wunder J., Furini A., Saedler H. and Varotto C. (2009) Evolution of *MIR168* paralogs in Brassicaceae. *BMC Evolutionary Biology*, **9**: 62-76.

Authors' contributions

Gazzani S. carried out homolog isolation, thermodynamic profile calculation and drafted the manuscript. Li M. isolated phylogenetic markers and participated in manuscript drafting. Maistri S. and Furini A. carried out *Arabidopsis* transformation. Scarponi E., Graziola M. and Barbaro E. participated in homolog isolation and sequencing. Wunder J. helped in phylogenetic reconstruction. Saedler H. participated in the design of the study and manuscript drafting. Varotto C. conceived and coordinate the study, took part in data analysis and drafted the manuscript. All the authors read and approved the final manuscript.

Journal of Integrative Plant Biology 2008, **50** (10): 1268–1280

· **Invited Review** ·

How Plants Cope with Cadmium: Staking All on Metabolism and Gene Expression

Giovanni DalCorso, Silvia Farinati, Silvia Maistri and Antonella Furini*

(*Dipartimento Scientifico e Tecnologico, Università degli Studi di Verona*, Strada Le Grazie 15, 37134 Verona, Italy)

Abstract

Environmental pollution is one of the major problems for human health. Toxic heavy metals are normally present as soil constituents or can also be spread out in the environment by human activity and agricultural techniques. Soil contamination by heavy metals as cadmium, highlights two main aspects: on one side they interfere with the life cycle of plants and therefore reduce crop yields, and on the other hand, once adsorbed and accumulated into the plant tissues, they enter the food chain poisoning animals and humans. Considering this point of view, understanding the mechanism by which plants handle heavy metal exposure, in particular cadmium stress, is a primary goal of plant-biotechnology research or plant breeders whose aim is to create plants that are able to recover high amounts of heavy metals, which can be used for phytoremediation, or identify crop varieties that do not accumulate toxic metal in grains or fruits. In this review we focus on the main symptoms of cadmium toxicity both on root apparatus and shoots. We elucidate the mechanisms that plants activate to prevent absorption or to detoxify toxic metal ions, such as synthesis of phytochelatins, metallothioneins and enzymes involved in stress response. Finally we consider new plant-biotechnology applications that can be applied for phytoremediation.

Key words: cadmium; heavy metals; metallothioneins; phytochelatins; phytoremediation; transporters.

DalCorso G, Farinati S, Maistri S, Furini A (2008). How plants cope with cadmium: staking all on metabolism and gene expression. *J. Integr. Plant Biol.* **50**(10), 1268–1280.

Available online at www.jipb.net

As sessile organisms plants have restricted mechanisms for stress avoidance and are subjected to environmental stresses that change growth conditions and alter (or sometimes disrupt) their metabolic homeostasis. Worldwide, these stresses are the most limiting factors for crop productivity: a large proportion of annual yield is lost due to pathogen attack and to unfavorable abiotic conditions such as drought, salinity and extreme temperatures. The average and record yields of many crops were compared in a classical study (Boyer 1982) and it was found that crop plants were reaching only 20% of their genetic yield potential. Diseases, insects and weeds contributed only in part, with the major yield reduction resulting from abiotic stresses. Therefore, understanding how plants cope with stresses and how, during the course of evolution, some plant species have acquired mechanisms of stress tolerance, will allow more stress-tolerant crops to be developed and significantly contribute towards increasing world food production to meet population growth requirements.

In recent decades, plant responses to stress conditions such as drought, salinity and temperature extremes have been the subject of intense molecular studies (Cushman and Bohnert 2000; Mittler 2006). Extensive expressed sequence tag (EST) collections and large scale EST sequencing initiatives for various crops have allowed these abiotic-stress genetic responses to be studied. By genetic engineering, the transfer of one or several structural genes controlled by a constitutive promoter has contributed towards protecting plants against environmental stresses (Smirnoff 1998). For instance, LEA (late embryogenesis abundant) proteins have been found in all plants in which they have been looked for and some are rapidly induced in vegetative tissues in response to water, cold or saline stress (Bies-Ethève et al. 2008). Their overexpression enhanced tolerance to salt and water stress (RoyChoudhury et al. 2007; Xiao et al. 2007). Improvement of freezing tolerance in transgenic plants has been achieved by expressing a

Received 14 Mar. 2008 Accepted 23 Jun. 2008

*Author for correspondence.

Tel: +39 45 802 7950;

Fax: +39 45 802 7929;

E-mail: <antonella.furini@univr.it>.

doi: 10.1111/j.1744-7909.2008.00737.x

cold-responsive (COR) gene of wheat (Shimamura et al. 2006). Furthermore, transcription factors play a crucial role in signal transduction pathways by controlling the expression of specific genes in response to environmental factors. Ectopic expression of the dehydration-responsive element (DRE)-binding protein DREB1A in *Arabidopsis* plants, resulted in improving freezing, salinity and drought tolerance and highlights the importance of regulatory controls for key stress-tolerance components (Kasuka et al. 1999). More recently, Nelson and co-workers demonstrated the potential of a transgene strategy for improving drought tolerance in crop plants (Nelson et al. 2007). They identified a transcription factor in *Arabidopsis* from the nuclear factor Y (NF) family, *AtNF-YB1*, that confers drought tolerance. A maize crop that constitutively expressed the orthologous gene *ZmNF-YB2* showed drought tolerance and improved yield under water-limited conditions.

Nowadays, plants also have to face rapid environmental changes mainly due to human activities causing air and soil pollution, acid precipitation, climate change etc. Plants thus need to adapt to changing environmental conditions in order to tolerate new stresses. Heavy metals, for instance, occur naturally in soils as rare elements. However, traffic, refuse dumping, and metal working industries contribute towards the spread of heavy metals in the environment. In agricultural soils, heavy metal pollution is an increasing problem due to soil amendment and the intense use of phosphate fertilizers that contain cadmium (Cd) as a contaminant (Polle and Schützendübel 2003). The latter is a highly toxic pollutant to prokaryotic and eukaryotic organisms also due to its solubility in water, which determines a rapid distribution in the environment. Uptake of Cd by crop plants is the main entry pathway into the food chain causing serious problems to human health (Buchet et al. 1990). Plants try to avoid Cd stress by preventing absorption; however the uptake depends on the soil metal concentration and pH. If taken up by plants, Cd is transported into the root by metal transporters or Ca channels (Perfus-Barbeoch et al. 2002). Within the cell Cd is detoxified preferentially by binding to S-containing ligands such as metallothioneins, glutathione and phytochelatins and the ligand-Cd complexes are most likely removed by sequestration from potentially sensitive organelles and structures (Cobbett 2000; Cobbett and Goldsbrough 2002; Clemens 2006). Nevertheless, exposure of plants to Cd stress may lead to the alteration of many cellular processes and structures (Hall 2002). Cd accumulation in plants causes reductions in photosynthesis, diminishes water and nutrient uptake (Sanità di Toppi and Gabbrielli 1999), inhibition of enzyme activities, disruption of cell transport processes, disturbance of cellular redox control (Clemens et al. 2001; Schützendübel and Polle 2002) and affects general root and shoot growth.

Cadmium is generally toxic to most plant species, and metallophytes in Cd-containing soils have evolved mechanisms of Cd exclusion by inhibiting its entry and hindering its transport to the above-ground tissue. However, a number of so-called hyperaccumulator plants, endemic to metal-rich soils, can accumulate Cd in their aerial organs to a level that is orders of magnitude higher than that normally found in plants, without showing any sign of phytotoxicity (Baker and Brooks 1989). The complex mechanisms evolved in these plants either to prevent Cd uptake or to enable metal extraction from soils, trafficking, allocation and cellular detoxification are research areas that have attracted the attention of many investigators. Indeed, to identify the genetic and physiological mechanisms of plants that are able to avoid Cd stress by preventing its absorption or translocation is needed to prevent Cd entering the food chain. Likewise, understanding the mechanism of Cd accumulation in the vegetative parts of hyperaccumulator plants is crucial to the promising approach of using plants as ecological remediation of Cd polluted environments (phytoremediation). In this review we have chosen to focus on plant mechanisms that allow Cd detoxification and absorption prevention. We will also discuss the potential biotechnological application in phytoextraction of Cd from polluted sites.

Cd Uptake and Transport

Higher plants can uptake Cd, depending on its availability and concentration, in soil or water; rather little is taken up directly from the atmosphere (Clemens 2006). Soil pH, the rhizosphere and presence of organic acids modulate the bio-availability of Cd (as well as of other heavy metals) for plant uptake (Benavides et al. 2005). For instance, it has been reported that Cd uptake in corn was lower in acid soils with high organic matter content (Benavides et al. 2005). The concentration of other nutrient elements (e.g. Ca, Zn and Fe) in the soil also influences Cd absorption; it has been shown that addition of Ca or Zn diminishes the Cd uptake (Cosio et al. 2004).

Since membrane potential, which might exceed −200 mV in root epidermal cells, provides a strong driving force for the uptake of cations (Benavides et al. 2005), toxic heavy metals compete with and gain access into the plant cell via the transport systems operating for micronutrient uptake: in particular, the uptake of Cd ions occurs via the same transmembrane carriers used to uptake Ca^{2+}, Fe^{2+}, Mg^{2+}, Cu^{2+} and Zn^{2+} (Roth et al. 2006; Papoyan et al. 2007). Due to its high mobility and water solubility, Cd readily enters the roots through the cortical tissue and can reach the xylem via an apoplastic and/or symplastic pathway, complexed to organic acids or phytochelatins (Salt et al. 1995). Once loaded into the tracheary elements, Cd complexes spread throughout the entire plant following the water stream. It has been hypothesized that Cd accumulation in developing fruits could occur via phloem-mediated transport, implicating a systemic diffusion of the heavy metals into the plant body (Benavides et al. 2005).

Cd Toxicity in Plants

Together with Ag, As, Hg, Pb and Sb, Cd is considered a non-nutrient element, since it has no known function in plant development and life, with the exception of the Cd-carbonic anhydrase of marine diatoms (Lane and Morel 2000). Furthermore, these heavy metals seem to be more or less toxic to eukaryotic organisms and microorganisms (Sanità di Toppi and Gabbrielli 1999; Benavides et al. 2005). The large majority of studies have been based on the application of extremely high Cd concentrations and so the consequences of acute Cd stress are well-documented. In higher plants, Cd negatively affects both plant growth and development, resulting in stunting and eventually plant death. The critical concentration, at which the metal causes injuries in plant physiology is in the range of 3 to 10 mg/kg dry mass (Bahlsberg-Pahlson 1989).

The bases of Cd toxicity are still not completely understood, but it might result from its high affinity for sulfydryls (e.g. threefold higher than Cu ions, Schützendübel and Polle 2002). Cd, binding to sulfydryl groups of structural proteins and enzymes, leads to misfolding, inhibition of activity and/or interference with redox-enzymatic regulation (Hall 2002). Another important toxicity mechanism is due to the chemical similarity between Cd^{2+} and functionally active ions situated in active sites of enzymes and signaling components. Thus, Cd^{2+} ions can interfere with homeostatic pathways for essential metal ions (Roth et al. 2006) and the displacing of divalent cations, such as Zn and Fe, from proteins could cause the release of "free" ions, which might trigger oxidative injuries via free Fe/Cu-catalyzed Fenton reaction (Polle and Schützendübel 2003).

It has to be noticed that *in vivo* Cd-related injuries on plants depend first on the plant species: hyperaccumulators or genetically resistant plants activate cellular mechanisms that weaken the impairment due to Cd stress. Moreover, time of Cd exposure and its magnitude together with external environmental conditions, contribute to modulating plant sensitivity to heavy metals (Sanità di Toppi and Gabbrielli 1999).

Photosynthesis and carbon assimilation

The most evident symptoms of Cd toxicity are leaf roll and chlorosis, water uptake imbalance and stomatal closure (Clemens 2006). Chlorosis might be due to changes in Fe : Zn ratio caused by Cd, as in corn leaves (Root et al. 1975) and to the negative effects on chlorophyll metabolism (Chaffei et al. 2004). At the cellular level, Cd damages the photosynthetic apparatus, particularly the light harvesting complex II and the two photosystems and causes a decrease in chlorophyll and carotenoid content (Figure 1A), leading to higher non-photochemical quenching (Sanità di Toppi and Gabbrielli 1999). Regarding stomatal closure, it has been shown that during Cd exposure, stomata close independently of water status. Stomatal closure can be actively driven by abscisic acid (ABA)-induced Ca^{2+} accumulation in the cytosol of the guard cells. The increase in cytosolic free Ca^{2+} promotes opening of plasma membrane anion and K^{+}_{out} channels. As more ions leave the cell, water follows and turgor is lost, with stomatal pore closure (MacRobbie and Kurup 2007). Being chemically similar to Ca ions, Cd probably enters guard cells through voltage-dependent Ca^{2+} channels and once in the cytosol, it mimics Ca^{2+} activity (Perfus-Barbeoch et al. 2002).

All together, stomatal closure, damage to the photosynthetic machinery and interference with pigment synthesis, cause a general depression of the photosynthetic efficiency lowering the effective quantum yield. Moreover, by inhibiting enzymes involved in CO_2 fixation, Cd decreases carbon assimilation (Perfus-Barbeoch et al. 2002).

Effects on nutrient uptake and root physiology

Cadmium, as do other heavy metals, imbalances the water uptake and nutrient metabolism (uptake, transport and use) at the root level interfering with the uptake of Ca, Mg, K and P (Benavides et al. 2005). The inhibition of the root Fe(III) reductase induced by Cd leads to a Fe(II) deficiency in cucumber and sugarbeet (Alcantara et al. 1994).

In different plant species (e.g. tomato, maize, pea and barley) Cd alters the activity of different enzymes involved in nitrogen metabolism (Nussbaum et al. 1988; Boussama et al. 1999). At the root level, the reduction of nitrate absorption may be due to transpiration inhibition. Moreover, both the nitrate reductase and nitrite reductase activity in roots and leaves are affected (Chaffei et al. 2004) as well as nitrate transport from roots to shoots (Sanità di Toppi and Gabbrielli 1999) leading to a reduced nitrate assimilation by the whole plant (Figure 1A). The activity of the enzymes responsible for the incorporation of ammonium molecules into the carbon skeleton (i.e. glutamine and glutamate synthetase) is also compromised (Chaffei et al. 2004). On the other hand, the activity of the glutamate dehydrogenase (GDH) is enhanced during Cd-stress (Boussama et al. 1999). Because high activity of GDH enzyme has been related with pathogen response and senescence induction (Osuji and Madu 1996; Masclaux et al. 2000) and changes in nitrogen metabolism due to Cd stress are similar to the ones induced during senescence, it has been hypothesized that Cd induces senescence-like symptoms at least in tomato, leading to nitrogen mobilization and a storage strategy (Chaffei et al. 2004).

Regarding sulfur metabolism, exposure to Cd induces a remarkable increase in the amount of thiol compounds, with a concomitant decrease in the activity of leaf Adenosine Triphosphate (ATP)-sulfurylase and O-acetylserine sulfurylase (Figure 1A), the first and the last enzymes involved in the sulfate assimilation pathway (Astolfi et al. 2004). It is noticeable that also nitrogen

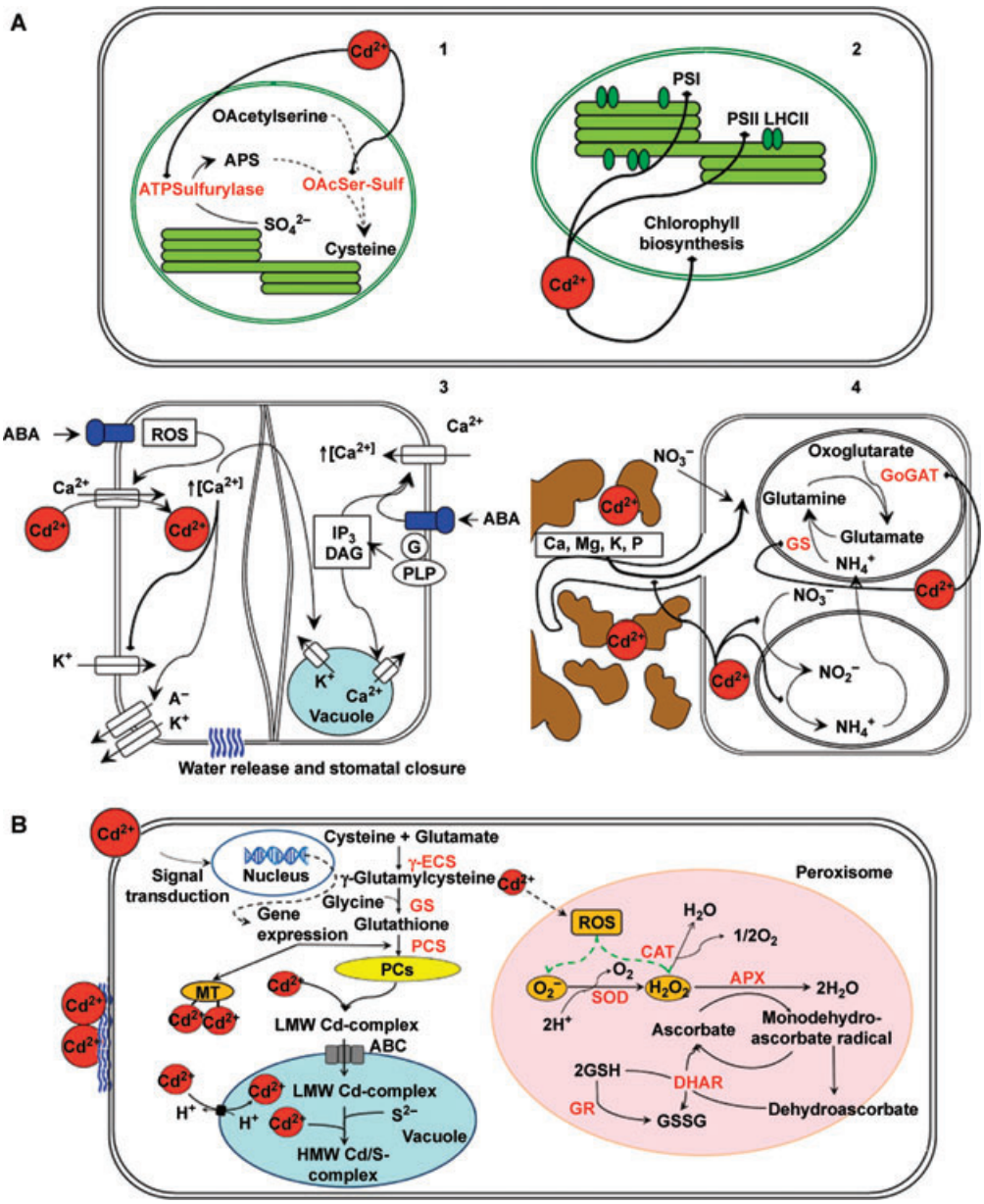


Figure 1. Main effects of Cd in plant physiology and plant responses to Cd stress.

(A) Effects of Cd on shoot cells: after uptake into the cell (through a still unidentified metal transporter) Cd inhibits sulfur metabolism **(1)** photosynthesis and chlorophyll biosynthesis **(2)**. **(3)** Mimicking Ca^{2+} ions, Cd enters stomatal guard-cell and activates the opening of the plasma membrane anion and K^{+}_{out} channels. As more ions leave the cell, water follows and turgor is lost, with stomatal pore closure (see the text for further explanation). **(4)** Cd induces inhibition of root enzymes involved in nitrogen assimilation: both nitrate and nitrite reductase activity are inhibited. Moreover, Cd interferes with activity of both GS and GOGAT enzymes, involved in ammonium assimilation.

(B) Constituents of the cell wall can immobilize Cd ions preventing cytosolic uptake. Once Cd enters the cytosol, it stimulates the synthesis of phytochelatins and probably metallothioneins. After complexation with Cd, the low-molecular-weight (LMW) complex enters the vacuole via a tonoplast-localized ATP-binding-cassette (ABC) transporter and further complex in high-molecular-weights (HMWs). On the right, particular of the Cd-induced activation of ROS-scavenging cycle taking place mainly in peroxisomes.

ABA, abscissic acid; APX, ascorbate peroxidase; CAT, catalase; DAG, diacylglycerol; DHAR, dehydro-ascorbate reductase; G, G-proteins, activating PLP enzyme; GOGAT, glutamate synthase; GR, glutathione reductase; GS, glutamine synthase; LHCII, light harvesting complex II; MT, metallothioneins; PCs, phytochelatins; PCS, phytochelatin synthetase; PLP, phospholipase protein, involved in DAG and IP_3 mediated signaling; PSI, photosystem I; ROS, reactive oxygen species; IP_3, inositol-3-phosphate; SOD, supeoxide dismutase; γ-ECS, γ-glutamylcysteine synthetase.

fixation and primary ammonia assimilation decreased in nodules of soybean plants during Cd treatments (Balestrasse et al. 2003).

Phenotypically, Cd exposure inhibits root growth and lateral root formation while it induces differentiation of numerous root hairs both in *Arabidopsis* and tobacco plants (S Farinati et al., unpubl. data, 2008). In tomato, Cd-treated roots were thicker and stronger and the root biomass was less affected than the leaves (Chaffei et al. 2004).

Symptoms on cellular homeostasis

In different plant species, such as bean and wheat, cytotoxicity of Cd exposure appears as chromosomal aberrations and inhibition of mitotic processes with consequent altered cell cycle and division (Benavides et al. 2005). Furthermore, Cd causes high mutation rates in *Arabidopsis thaliana*, floral anomalies (banding), poor seed production and malformed embryos (Ernst et al. 2008). It also induces vacuolization and mitochondrial degeneration (Silverberg 1976) affecting cell metabolism and aerobic respiration.

Although Cd does not participate directly in cellular redox reactions (i.e. Cd ions do not alter their oxidation state, since they do not take part in Fenton and Haber-Weiss reactions, Clemens 2006), its exposure drives oxidative injuries, such as lipid peroxidation, which leads to alteration in the membranes functionality, and protein carbonylation (Schützendübel et al. 2001; Romero-Puertas et al. 2002) and converges into a general redox homeostasis impairment. Cd unbalances the activity of antioxidative enzymes and affects catalase and super-oxide dismutase (SOD) activity triggering H_2O_2 and O_2^- (reactive oxygen species, ROS) over-accumulation (Romero-Puertas et al. 2004). It is still not clear if the over-production of ROS during Cd treatment is the cause of redox cellular imbalance or if this is a specific stress mechanism activated by the plant cell to cope with the heavy metal ions (Romero-Puertas et al. 2004). It was also shown that Cd induces peroxisome-senescence in leaves activating the glyoxylate cycle enzymes, malate synthase and isocitrate lyase, as well as peroxisomal peptidases, the latter being well-known as leaf senescence-associated factors (Chaffei et al. 2004).

Furthermore, Cd interferes with plasma membrane ion transporters and ATPase (Sanità di Toppi and Gabbrielli 1999) disturbing ion and metabolite movement and accumulation. In addition, Cd exposure inhibits the activity of metabolic enzymes such as glucose-6-phosphate dehydrogenase, glutamate dehydrogenase, malic enzyme, isocitrate dehydrogenase, Rubisco and carbonic anhydrase (Sanità di Toppi and Gabbrielli 1999).

Cd effects on intracellular signaling

A secondary effect due to the accumulation of ROS in the cell compartments is the alteration of the signaling mediated by H_2O_2 and other oxygen species. It was widely accepted that H_2O_2 can play a role as signal molecule in triggering the induction of defense mechanisms against both abiotic stresses, such as temperature and ozone (Dat et al. 2000; Sharma et al. 1996) and pathogen attack as infections due to bacteria or powdery mildew fungi (Thordal-Christensen et al. 1997; Bestwick et al. 1998). Interfering with H_2O_2 accumulation, Cd meddles with the signal transduction pathways in which ROS are involved. Being chemically very similar to Zn^{2+}, Cd^{2+} ions can hamper the activity of Zn-finger transcription factors, substituting Zn ions and consequently interfering with transcription mechanisms (Sanità di Toppi and Gabbrielli 1999). With similar mechanisms, Cd^{2+} replace Ca^{2+} ions in calmodulin proteins, causing the perturbation of intracellular calcium level and altering the calcium-dependent signaling (Ghelis et al. 2000; Perfus-Barbeoch et al. 2002).

Plant Responses to Cd

Plants, like all other organisms, have evolved a complex network of homeostatic mechanisms to minimize the damages from exposure to nonessential metal ions. To avoid Cd toxicity, land plants developed active and passive strategies of exclusion of the heavy metal ion from the cellular environment. As first defense to Cd stress, plant exudates such as malate or citrate bind to metal ions in the soil matrix excluding them from the root absorption (Delhaize and Ryan 1995). Second, the cell wall (through pectic sites and hystidyl groups) and extracellular carbohydrates (callose, mucilage) can play a significant role in immobilizing toxic ions and preventing their uptake into the cytosol (Sanità di Toppi and Gabbrielli 1999). Nevertheless, as soon as the concentration of the toxic element rises above the physical adsorption limit of these barriers, active metabolism takes charge producing chelating compounds (phytochelatins and, in some cases, metallothioneins) involved in the detoxification and compartmentalization of the heavy metals in specific cellular compartments. Moreover, as for other abiotic stresses, Cd resistance involves the synthesis of stress-related proteins and signal molecules (heat shock proteins, salicylic and abscisic acids, ethylene) (Sanità di Toppi and Gabbrielli 1999). The signal transduction pathway is characterized by a complicated interaction of genes in which transcription factors have essential roles since regulation of their expression may strongly affect plant stress response (Uno et al. 2000). With the recent introduction of genomics technology it has been possible to identify numerous putative genes involved in response to Cd-stress: for example, several genes induced in *Brassica juncea* after Cd treatments were identified by cDNA Amplified Fragment Lenght Polimorfism (AFLP) (Fusco et al. 2005). Changes in the transcriptome of *Arabidopsis* plants exposed to Cd and Pb were studied by Affymetrix DNA array (Kovalchuk et al. 2005).

Cd induces modulation of gene expression

Responses to heavy metal stress depend on an intricate signal transduction pathway within the cell that begins with the sensing of heavy metal (or heavy metal associated symptoms) and converges in transcription regulation of metal-responsive genes (Singh et al. 2002). Still much remains unknown about the molecular components of the metal-induced signal transduction, and only recently thanks to differential-expression analyses has it been possible to identify transcription factors (TFs) putatively responsive to heavy metal stress (Fusco et al. 2005). As commonly found for other stress-related TFs, heavy metal responsive TFs also share the same signal transduction pathway and are therefore activated by abiotic stresses such as cold, dehydration, Salicylic Acid (SA) and H_2O_2 (Singh et al. 2002). In addition, cross-talk also exists between Cd tolerance mechanisms and pathogen defense signaling (Suzuki et al. 2001).

Cadmium affects the expression of ERF proteins that belong to the APETALA2 (AP2)/ethylene-responsive-element-binding protein (EREBP) family. Members of these TFs can bind to several pathogenesis-related promoters and dehydration-responsive elements (DRE motif) (Singh et al. 2002). It has been shown that *ERF1* and *ERF2* genes are induced after 2 h of Cd-treatment in *A. thaliana* roots (Weber et al. 2006). Moreover, it has been reported that DREB2A is induced by Cd: DREB2A specifically interacts with the DRE motif in the promoter region of the rd29A and activates its transcription in Cd-exposed plants. Rd29A is already known to be induced by cold, salt and dehydration stresses (Suzuki et al. 2001). OBF5, a bZIP-type DNA binding protein, was shown to be Cd-induced: it binds to the promoter region of the gene coding for the glutathione S-transferase, an enzyme involved in ROS scavenging and xenobiotic detoxification (Suzuki et al. 2001). Furthermore, it has recently been demonstrated that the expression of BjCdR15, a bZIP protein identified in *B. juncea*, is induced after short Cd treatment (Fusco et al. 2005). This TF controls the expression of several metal transporters, is involved in long distance root-to-shoot Cd transportation and its overexpression in *A. thaliana* and tobacco plants enhances Cd tolerance and accumulation in the shoot (S Farinati et al., unpubl. data, 2008).

WRKY53, a TF belonging to the WRKY family, was isolated as being differentially expressed in Cd-treated *Thlaspi caerulescens* plants. This gene is also modulated by other environmental stresses such as salinity, drought, cold and salicylic acid and seems to participate in the stress-related signal transduction pathway regulating the activity of other TFs rather than directly activating gene expression (Wei et al. 2008).

MYeloBlastosis Protein (MYB) proteins, and in particular members of the R2R3 MYB subgroup also respond to heavy-metal stress: in *A. thaliana* MYB4 is more highly expressed after Cd and Zn-treatment (Van de Mortel et al. 2008), while MYB43, MYB48 and MYB124 proteins are specifically induced by Cd in roots (Weber et al. 2006). In *T. caerulescens* MYB28 is strongly expressed under Zn deficiency and high Cd concentrations and is involved in the regulation of glucosinolate (GSL) synthesis (Van de Mortel et al. 2008). GSL plays an important role as a storage form of sulfur and its biosynthesis responds to changes in nutritional status, biotic and abiotic stresses (Hirai et al. 2007). As already mentioned, Cd interferes with nutrient uptake and sulfur metabolism.

The modulation of TF belonging to different groups may indicate the complexity of the response of plants to Cd, from the signal perception to the intracellular transduction cascade triggering the activation of genes responsible for Cd uptake, transport and detoxification.

Phytochelatins

Cadmium can induce the synthesis of small metal-binding peptides defined as phytochelatins (PCs). PCs have the general structure $(\gamma\text{-Glu-Cys})_n\text{-X}$ where n is a variable number from 2 to 11 and X an amino acid such as Gly, β-Ala, Ser, Glu or Gln (Cobbett and Goldsbrough 2002). Due to the presence of the thiolic groups of Cys, PCs chelate Cd and form several complexes with molecular weight of about 2 500 or 3 600 Da, protecting the cytosol from free Cd ions (Cobbett 2000). Glutathione is the building block for PCs synthesis, which is catalyzed by the cytosolic PCs synthetase (PCS). It has been shown that PCS is constitutively expressed and post-translationally activated by heavy metals (Cobbett and Goldsbrough 2002).

Due to their metal ion affinity, PCs are supposed to play a role in cellular homeostasis and trafficking of essential nutrients such as Cu and Zn (Thumann et al. 1991) and they are required for detoxification of toxic metals, particularly to Cd, as confirmed in both *Arabidopsis* and *Schizosaccharomyces pombe*, by the Cd-sensitive phenotype of *cad1* mutants defective in PCS activity (Ha et al. 1999). Regardless, an excessive amount of PC does not confer, *per se*, any hyper-tolerance; indeed, although an enhanced PCs synthesis seems to increase heavy metals accumulation in transgenic plants (Pomponi et al. 2006), an excessive expression of *AtPCS* genes determines hypersensitivity to Cd stress (Lee et al. 2003). After synthesis, PCs bind the heavy metal ions and facilitate their transport as complexes into the vacuole (Clemens 2006) where they eventually form high-molecular-weight (HMW) complexes (Figure 1B). Several studies demonstrated that in *Arabidopsis* the transport of HMW complexes across the tonoplast is mediated by ATP-binding-cassette (ABC) transporters (Cobbett and Goldsbrough 2002). It was also reported that PCs play a role in Cd transport from root to shoot and it was demonstrated that a PCs-dependent "overflow protection mechanism" would contribute to keeping Cd accumulation low in the root, causing extra Cd transport to the shoot (Gong et al. 2003).

Metallothioneins

Metallothioneins (MTs) are other cysteine-rich peptides with a low molecular weight able to bind metal ions by means of mercaptide bonds. Differently from PCs, MTs are products of mRNA translation, induced in response to heavy metal stress (Cobbett and Goldsbrough 2002). MT proteins in vertebrates are characterized by a stretch of 20 Cys residues highly conserved, whereas plant and fungi isoforms do not contain this structure (Cherian and Chan 1993).

Regarding their metal binding activity, the pea MT (PsMTa) can bind Cd, Zn and Cu when expressed in *Escherichia coli* (Tommey et al. 1991). Moreover, *Arabidopsis* MTs restore tolerance to copper in MT-deficient yeast strains (Zhou and Goldsbrough 1994). Although the role of plant MTs in Cd tolerance is still almost unknown, there is some evidence that supports their participation in Cu homeostasis (Cobbett and Goldsbrough 2002). Moreover, overexpression of mouse MT in tobacco plants enhances Cd tolerance *in vitro* (Pan et al. 1994), whereas *Brassica juncea* MT2, ectopically expressed in *Arabidopsis thaliana*, confers increased tolerance to Cd and Cu (Zhigang et al. 2006).

In terms of transcript amount, many plant MT genes are expressed at very high levels in all tissues. *Arabidopsis* MT1a and MT2a seem to accumulate in trichomes, being involved in sequestration of heavy metal ions in these structures (Salt et al. 1995). Since *Arabidopsis* MT expression has been detected in phloem elements, a role in metal ion transport has been postulated (Garcia-Hernandez et al. 1998). Finally, MT genes are expressed during various stages of plant development and in response to different environmental conditions (Rauser 1999). For instance, the MT gene of wheat and rice can be induced by a variety of metal ions, such as Cu, Cd and Al, and abiotic stresses, such as high temperature and deficiency of nutrients (Cobbett and Goldsbrough 2002). A number of MT genes have been isolated from ripening fruits (Rauser 1999) and they probably have a role in normal development processes.

Metal ion transporters

Land plants possess a highly effective metal ions uptake system that allows the acquisition of metal ions and other inorganic nutrients from soil by plant roots. Therefore metal transporters, situated in the tonoplast or plasma membrane, play a central role in the maintenance of metal homeostasis within physiological limits. In fact, Cd tolerance is correlated with its extrusion or intracellular compartmentalization mediated by the activation of specific transport processes.

Generally, metal transporters appointed to ion import show low selectivity. For example A*t*IRT1 (localized in the plasma-membrane of root cells) is the primary root iron uptake system in *Arabidopsis* but can transport significant amounts of Cd (Korshunova et al. 1999). On the other hand, intracellular transporters that export metal ions from the cytosol to both vacuoles or outside the cell, are highly selective. For instance, tonoplastic transporters A*t*MTP1 and A*t*MTP3 specifically export Zn into the vacuole (Krämer et al. 2007).

An important group of metal transporters is the ZIP (ZRT, IRT-like protein) family, plasma-membrane transport proteins that are induced both in roots and shoots of *Arabidopsis* in response to Zn-limiting conditions. ZIP members have now been identified in several plant species, as well as in bacteria, fungi and animals and results indicate that they are involved in divalent cations transport across the membranes (Lòpez-Millàn et al. 2004). Members of this family are thought to be implicated in Cd uptake from the soil into the root cell and in cadmium root-to-shoot transport, being involved in the xylem unloading process (Krämer et al. 2007). Enhanced root metal uptake mediated by ZIP transporters seems to be a factor necessary, but not sufficient, for hyperaccumulation in the model species *Arabidopsis halleri* and *T. caerulescens* (Krämer et al. 2007) and accumulation capacity in these plants varies with the expression of these proteins. For instance, in *A. halleri*, ZIP9 has a high expression level in roots already under Zn-sufficient conditions, while it is upregulated in shoots in response to Zn-deficiency (Krämer et al. 2007). On the contrary, in *A. thaliana*, ZIP9 is induced during Zn-deficiency in both root and shoot. Similarly, ZIP6 is highly expressed in hyperaccumulators under Zn-deficiency, whereas it is not induced in *A. thaliana* (Becher et al. 2004; Filatov et al. 2006).

The family of natural resistance-associated macrophage protein (NRAMP) metal ion transporters represents another important group of transmembrane protein involved in metal transport and homeostasis. These transporters are considered as "general metal ion transporters" due to their ability to transport Mn^{2+}, Zn^{2+}, Cu^{2+}, Fe^{2+}, Cd^{2+}, Ni^{2+} and Co^{2+} (Nevo and Nelson 2006). Like the ZIP transporters, members of the NRAMP family share remarkable protein sequence identity among plants, yeast and mammalians (Nevo and Nelson 2006). By cDNA microarray, it has been shown that the expression level of NRAMP genes is higher in hyperaccumulator species (Chiang et al. 2006). They are expressed in both root and shoot and are implicated in the transport of metal cations across the plasma-membrane into the cytosol or across the tonoplast (Krämer et al. 2007). In *A. thaliana* these metal transporters participate principally in Fe homeostasis. In heterologous systems, three members of the *Arabidopsis* NRAMP family, *At*NRAMP1, *At*NRAMP3 and *At*NRAMP4, can mediate uptake of Fe, Mn and Cd (Curie et al. 2000). Interestingly, the overexpression of *AtNramp3* results in Cd hypersensitivity of *Arabidopsis* root growth and in an increased accumulation of Fe (Thomine et al. 2000). These results lead us to suppose that NRAMP metal transporters are able to transport both Fe and Cd *in planta* (Thomine et al. 2000).

Concerning efflux systems, metal transporters P_{1B}-ATPases (HMA) translocate metal ions out of the cytoplasm (both

outside the plasma membrane or into the vacuole) hydrolyzing ATP. As already mentioned, export metal-transporters are more selective than import-transporters: indeed, HMA members (e.g. HMA2, HMA3 and HMA4) export Zn and Cd exclusively (Krämer et al. 2007). Recent works highlighted that members of this family (*Ah*HMA4, *Ah*HMA3 and *Tc*HMA4 deriving from hyperaccumulator species *A. halleri* and *T. caerulescens* respectively) are able to confer Cd or Zn tolerance when expressed in yeast (Bernard et al. 2004; Papoyan and Kochian 2004). Therefore, it has been proposed that *Ah*HMA4, *Tc*HMA4 and probably *At*HMA4, its homolog in *A. thaliana*, may contribute to Cd and Zn homeostasis extruding the metal ions from the cytosolic compartment (Krämer et al. 2007). Furthermore, their expression mainly in the vascular system of root and shoot suggests an implication of these transporters in metal root-to-shoot transport (Verret et al. 2004).

Recently, ABC transporters have been shown to be implicated in a range of processes that encompasses polar auxin transport, lipid catabolism, disease resistance, stomatal function, xenobiotic and metal detoxification (Kim et al. 2006; Rea 2007). Examples are the ABC family of the mitochondria in *Arabidopsis* (*At*ATM). It has been found that *At*ATM3 is upregulated in roots of plants treated with Cd and Pb. Moreover, *At*ATM3-overexpressing plants were more tolerant to Cd, whereas *AtATM3* mutants showed increased sensitivity. The *At*ATM3 homolog in *Schizosaccharomyces pombe* (HMT1) is a tonoplast transporter exporting Cd-phytochelatin complexes. Similarly, it has been hypothesized, but has still to be demonstrated, that *At*ATM3 has a role in extruding Cd-GSH complexes formed in the mitochondria and that the sensitivity of the mutant is due to the oxidative damage of Cd accumulation in this organelle (Kim et al. 2006). *At*PDR8 is another ABC transporter in *A. thaliana* involved in metal homeostasis: it was demonstrated that *At*PDR8 participates in both Cd tolerance and pathogen resistance (Kobae et al. 2006; Stein et al. 2006). Not only is its expression induced by Cd, but its overexpression induces lower Cd accumulation in root and shoots. *At*PDR8 is mainly localized in the membrane of root hair and epidermis (Kim et al. 2007). It is proposed that *At*PDR8 might confer Cd tolerance by pumping it out of the plasma membrane to the apoplast (Kim et al. 2007).

Finally, members of the "cation diffusion facilitator" (CDF) transporter group seem to mediate vacuolar sequestration, storage and transport of metal ions from the cytoplasm to the outer compartment (Krämer et al. 2007). CDF transporters have been characterized in both prokaryotes and eukaryotes and can transport across membranes divalent metal cations such as Zn, Cd, Co, Fe, Ni or Mn (Montanini et al. 2007).

Enzymes

As already mentioned, toxicity of heavy metals determines altered activity and accumulation of different enzymes (Prasad 1995). For example, Cd inhibits the activity of enzymes involved in carbon assimilation (e.g. Rubisco) probably through reaction with Sulphidric Groups (SH) groups of the protein interfering with its folding or activity (Prasad 1995). Furthermore, treatment with Cd increases Mg dependent ATPase activity and induces diacylglycerol (DAG) kinase in roots of *B. juncea*, suggesting that Cd may activate the lipid signaling pathway (Lang et al. 2005). It has been reported that Cd can affect protein kinase expression in *Arabidopsis* (Suzuki et al. 2001) and that mitogen activated protein kinase (MAPK) cascade is involved in the Cd-signaling pathway in rice and alfalfa plants (Romero-Puertas et al. 2007). Under Cd stress, enzymes involved in primary nitrogen assimilation and nitrogen mobilization are impaired (Chaffei et al. 2004). It has been shown that the total glutamine-synthetase (GS) activity decreases (Figure 1A). Specifically, in shoots the plastidic GS isoform is decreased both in activity and expression, whereas the gene transcription of the cytosolic isoform is increased. In roots, the mRNA of the cytosolic GS isoform accumulates. This suggests that when Cd affects the plastidic-GS activity, plants induce the cytosolic isoform to compensate and maintain glutamine biosynthesis (Chaffei et al. 2004). On the other hand, a response mechanism to overcome heavy metal stress is the production of PC by PCS. PCS is activated, both *in vivo* and *in vitro*, by a wide range of metals and metalloids, such as Cd, Ag, Pb, Cu, Hg, Zn, As and Au (Schat et al. 2002). The activation mechanism is still unknown and PCS was believed to sense heavy metals directly binding the metal ion, but it has been proved that its catalytic activation does not depend on this binding (Vatamaniuk et al. 2000). The Cd-induced expression of PCS genes has been examined in *A. thaliana* and in *B. juncea* and the results are, in most cases, contradictory. Cazale and Clemens (2001) demonstrated that the *AtPCS1* and *AtPCS2* genes are constitutively expressed and not transcriptionally regulated by Cd, whereas other authors found that the *AtPCS1* level of transcript, but not of protein, is responsive to Cd (Lee et al. 2002). Furthermore the level of PCS protein was enhanced in leaves, but not in roots, of *B. juncea* after prolonged Cd exposure (Heiss et al. 2003) suggesting that the effects of Cd on PCS expression may also vary with the plant organ and species.

Cadmium causes oxidative stress by inducing generation of ROS and by disturbing the antioxidative systems in their scavenging (Schützendübel et al. 2001; Romero-Puertas et al. 2004). Catalase (CAT) represents a key enzyme for the defense responses against oxidative stress (Figure 1B). It is present only in peroxisomes and catalyzes the H_2O_2 breakdown (Buchanan et al. 2000). In *B. juncea* four distinct CAT sequences have been cloned and it has been shown that Cd exposure causes an increase of *CAT3* transcript. This induction could be useful to limit high H_2O_2 concentration in order to protect the cell from oxidative stress (Lang et al. 2005). In pea plants, Cd-induced oxidation of CAT protein determines reduced CAT activity and

protein content. As a compensatory mechanism, in response to Cd, CAT transcription is upregulated (Romero-Puertas et al. 2007).

The cycle ascorbic acid-glutathione is activated as a ROS scavenging mechanism. The main enzymes of these reactions are modulated by Cd, which induces increased activity of the ascorbate peroxidase (the first enzyme of the cycle) in *Phaseolus vulgaris* and in *Pisum sativum* (Romero-Puertas et al. 2007). In addition, another enzyme taking part in the cycle, the glutathione reductase (GR), is differently induced in roots and leaves of Cd treated pea plants (Yannarelli et al. 2007).

Super-oxide dismutase enzyme plays a role in protecting cells against ROS accumulation. SOD activity was induced in tomato seedlings after prolonged Cd treatment (Dong et al. 2006). Moreover, a significant increase of SOD activity was shown in wheat leaves, but only under exposure to high Cd concentration, probably due to the high production of superoxide (Lin et al. 2007). Nevertheless, it has to be considered that previous studies showed that SOD activity decreased in response to Cd toxicity in pea plants (Romero-Puertas et al. 2007).

Hyperaccumulator Plants: A New Frontier of Plant Biotechnology

Heavy metal hyperaccumulators are a unique group of plants that can accumulate high amounts of various toxic elements in their tissues (Reeves and Baker 2000). Hyperaccumulation is an active process that depends on an internal hypertolerance mechanism to resist the cytotoxic levels of the accumulated metals and on a powerful scavenging mechanism for the efficient uptake of the pollutants (Salt 2006). To date, there are approximately 400 known metal hyperaccumulator plants (Eapen and D'Souza 2005). Most of them are Ni and/or Zn hyperaccumulators, whereas only a few species are known to hyperaccumulate Cd. The most common are *T. caerulescens*, *Thlaspi praecox*, *A. halleri* and *Sedum alfredii* (Van de Mortel et al. 2008). *Thlaspi* species are polymetallic hyperaccumulators known to accumulate high amounts of Zn, Cd, Ni and Pb (Mari et al. 2006), whereas *A. halleri* is able to tolerate Zn, Cd and Pb and hyperaccumulates Zn and Cd (Van Rossum et al. 2004). *S. alfredii* has been identified as a Zn hyperaccumulator, and recently it was confirmed to also hyperaccumulate Cd (Zhou and Qiu 2005).

Non-hyperaccumulator plants normally accumulate heavy metals in roots, whereas hyperaccumulator plants are able to transport most of the absorbed toxic elements to the shoots (Lasat et al. 1998). Metal translocation from root to shoot through the xylem is therefore a key determinant of the hyperaccumulation phenotype. In this respect, it has been recently demonstrated that the metal transporter HMA4 is essential for the root-to-shoot transport. Interestingly, HMA4 is expressed at higher levels in the hyperaccumulator *A. halleri*, in comparison with non-tolerant *A. thaliana* (Hanikenne et al. 2008).

At the molecular level, amino and organic acids have been proposed to play a role in heavy metal hyperaccumulation or tolerance (Sharma and Dietz 2006), although no clear mechanisms of metal long-distance trafficking related to metal hyperaccumulation have been described (Mari et al. 2006). Phytoremediation is an emerging *in situ*, cost-effective and ecological technology that exploits the ability of plants to accumulate heavy metals in their above-ground tissues to reclaim polluted environments (Alkorta et al. 2004). In this respect, hyperaccumulator plants have a direct performer role of pollutant removal or indirectly represent sources of genes for the improvement of non-hyperaccumulator plants. Ideal hyperaccumulator plants, in fact, should have some specific features such as high biomass, rapid growth rate, highly branched and extended root apparatus and easy harvesting. However, natural hyperaccumulator plants have generally low biomass and slow growth rate. This restriction may be overcome by transferring the genetic potential responsible for hyperaccumulation from hyperaccumulator species to plants with appropriate traits for phytoremediation, to confer enhanced capacity for pollutant accumulation and tolerance. Poplar has recently emerged as a model system (its genome is under sequencing) and a good candidate for phytoremediation purposes. The transgenic yellow poplar expressing a bacterial mercury reductase, developed for enhanced mercury phytoremediation is well known (Rugh et al. 1998). Moreover, Indian mustard (*B. juncea*) is also a suitable target species, because of its large biomass production, a relatively high metal accumulation and the already well-established transformation technology.

Typically, chelation of the metal ion, transport of metal or its complexes and subsequent compartmentalization in vacuoles are the processes where biotechnology can play a part in enhancing the phytoremediation capacity of plants. For example, transferring a single gene involved in metal transport, such as *HMA4*, from *A. halleri* to *A. thaliana* has increased the shoot metal uploading in this non-accumulator species (Hanikenne et al. 2008). Regarding metal-conjugates transport, plants overexpressing specific transport proteins (such as members of the CDF group, Krämer et al. 2007) might acquire higher detoxification and compartmentalization of GS-Cd conjugates into the vacuoles. Transgenic *B. juncea* plants engineered to produce more glutathione and phytochelatins accumulated significantly more Cd than wild-type plants (Bennett et al. 2003). *A. thaliana* and tobacco plants engineered with the MT gene information developed Cd tolerance and accumulation (Eapen and D'Souza 2005). Furthermore, Cd tolerance and accumulation is also enhanced by overexpressing the γ-glutamylcysteine synthetase, an enzyme with an important role in controlling glutathione synthesis and therefore metal chelation (Figure 1B) (Zhu et al. 1999). Another study revealed that the expression of the *AtPCS1* gene increased Cd and As

tolerance and accumulation in *B. juncea* (Gasic and Korban 2007) and in tobacco plants (Pomponi et al. 2006). Recently, it has been verified that a bZIP transcription factor isolated as differentially expressed in response to Cd treatment in *B. juncea* (Fusco et al. 2005), enhances Cd accumulation and tolerance in transgenic *Arabidopsis* and tobacco plants (S Farinati et al., unpubl. data, 2008). Moreover, the comparison between hyper-accumulator with non-accumulator sister species (e.g. *A. halleri* with *A. thaliana*) suggests that the hyper-accumulating features could reside in sequence mutations, gene copy number and/or in different expression levels of the proteins that contribute to the metal tolerance (Plaza et al. 2007; Hanikenne et al. 2008). These findings highlight that probably part of the genetic potential for metal detoxification is already present in most plant species and those small sequence changes that influence both metal sensing and activation of appropriate responses make the difference.

References

Alcantara E, Romera FJ, Canñete M, De La Guardia MD (1994). Effects of heavy metals on both induction and function of root Fe(III) reductase in Fe-deficient cucumber (*Cucumis sativus* L.) plants. *J. Exp. Bot.* **45**, 1893–1898.

Alkorta I, Hernández-Allica J, Becerril JM, Amezaga I, Albizu I, Garbisu C (2004). Recent findings on the phytoremediation of soils contaminated with environmentally toxic heavy metals and metalloids such as zinc, cadmium, lead, and arsenic. *Rev. Environ. Sci. Biotechnol.* **3**, 71–90.

Astolfi S, Zuchi S, Passera C (2004). Role of sulphur availability on cadmium-induced changes of nitrogen and sulphur metabolism in maize (*Zea mays* L.) leaves. *J. Plant Physiol.* **161**, 795–802.

Baker AJM, Brooks RR (1989). Terrestrial higher plants which hyper-accumulate metallic elements: a review of their distribution, ecology and phytochemistry. *Biorecovery* **1**, 81–126.

Balestrasse KB, Benavides MP, Gallego SM, Tomaro ML (2003). Effect of cadmium stress on nitrogen metabolism in nodules and roots of soybean plants. *Funct. Plant Biol.* **30**, 57–64.

Balsberg Pahlsson AM (1989). Toxicity of heavy metals (Zn, Cu, Cd, Pb) to vascular plants. *Water Air Soil Poll.* **47**, 287–319.

Becher M, Talke IN, Krall L, Krämer U (2004) Crossspecies microarray transcript profiling reveals high constitutive expression of metal homeostasis genes in shoots of the zinc hyperaccumulator *Arabidopsis halleri*. *Plant J*. **37**, 251–268.

Benavides MP, Gallego SM, Tomaro ML (2005). Cadmium toxicity in plants. *Braz. J. Plant Physiol*. **17**, 21–34.

Bennett LE, Burkhead JL, Hale KL, Terry N, Pilon M, Pilon-Smits EHA (2003). Analysis of transgenic Indian mustard plants for phytoremediation of metal-contaminated mine tailings. *J. Environ. Qual.* **32**, 432–440.

Bernard C, Roosens N, Czernic P, Lebrun M, Verbruggen N (2004). A novel CPx-ATPase from the cadmium hyperaccumulator *Thlaspi caerulescens*. *FEBS Lett*. **569**, 140–148.

Bestwick CS, Brown IR, Mansfield JW (1998). Localized changes in peroxidase activity accompany hydrogen peroxide generation during the development of a non-host hypersensitive reaction in lettuce. *Plant Physiol.* **118**, 1067–1078.

Bies-Ethève N, Gaubier-Comella P, Debures A, Lasserre E, Jobet E, Raynal M et al. (2008). Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.* **67**, 107–124.

Boussama N, Ouariti O, Suzuki A, Ghorbel MH (1999). Cd-stress on nitrogen assimilation. *J. Plant Physiol.* **155**, 310–317.

Boyer JS (1982). Plant productivity and environment. *Science* **218**, 443–448.

Buchanan BB, Gruissem W, Jones RL (2000). *Biochemistry and Molecular Biology of Plants*. American Society of Plant Physiologists, Rockville, Maryland.

Buchet JP, Lauwerys R, Roels H, Bernard A, Bruaux P, Claeys F et al. (1990). Renal effects of cadmium body burden of the general population. *Lancet* **336**, 699–702.

Cazale AC, Clemens S (2001). *Arabidopsis thaliana* expresses a second functional phytochelatin synthase. *FEBS Lett.* **507**, 215–219.

Chaffei C, Pageau K, Suzuki A, Gouia H, Ghorbel HM, Mascalaux-Daubresse C (2004). Cadmium toxicity induced changes in nitrogen management in *Lycopersicon esculentum* leading to a metabolic safeguard through an amino acid storage strategy. *Plant Cell Physiol.* **45**, 1681–1693.

Cherian GM, Chan HM (1993). Biological functions of metallothioneins: a review. In: Suzuki KT**,** Imura N**,** Kimura M, eds. *Metallothionein III*. Birkhauser, Basel. pp. 87–109.

Chiang HC, Lo JC, Yeh QC (2006). Genes associated with heavy metal tolerance and accumulation in Zn/Cd hyperaccumulator *Arabidopsis halleri*: a genomic survey with cDNA microarray. *Environ. Sci. Technol.* **40**, 6792–6798.

Clemens S, Schroeder JI, Degenkolb T (2001). *Caenorhabdites elegans* expresses a functional phytochelatin synthase. *Eur. J. Biochem.* **268**, 3640–3643.

Clemens S (2006). Evolution and function of phytochelatin synthases. *J. Plant Physiol.* **163**, 319–332.

Cobbett CS (2000). Phytochelatins and their roles in heavy metal detoxification. *Plant Physiol.* **123**, 825–832.

Cobbett CS, Goldsbrough P (2002). Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **53**, 159–182.

Cosio C, Martinoia E, Keller C (2004). Hyperaccumulation of cadmium and zinc in *Thlaspi caerulescens* and *Arabidopsis halleri* at the leaf cellular level. *Plant Physiol.* **134**, 716–725.

Curie C, Alonso JM, Le Jean M, Ecker JR, Briat JF (2000). Involvement of NRAMPI from *Arabidopsis thaliana* in iron transport. *Biochem. J.* **347**, 749–755.

Cushman JC, Bohnert H (2000). Genomic approaches to plant stress tolerance. *Curr. Opin. Plant Biol.* **3**, 117–124.

Dat JF, Vandenabeele S, Vranová E, Van Montagu M, Inzé D, Van Breusegem F (2000). Dual action of the active oxygen species during plant stress responses. *Cell Mol. Life Sci.* **57**, 779–795.

Delhaize EP, Ryan R (1995). Aluminium toxicity and tolerance in plants. *Plant Physiol.* **107**, 315–321.

Dong J, Wu F, Zhang G (2006). Influence of cadmium on antioxidant capacity and four microelement concentrations in tomato seedlings (*Lycopersicon esculentum*). *Chemosphere* **64**, 1659–1666.

Eapen S, D'Souza SF (2005). Prospects of genetic engineering of plants for phytoremediation of toxic metals. *Biotechnol. Adv.* **23**, 97–114.

Ernst WH, Krauss GJ, Verkleij JAC, Wesenberg D (2008). Interaction of heavy metals with the sulphur metabolism in angiosperms from an ecological point of view. *Plant Cell Environ.* **31**, 123–143.

Filatov V, Dowdle J, Smirnoff N, Ford-Lloyd B, Newbury HJ, Macnair MR (2006). Comparison of gene expression in segregating families identifies genes and genomic regions involved in a novel adaptation, zinc hyperaccumulation. *Mol. Ecol.* **15**, 3045–3059.

Fusco N, Micheletto L, Dal Corso G, Borgato L, Furini A (2005). Identification of cadmium-regulated genes by cDNA-AFLP in the heavy metal accumulator *Brassica juncea* L. *J. Exp. Bot.* **56**, 3017–3027.

Garcia-Hernandez M, Murphy A, Taiz L (1998). Metallothioneins 1 and 2 have distinct but overlapping expression patterns in *Arabidopsis*. *Plant Physiol.* **118**, 387–397.

Gasic K, Korban SS (2007). Transgenic Indian mustard (*Brassica juncea*) plants expressing an *Arabidopsis* phytochelatin synthase (AtPCS1) exhibit enhanced As and Cd tolerance. *Plant Mol. Biol.* **64**, 361–369.

Ghelis T, Dellis O, Jeannette E, Bardat F, Miginiac E, Sott B (2000). Abscisic acid plasmalemma perception triggers a calcium influx essential for RAB18 gene expression in *Arabidopsis thaliana* suspension cells. *FEBS Lett.* **483**, 67–70.

Gong JM, Lee DA, Schroeder JI (2003). Long-distance root-to-shoot transport of phytochelatins and cadmium in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **100**, 10118–10123.

Ha SB, Smith AP, Howden R, Dietrich WM, Bugg S, O'Connell MJ et al. (1999). Phytochelatin synthase genes from *Arabidopsis* and the yeast *Schizosaccharomyces pombe*. *Plant Cell* **11**, 1153–1164.

Hall JL (2002). Cellular mechanisms for heavy metal detoxification and tolerance. *J. Exp. Bot.* **53**, 1–11.

Hanikenne M, Talke IN, Haydon MJ, Lanz C, Nolte A, Motte P et al. (2008). Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. *Nature* **453**, 391–395.

Heiss S, Wachter A, Bogs J, Cobbett C, Rausch T (2003). Phytochelatin synthase (PCS) protein is induced in *Brassica juncea* leaves after prolonged Cd exposure. *J. Exp. Bot.* **54**, 1833–1839.

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A et al. (2007). Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. USA* **104**, 6478–6483.

Kasuka M, Liu Q, Setsuko M, Yamaguchi-Shinozaki K, Shinozaki K (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat. Biotechnol.* **17**, 287–291.

Kim DY, Bovet L, Kushnir S, Noh EW, Martinoia E, Lee Y (2006). AtATM3 is involved in heavy metal resistance in *Arabidopsis*. *Plant Physiol.* **140**, 922–932.

Kim DY, Bovet L, Maeshima M, Martinoia E, Lee Y (2007). The ABC transporter AtPDR8 is a cadmium extrusion pump conferring heavy metal resistance. *Plant J*. **50**, 207–18.

Kobae Y, Sekino T, Yoshioka H, Nakagawa T, Martinoia E, Maeshima M (2006). Loss of AtPDR8, a plasma membrane ABC transporter of *Arabidopsis thaliana*, causes hypersensitive cell death upon pathogen infection. *Plant Cell Physiol.* **47**, 309–318.

Korshunova YO, Eide D, Clark WG, Guerinot ML, Pakrasi HB (1999). The IRT1 protein from *Arabidopsis thaliana* is a metal transporter with a broad substrate range. *Plant Mol. Biol.* **40**, 37–44.

Kovalchuk I, Titov V, Hohn B, Kovalchuk O (2005). Transcriptome profiling reveals similarities and differences in plant responses to cadmium and lead. *Mutat. Res.* **570**, 149–161.

Krämer U, Talke IN, Hanikenne M (2007). Transition metal transport. *FEBS Lett.* **581**, 2263–2272.

Lane TW, Morel FM (2000). A biological function for cadmium in marine diatomeas. *Proc. Natl. Acad. Sci. USA* **97**, 4627–4631.

Lang M, Zhang Y, Chai T (2005). Identification of genes up-regulated in response to Cd exposure in *Brassica juncea* L. *Gene* **363**, 151–158.

Lasat MM, Baker A, Kochian LV (1998). Altered Zn compartmentation in the root symplasm and stimulated Zn absorption into the leaf as mechanisms involved in Zn hyperaccumulation in *Thlaspi caerulescens*. *Plant Physiol.* **118**, 875–883.

Lee S, Moon JS, Domier LL, Korban SS (2002). Molecular characterization of phytochelatin synthase expression in transgenic *Arabidopsis*. *Plant Physiol. Biochem.* **40**, 727–733.

Lee S, Moon JS, Ko TS, Petros D, Goldsbrough PB, Korban SS (2003). Overexpression of *Arabidopsis* phytochelatin synthase paradoxically leads to hypersensitivity to cadmium stress. *Plant Physiol*. **131**, 656–63.

Lin R, Wang X, Luo Y, Du W, Guo H, Yin D (2007). Effects of soil cadmium on growth, oxidative stress and antioxidant system in wheat seedlings (*Triticum aestivum* L.). *Chemosphere* **69**, 89–98.

Lòpez-Millàn AF, Ellis DR, Grusak MA (2004). Identification and characterization of several new members of the ZIP family of metal ion transporters in *Medicago truncatula*. *Plant Mol. Biol.* **54**, 583–596.

MacRobbie EA, Kurup S (2007). Signalling mechanisms in the regulation of vacuolar ion release in guard cells. *New Phytol.* **175**, 630–640.

Mari S, Gendre D, Pianelli K, Ouerdane L, Lobinski R, Briat JF et al. (2006). Root-to-shoot long-distance circulation of nicotianamine and nicotianamine-nickel chelates in the metal hyperaccumulator *Thlaspi caerulescens*. *J. Exp. Bot.* **57**, 4111–4122.

Masclaux C, Valadier M, Brugière N, Morot-Gaudry J, Hirel B (2000). Characterization of the sink/source transition in tobacco (*Nicotiana tabacum* L.) shoots in relation to nitrogen management and leaf senescence. *Planta* **211**, 510–518.

Mittler R (2006). Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* **11**,15–19.

Montanini B, Blaudez D, Jeandroz S, Sanders D, Chalot M (2007). Phylogenetic and functional analysis of the Cation Diffusion

Facilitator (CDF) family: improved signature and prediction of substrate specificity. *BMC Genomics* **23**, 107.

Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, Warner DC et al. (2007). Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yield on water-limited acres. *Proc. Natl. Acad. Sci. USA* **104**, 1645–16455.

Nevo Y, Nelson N (2006). The NRAMP family of metal-ion transporters. *Biochim. Biophys. Acta* **1763**, 609–620.

Nussbaum S, Schmutz D, Brunold C (1988). Regulation of assimilatory sulfate reduction by cadmium in *Zea mays* L. *Plant Physiol.* **88**, 1407–1410.

Osuji GO, Madu WC (1996). Ammonium ion salvage by glutamate dehydrogenase during defense response in maize. *Phytochemistry* **42**, 1491–1498.

Pan A, Yang M, Tie F, Li L, Chen Z, Ru B (1994). Expression of mouse metallothionein-I gene confers cadmium resistance in transgenic tobacco plants. *Plant Mol. Biol.* **24**, 341–51.

Papoyan A, Kochian LV (2004). Identification of *Thlaspi caerulescens* genes that may be involved in heavy metal hyperaccumulation and tolerance. Characterization of a novel heavy metal transporting ATPase. *Plant Physiol.* **136**, 3814–3823.

Papoyan A, Pineros M, Kochian LV (2007). Plant Cd^{2+} and Zn^{2+} status effects on root and shoot heavy metal accumulation in *Thlaspi caerulescens*. *New Phytol.* **175**, 51–58.

Perfus-Barbeoch L, Leonhardt N, Vavaddeur A, Forestier C (2002). Heavy metal toxicity: cadmium permeates through calcium channels and disturbs the plant water status. *Plant J.* **32**, 539–548.

Plaza S, Tearall KL, Zhao FJ, Buchner P, McGrath SP, Hawkesford MJ (2007). Expression and functional analysis of metal transporter genes in two contrasting ecotypes of the hyperaccumulator *Thlaspi caerulescens*. *J. Exp. Bot.* **58**, 1717–1728.

Polle A, Schützendübel A (2003). Heavy metal signalling in plants: linking cellular and organismic responses. In: Hirt H, Shinozaki K, eds. *Plant Responses to Abiotic Stress*. Springer-Verlag, Berlin-Heidelberg. pp. 187–215.

Pomponi M, Censi V, Di Girolamo V, De Paolis A, Sanità di Toppi L, Aromolo R et al. (2006). Overexpression of *Arabidopsis* phytochelatin synthase in tobacco plants enhances Cd^{2+} tolerance and accumulation but not translocation to the shoot. *Planta* **223**, 180–190.

Prasad MNV (1995). Cadmium toxicity and tolerance in vascular plants. *Environ. Exp. Bot.* **35**, 525–545.

Rauser WE (1999). Structure and function of metal chelators produced by plants: the case for organic acids, amino acids, phytin and metallothioneins. *Cell Biochem. Biophys.* **31**, 19–48.

Rea PA (2007). Plant ATP-binding cassette transporters. *Annu. Rev. Plant Biol.* **58**, 347–375.

Reeves RD, Baker AJM (2000). Metal-accumulating plants. In: Raskin I, Ensley BD, eds. *Phytoremediation of Toxic Metals: Using Plants to Clean Up the Environment*. John Wiley & Sons Inc., New York. pp. 193–229.

Romero-Puertas MC, Corpas FJ, Rodriguez-Serrano M, Gomez M, del Rio LA, Sandalio LM (2007). Differential expression and regulation of antioxidative enzymes by cadmium in pea plants. *J. Plant Physiol.* **164**, 1346–1357.

Romero-Puertas MC, Palma JM, Gomez LA, del Rio LA, Sandalio LM (2002). Cadmium causes oxidative modification of proteins in plants. *Plant Cell Environ.* **25**, 677–686.

Romero-Puertas MC, Rodriguez-Serrano M, Corpas FJ, Gomez M, del Rio LA, Sandalio LM (2004). Cadmium-induced subcellular accumulation of O^{2-} and H_2O_2 in pea leaves. *Plant Cell Environ.* **27**, 1122–1134.

Root RA, Miller RJ, Koeppe DE (1975). Uptake of cadmium—its toxicity and effect on the iron-to-zinc ratio in hydroponically grown corn. *J. Environ. Qual.* **4**, 473–476.

Roth U, Von Roepenack-Lahaye E, Clemens S (2006). Proteome changes in *Arabidopsis thaliana* roots upon exposure to Cd^{2+}. *J. Exp. Bot.* **57**, 4003–4013.

RoyChoudhury A, Roy C, Sengupta DN (2007). Transgenic tobacco plants overexpressing the heterologous lea gene Rab 16A from rice during high salt and water deficit display enhanced tolerance to salinity stress. *Plant Cell Rep.* **26**, 1839–1859.

Rugh CL, Senecoff JE, Meagher RB, Merkle SA (1998). Development of transgenic yellow poplar for mercury phytoremediation. *Nat. Biotech.* **16**, 925–928.

Salt DE (2006). An extreme plant lifestyle: metal hyperaccumulation. In: Taiz L, Zeiger E, eds. *Companion to Plant Physiology*. Sinauer Associates, Inc. Sunderland, MA.

Salt DE, Prince RC, Pickering IJ, Raskin I (1995). Mechanisms of cadmium mobility and accumulation in Indian mustard. *Plant Physiol.* **109**, 1427–1433.

Sanità di Toppi L, Gabbrielli R (1999). Response to cadmium in higher plants. *Environ. Exp. Bot.* **41**, 105–130.

Schat H, Llugany M, Vooijs R, Hartley-Whitaker J, Bleeker PM (2002). The role of phytochelatins in constitutive and adaptive heavy metal tolerances in hyperaccumulator and non-hyperaccumulator metallophytes. *J. Exp. Bot.* **53**, 2381–2392.

Schützendübel A, Polle A (2002). Plant responses to abiotic stresses: heavy metal-induced oxidative stress and protection by mycorrhization. *J. Exp. Bot.* **53**, 1351–1365.

Schützendübel A, Schwanz P, Teichmann T, Gross K, Langenfeld-Heyser R, Godbold DL et al. (2001). Cadmium-induced changes in antioxidative systems, hydrogen peroxide content, and differentiation in Scots pine roots. *Plant Physiol.* **75**, 887–898.

Sharma SS, Dietz KJ (2006). The significance of amino acids and amino acid-derived molecules in plant responses and adaptation to heavy metal stress. *J. Exp. Bot.* **57**, 711–726.

Sharma Y, León J, Raskin I, Davis KR (1996). Ozone-induced responses in *Arabidopsis thaliana*: the role of salicylic acid in the accumulation of defence-related transcripts and induced resistance. *Proc. Natl. Acad. Sci. USA* **93**, 5099–5104.

Shimamura C, Ohno R, Nakamura C, Takumi S (2006). Improving of freezing tolerance in tobacco plants expressing a cold-responsive and chloroplast-targeting protein WCOR15 of wheat. *J. Plant Physiol.* **163**, 213–219.

Silverberg BA (1976). Cadmium-induced ultrastructural changes in mitochondria of freshwater green algae. *Phycologia* **15**, 155–159.

Singh BK, Foley RC, Onate-Sanchez L (2002). Transcription factors in plant defense and stress responses. *Curr. Opin. Plant Biol.* **5**, 430–436.

Smirnoff N (1998). Plant resistance to environmental stress. *Curr. Opin. Biotech.* **9**, 214–219.

Stein M, Dittgen J, Sanchez-Rodriguez C, Hou BH, Molina A, Schulze Lefert P et al. (2006). *Arabidopsis* PEN3/PDR8, an ATP binding cassette transporter, contributes to nonhost resistance to inappropriate pathogens that enter by direct penetration. *Plant Cell* **18**, 731–746.

Suzuki N, Koizumi N, Sano H (2001). Screening of cadmium-responsive genes in *Arabidopsis thaliana*. *Plant Cell Environ.* **24**, 1177–1188.

Thomine S, Wang R, Ward JM, Crawford NM, Schroeder JI (2000). Cadmium and iron transport by members of a plant metal transporters family in *Arabidopsis* with homology to Nramp genes. *Proc. Natl. Acad. Sci. USA* **97**, 4991–4996.

Thordal-Christensen H, Zang Z, Wei Y, Collinge DB (1997). Subcellular localization of H_2O_2 accumulation in papillae and hypersensitive response during the barley powdery mildew interaction. *Plant J.* **11**, 1187–1194.

Thumann J, Grill E, Winnacker EL, Zenk MH (1991). Reactivation of metal-requiring apoenzymes by phytochelatin-metal complexes. *FEBS Lett.* **284**, 66–69.

Tommey AM, Shi J, Lindsay WP, Urwin PE, Robinson NJ (1991). Expression of the pea gene PsMTa in *E. coli*—metal binding properties of the expressed protein. *FEBS Lett.* **292**, 48–52.

Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K (2000). *Arabidopsis* basic leucine zipper transcription factors involved in an abscissic acid dependent signal transduction pathway under drought and high-salinity. *Proc. Natl. Acad. Sci. USA* **97**, 11632–11637.

Van de Mortel JE, Schat H, Moerland PD, Ver Loren van Themaat E, Van Der Ent S, Blankestijn H et al. (2008). Expression differences for genes involved in lignin, glutathione and sulphate metabolism in response to cadmium in *Arabidopsis thaliana* and the related Zn/Cd-hyperaccumulator *Thlaspi caerulescens*. *Plant Cell Environ.* **31**, 301–324.

Van Rossum F, Bonnin I, Fenart S, Pauwels M, Petit D, Saumitou-Laprade P (2004). Spatial genetic structure within a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy-metal-tolerant species. *Mol. Ecol.* **13**, 2959–2967.

Vatamaniuk OK, Mari S, Lu Y-P, Rea PA (2000). Mechanism of heavy metal ion activation of phytochelatin (PC) synthase-blocked thiols are sufficient for PCsynthase-catalyzed transpeptidation of glutathione and related thiol peptides. *J. Biol. Chem.* **275**, 31451–31459.

Verret F, Gravot A, Auroy P, Leonhardt N, David P, Nussaume L et al. (2004). Overexpression of AtHMA4 enhances root-to-shoot translocation of zinc and cadmium and plant metal tolerance. *FEBS Lett.* **576**, 306–312.

Weber M, Trampczynska A, Clemens S (2006). Comparative transcriptome analysis of toxic metal responses in *Arabidopsis thaliana* and the Cd-hypertolerant facultative metallophyte *Arabidopsis halleri*. *Plant Cell Environ.* **29**, 950–963.

Wei W, Zhang Y, Han L, Guan Z, Chai T (2008). A novel WRKY transcriptional factor from *Thlaspi caerulescens* negatively regulates the osmotic stress tolerance of transgenic tobacco. *Plant Cell Rep.* **27**, 795–803.

Xiao B, Huang Y, Tang N, Xiong L (2007). Over-expression of a LEA gene in rice improves drought resistance under the field conditions. *Theor. Appl. Genet.* **115**, 35–46.

Yannarelli GG, Fernàndez-Alvarez AJ, Santa-Cruz DM, Tomaro ML (2007). Glutathione reductase activity and isoforms in leaves and roots of wheat plants subjected to cadmium stress. *Phytochemistry* **68**, 505–512.

Zhigang A, Cuijie L, Yuangang Z, Yejie D, Wachter A, Gromes R et al. (2006). Expression of BjMT2, a metallothionein 2 from *Brassica juncea*, increases copper and cadmium tolerance in *Escherichia coli* and *Arabidopsis thaliana*, but inhibits root elongation in *Arabidopsis thaliana* seedlings. *J. Exp. Bot.* **57**, 3575–3582.

Zhou J, Goldsbrough PB (1994). Functional homologs of animal and fungal metallothionein genes from *Arabidopsis*. *Plant Cell* **6**, 875–884.

Zhou W, Qiu B (2005). Effects of cadmium hyperaccumulation on physiological characteristics of *Sedum alfredii* Hance (Crassulaceae). *Plant Sci.* **169**, 737–745.

Zhu YL, Pilon-Smits EHA, Tarun SA, Weber SU, Jouanin L, Terry N (1999). Cadmium tolerance and accumulation in Indian mustard is enhanced by overexpressing γ-glutamylcysteine synthetase. *Plant Physiol*. **121**, 1169–1177.

(Handling editor: Zhizhong Gong)

BMC Evolutionary Biology

BioMed Central

Research article

Evolution of *MIR168* paralogs in Brassicaceae

Silvia Gazzani[†1], Mingai Li[†1], Silvia Maistri[2], Eliana Scarponi[3], Michele Graziola[1], Enrico Barbaro[1], Jörg Wunder[1], Antonella Furini[2], Heinz Saedler[4] and Claudio Varotto*[1]

Address: [1]Environment and Natural Resources Area, Fondazione Edmund Mach, via Mach 1, 38010 San Michele all'Adige (TN), Italy, [2]Department of Science and Technology, University of Verona, strada le Grazie 15, 37134 Verona, Italy, [3]Immuno-hematology and Transfusion Service, Santa Chiara Hospital, Largo Medaglie D'Oro 9, 38100 Trento, Italy and [4]Department of Molecular Plant Genetics, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Cologne, Germany

Email: Silvia Gazzani - silvia.gazzani@iasma.it; Mingai Li - mingai.li@iasma.it; Silvia Maistri - maistri@sci.univr.it; Eliana Scarponi - elys2118@hotmail.com; Michele Graziola - micky78g@yahoo.it; Enrico Barbaro - enrico.barbaro@iasma.it; Jörg Wunder - wunder@mpiz-koeln.mpg.de; Antonella Furini - antonella.furini@univr.it; Heinz Saedler - saedler@mpiz-koeln.mpg.de; Claudio Varotto* - claudio.varotto@iasma.it

* Corresponding author †Equal contributors

Published: 23 March 2009

BMC Evolutionary Biology 2009, **9**:62 doi:10.1186/1471-2148-9-62

Received: 26 August 2008
Accepted: 23 March 2009

This article is available from: http://www.biomedcentral.com/1471-2148/9/62

Abstract

Background: In plants, expression of ARGONAUTE1 (AGO1), the catalytic subunit of the RNA-Induced Silencing Complex responsible for post-transcriptional gene silencing, is controlled through a feedback loop involving the miR168 microRNA. This complex auto-regulatory loop, composed of miR168-guided AGO1-catalyzed cleavage of *AGO1* mRNA and AGO1-mediated stabilization of miR168, was shown to ensure the maintenance of AGO1 homeostasis that is pivotal for the correct functioning of the miRNA pathway.

Results: We applied different approaches to studying the genomic organization and the structural and functional evolution of *MIR168* homologs in Brassicaeae. A whole genome comparison of Arabidopsis and poplar, phylogenetic footprinting and phylogenetic reconstruction were used to date the duplication events originating *MIR168* homologs in these genomes. While orthology was lacking between Arabidopsis and poplar *MIR168* genes, we successfully isolated orthologs of both loci present in Arabidopsis (*MIR168a* and *MIR168b*) from all the Brassicaceae species analyzed, including the basal species *Aethionema grandiflora*, thus indicating that (1) independent duplication events took place in Arabidopsis and poplar lineages and (2) the origin of *MIR168* paralogs predates both the Brassicaceae radiation and the Arabidopsis alpha polyploidization. Different phylogenetic footprints, corresponding to known functionally relevant regions (transcription starting site and double-stranded structures responsible for microRNA biogenesis and function) or for which functions could be proposed, were found to be highly conserved among *MIR168* homologs. Comparative predictions of the identified microRNAs also indicate extreme conservation of secondary structure and thermodynamic stability.

Conclusion: We used a comparative phylogenetic footprinting approach to identify the structural and functional constraints that shaped *MIR168* evolution in Brassicaceae. Although their duplication happened at least 40 million years ago, we found evidence that both *MIR168* paralogs have been maintained throughout the evolution of Brassicaceae, most likely functionally as indicated by the extremely high conservation of functionally relevant regions, predicted secondary structure and thermodynamic profile. Interestingly, the expression patterns observed in Arabidopsis indicate that *MIR168b* underwent partial subfunctionalization as determined by the experimental characterization of its expression pattern provided in this study. We found further evolutionary evidence that pre-miR168 lower stem (the RNA-duplex structure adjacent to the miR-miR* stem) is significantly longer than animal lower stems and probably plays a relevant role in multi-step miR168 biogenesis.

Background

MicroRNAs (miRNAs) are a large class of recently discovered short non-coding RNAs (19–25 nt long) involved in post-transcriptional regulation of protein-coding genes. In plants they repress gene expression by catalytic mRNA degradation on the basis of sequence homology between the microRNA itself and a target sequence. Through this function they act as major players in the regulation of a series of fundamental processes in plant growth and development, in response to biotic and abiotic stress and in the regulation of components of the plant silencing machinery itself [1-4]. In plants, RNA polymerase II produces a long primary transcript (pri-miRNA) folded in a typical stem-loop structure [5,6] that is processed by a Dicer-like RNAse III ribonuclease (DCL1), first in a shorter miRNA precursor (pre-miRNA) and then in the miRNA:miRNA* duplex [7-9]. The miRNA:miRNA* duplex is transported to the cytoplasm and the mature miRNA is incorporated in the RNA-Induced Silencing Complex (RISC) where it drives the slicer ARGONAUTE1 (AGO1) to silence the target mRNA [5,10,11].

Plant miRNAs have been found in a wide variety of species and several miRNA families are evolutionarily highly conserved, ranging from mosses and ferns to dicots [1,12-16]. The members of each miRNA family normally retain a complete or almost complete conservation of miRNA and miRNA* sequences and of the structure formed by their pairing. Generally strong conservation constraints characterize the sequences and structure of the pre-miRNA hairpin structure, whereas the conservation constraints on loop and flanking sequences are less tight [1]. This is due to the fact that in plants miRNA processing depends on pre-miRNA structure rather than on sequence and in particular on the structure of the flanking sequences (lower stem) rather than on the mature miRNA itself [17]. A detailed analysis of miR163 biogenesis has revealed that the release of the mature microRNA requires at least three DCL1 cleavage steps spaced by 21 nucleotide intervals each, starting from the base of its unusually long lower stem [9]. Similar studies in animals have shown that structural features of the lower stem are essential for cleavage of pri-miRNA by Drosha (which acts in animals as DCL1 does in plants; [18]).

In contrast to the complexity that regulatory cascades of transcription factors can reach [8,19], plant microRNAs are organized according to a simple, two-level hierarchy: only three of them, miR162, miR168 and miR403 [20], control their own expression and that of the other miRNAs by targeting specific proteins involved in the post-transcriptional gene silencing pathway. In particular, miR168 regulates the function of all miRNAs by targeting *AGO1* expression, therefore modulating its actual levels and consequently RISC activity [21,22]. *MIR168* is present in a low copy number in different plant species [23,24] and in the Arabidopsis genome two *MIR168* paralogs (*MIR168a* and *MIR168b*) are present. Only *MIR168a*, for which the primary transcript has been isolated [23], was shown to be involved in *AGO1* post-transcriptional gene silencing in Arabidopsis. A miR168a-resistant version of *AGO1* showed increased levels of *AGO1* mRNA, the over-accumulation of miR168 and developmental defects partially overlapping with those observed in *dcl1*, *hen1* and *hyl1* mutants [25]. A complex feedback loop, involving on the one hand cleavage of *AGO1* transcripts directed by miR168 and on the other hand stabilization of miR168 through AGO1 association, was shown to maintain AGO1 homeostasis which is pivotal for miRNA-mediated post-transcriptional gene silencing [26]. The overlapping expression patterns of *MIR168a* and *AGO1* and the restored development and fertility in *ago1* mutants expressing *miR168a-promoter:AGO1* fusion support this model [26].

Despite the relevance of *MIR168a* in plant development, up to now no detailed comparative study has been carried out to characterize its evolution, nor has the function of its paralog *MIR168b* been determined. In this study we applied phylogenetic footprinting to the characterization of the genomic organization, and structural and functional evolution of *MIR168* sets of orthologs in Brassicaceae. We found that, despite having originated before Brassicaceae radiation, *MIR168a* and *MIR168b* paralogs have been maintained, most likely as functional, throughout Brassicacea evolution, with *MIR168b* having undergone a partial sub-functionalization. We also provide evolutionary evidence that the lower stem in the pre-miRNA structure (the RNA-duplex structure adjacent to the miR-miR* stem) is significantly longer than lower stems in animals and propose the hypothesis that, similarly to mir163, it may play a relevant role in multi-step miR168 biogenesis.

Results

Synteny of MIR168a *and* MIR168b *loci in* A. thaliana *and* P. trichocarpa

In the genomes of both *A. thaliana* (Ath) and *P. trichocarpa* (Ptc) two *MIR168* loci have been identified, called *MIR168a* and *MIR168b*, located respectively on chromosome 4 and 5 in Arabidopsis and on linkage_group_III and scaffold_86 in poplar [8,27].

Analyses of synteny conservation were carried out by searching in poplar for the putative orthologs of the 20 Arabidopsis genes flanking *MIR168a* and *MIR168b* by screening for Reciprocal Best Matches (RBM) in BLASTP searches [28] (see Methods; Fig. 1A and Additional File 1). The queries from the former analyses were then used to identify recent segmental duplications (see Methods).

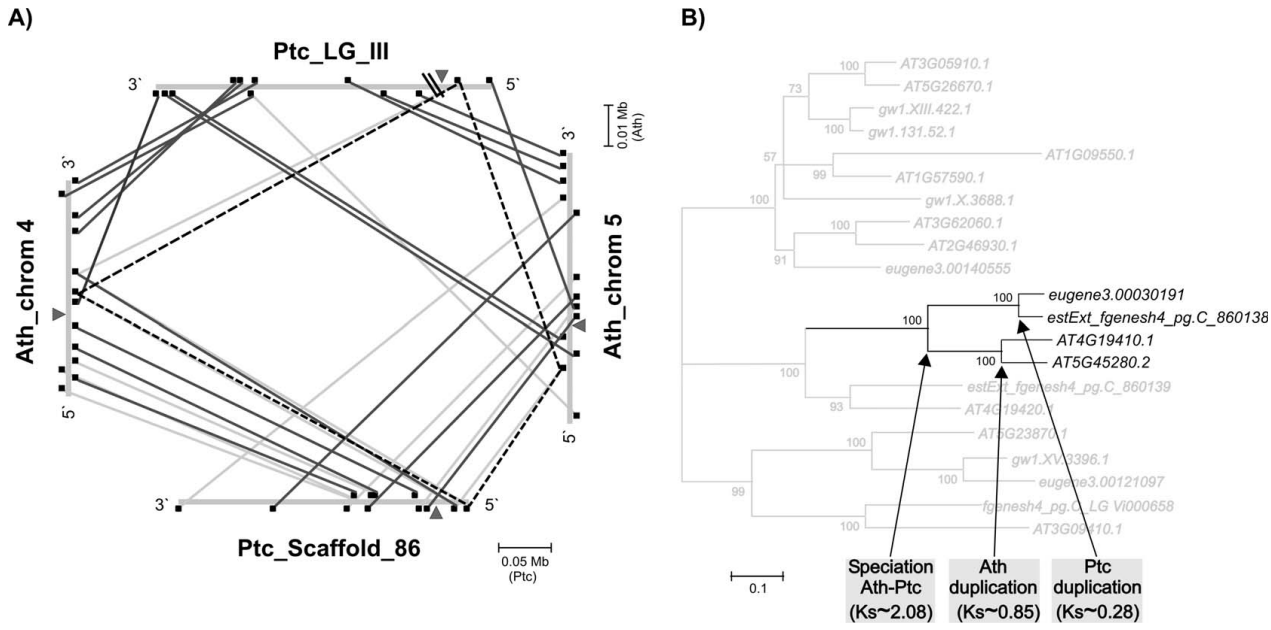


Figure 1

Synteny conservation and duplication dating of *MIR168* paralogs. A) Synteny conservation of the genomic regions encompassing *MIR168a* and *MIR168b* in *A. thaliana* and *P. trichocarpa*. Arrows represent *MIR168a* and *MIR168b*; the squares represent coding genes with at least one homolog in both genomes; the black lines represent RBMs and the gray lines connect BLASTP hits with lower homology within the same syntenic regions. Dashed lines connect *At4g19410* homologs; diagonal lines on Ptc_LG_III represent a 7 Mbp long region not syntenic to Arabidopsis. B) Phylogenetic reconstruction of *At4g19410* homologs in the Arabidopsis and poplar genomes. The portion of the linearized tree representing the homologs of *At4g19410* located in the same genomic regions as *MIR168a* and *MIR168b* is highlighted in black. Values at the branch roots correspond to majority rule consensus bootstrap values ≥ 50%. Ath: *A. thaliana*; Ptc: *P. trichocarpa*; Ks: number of synonymous nucleotide substitutions per synonymous site.

Assuming orthology among the Arabidopsis and poplar genomic regions encompassing the *MIR168* loci, the surrounding RBM pairs should be found mainly among the same pair of chromosomes. The uneven distribution of loci forming RBM pairs, however, indicated that the *MIR168* loci may have been the result of independent duplication events.

Dating of duplication events

Only two Arabidopsis paralogs formed RBM pairs in poplar (*At4g19410* and *Eugene3.00030191*; *At5g45280* and *EstExt_fgenesh4_pg.C_860138;* Fig. 1B). To determine the chronological order of these duplications, we carried out a phylogenetic reconstruction of all the genes that are homologous to the RBM pairs in the two genomes. The results show that the splitting of the two species predated two duplication events that took place independently in the Arabidopsis and poplar lineages. The two Arabidopsis paralogs, *At4g19410* and *At5g45280*, displayed a rate of synonymous substitution (Ks) of 0.85, a higher value than that observed for paralogs resulting from the Arabidopsis alpha whole genome duplication [29]. The two poplar paralogs, *Eugene3.00030191* and *Est Ext _fge nesh 4 _pg. C_860138*, were confirmed to have diverged more recently (Ks = 0.28). The divergence between poplar and Arabidopsis homologs ranged between Ks = 1.91 and Ks = 2.33. Based on the estimated divergence time between Cleomaceae and Brassicaceae (Ks = 0.82, corresponding to about 41 million years ago [29]), this should correspond to a poplar-Arabidopsis divergence time of about 105 million years, in full agreement with the 100–120 million year range provided by previous reports [30].

This dating agrees with the observation that synteny conservation between Arabidopsis and poplar is higher than between Arabidopsis chromosomes. Taken together, these results indicate that no orthologous relationship can be inferred between Arabidopsis and poplar *MIR168* homologs.

Genomic characterization of MIR168 *loci in Brassicaceae species*

On the basis of these results we focused on analysis of the evolution and conservation between species of the two

MIR168 homologs in a group of 16 Brassicaceae species (Table 1).

MIR168a and *MIR168b* homologs were amplified through a gene-to-gene amplification based on their up- and downstream genes in Arabidopsis. The intergenic region downstream of *MIR168a* was amplified from all the species with an amplification rate double than that of the upstream intergenic region (Table 1). In the case of *MIR168b* the intergenic regions were fully isolated (from the upstream to the downstream gene) in most of the species. The taxonomic distance of the single species from Arabidopsis did not significantly affect the isolation of intergenic regions.

The isolation of intergenic regions and the level of sequence conservation between species highlighted by their multiple alignments indicate: (1) general micro-synteny conservation in the regions surrounding *MIR168a* and *MIR168b* and (2) conservation of the orthologous relationship of all isolated *MIR168a* and *MIR168b* genes at the family level (Table 1).

MIR168a *and* MIR168b *phylogenetic footprinting*

A clear phylogenetic footprint was identified in all species ~100–150 bp upstream of the mature miR168a (Additional File 2A) in correspondence with Arabidopsis *MIR168a* transcription start site (TSS; GenBank accession DQ108858.1). On the contrary, the use of different alignment programs failed to identify a highly conserved footprint corresponding to *MIR168b* TSS. The location of *MIR168b* TSS in Arabidopsis was therefore determined by sequencing 21 RACE products obtained from *pMIR168b1::GFP-GUS* transgenic lines. The 5' end of all clones mapped in three points of a region ~60–110 bp upstream of the mature miR168b proximal to a TATA-like motif (consensus ATTAAATACC) conserved in both paralogs (Additional File 2B; positions 28–51). The three TSS conformed in all cases to the TA class of dinucleotides identified by the YR Rule [31]. This poorly conserved footprint could be identified by manual editing of a multiple sequence alignment performed with clustalW, thus indicating a lower functional constraint on *MIR168b* as compared to *MIR168a* transcription.

Detailed analysis of pre-miR168a and pre-miR168b and flanking sequences revealed a considerable conservation of the pre-miRNA sequences at both loci (Additional File 2C and 2D). Both miR168 and miR168* were completely or almost completely conserved between orthologs and paralogs in all species (Additional File 2C and 2D). The ~20 bp flanking regions preceding the mature miR168 and following the miR168* showed a significant level of sequence conservation between orthologs and also, although to a lower extent, between paralogs (Additional File 2E, 2F and 2G).

Table 1: Summary of *MIR168* homolog isolation from Brassicaeae

		MIR168a		*MIR168b*	
Species	**Code**	**Upstream IR (Kbp)**	**Downstream IR (Kbp)**	**Upstream IR (Kbp)**	**Downstream IR (Kbp)**
Aethionema grandiflora Boiss & Hohen. [b]	Agr	n.d.	2.0	1.3	1.3
Alyssum montanum L.	Amo	n.d.	4.5	n.d.	1.4
Arabidopsis lyrata (L.) O'Kane and Al-Shehbaz	Aly	n.d.	n.d.	n.d.	n.d.
Arabidopsis thaliana (L.) Heynh. [a,b]	Ath	2.2	2.4	0.7	2.2
Cardamine alpina Willd.	Cal	n.d.	2.0	n.d.	3.5
Cardamine flexuosa With. [b]	Cfl	n.d.	2.5	0.7	1.8
Capsella grandiflora (Fauché & Chaub.) Boiss. [b]	Cgr	3.5	3.0	0.7	1.5
Cardamine hirsuta L. [b]	Chi	n.d.	2.3	1.1	1.5
Cardamine impatiens L. [b]	Cim	n.d.	2.3	0.8	2.0
Calepina irregularis (Asso) Thell.	Cir	3.5	2.3	1.0	1.3
Diplotaxis tenuifolia (L.) DC. [b]	Dte	n.d.	1.5	0.6	0.8
Erysimum cheiri L. Crantz [a,b]	Ech	4.0	3.0	0.5	2.0
Malcolmia maritima (L.) Ait. f. [b]	Mma	4.0	3.0	0.5	1.5
Pseudoturritis turrita (L.) Al-Shehbaz [a]	Ptu	3.5	2.5	0.5	1.8
Rorippa austrica (Crantz) Spach [b]	Rau	n.d.	2.5	0.8	1.5
Thellungiella halophila (C.A. Mey.) O.E. Schulz [a,b]	Tha	3.0	2.5	0.7	2.7

Amplification of the intergenic regions upstream and downstream of *MIR168a* and *MIR168b* in Brassicaceae species. IR: intergenic region; Kbp: kilobasepair. n.d.: not determined.

[a], [b]: whole *MIR168a* or *MIR168b* loci obtained from the upstream to the downstream gene.

A completely conserved 9 bp long motif (5'-TCAGATCTG-3') was isolated in both *MIR168a* and *MIR168b* just downstream of the pre-microRNA (Additional File 2E). Despite being a palindromic structure, it was not involved in any predicted secondary structure. Searches for this motif in the Athamap database [32] showed a high quality match with the binding site of the tobacco AGP1 transcription factor [33]. No significant over-representation of the 9 bp motif downstream of microRNA loci was detected as compared with coding genes (the P-value of a two-tailed G-test for patterns with a maximum of one mismatch was p = 0.066). An identical pattern was also detected in *MIR396a* downstream of, but at a higher distance as compared with *MIR168*. To check for over-representation of this motif in specific groups of microRNAs, 94 microRNA superfamilies were defined based on the classification of their targets. The application of random permutation resampling approach led to the identification of only one superfamily which showed an enrichment in this motif (p = 0.00016, α = 0.0036 at the 0.05 level applying the Bonferroni correction with k = 14 superfamily classes tested; see Methods). This superfamily encompasses both *MIR168* paralogs and *MIR403*, a microRNA targeting *ARGONAUTE2* (*AGO2*) that is a member of the *ARGONAUTE* family of slicers responsible for mRNA cleavage in PTGS.

A footprint specific to *MIR168b* was located about 25 bp downstream of the TSS (Additional File 2B; positions 85–118). The footprint matched the binding sites of *AGAMOUS LIKE 1* (*AGL1*; AT3G58780) and *AGAMOUS LIKE 2* (*AGL2*; AT5G15800), two MADS-box domain transcription factors involved in floral organ identity and meristem determinacy [34-36]. The presence of a 14 bp insertion in the basal species *Aethionema grandiflora* prompted us to separately consider two sub-motifs (consensus TGCCAGATAT and GGTAACTGTT). Their occurrence upstream of Arabidopsis microRNAs was not significantly over-represented compared to 5'UTRs of all Arabidopsis coding genes (p = 0.64, p = 0.54, respectively). No statistical support for their preferential occurrence in the 5' region of specific microRNA superfamilies was found at the 0.05 level (data not shown).

Phylogenetic reconstruction of MIR168a *and* MIR168b

Phylogenetic reconstruction with all Brassicaceae *MIR168* homologs confirmed the successful isolation of orthologs of Arabidopsis *MIR168a* and *MIR168b*. The limited amount of parsimony-informative sites, however, could not provide a phylogenetic reconstruction resolved enough to compare the evolutionary rates of the single *MIR168* loci (data not shown). Two data partitions were created by concatenating *MIR168a* with *MIR168b* and *ITS* with *EIF3E* [37]. The resulting phylogenetic reconstructions of *MIR168* as compared with the *ITS-EIF3E* neutral markers showed slightly incongruent topologies that are the consequence of the overall lower resolution provided by the *MIR168* partition (Fig. 2).

Comparative analysis of predicted pre-miR168a and pre-miR168b structures

Secondary structures for pre-miR168a and pre-miR168b plus 50 bp of flanking sequences on each side were predicted based on free energy minimization [38,39]. The consensus of the most conserved portion of these regions, including about 20 bp upstream of mature miR168 and downstream of miR168*, is shown in Figure 3. The mature microRNA-microRNA* secondary structure (upper stem) was completely conserved in the case of *MIR168a* and almost completely conserved in the case of *MIR168b* (Fig. 3A and 3B). The structure adjacent to the upper stem (lower stem) was also highly conserved in *MIR168a* and *MIR168b*. In *MIR168a* it ranged from 18 to 19 bp, with two mismatches and one bulge loop (the two mismatches typically at positions -4 and -14, the bulge loop at position -11; Fig. 3A). The lower stem of the predicted *MIR168b* structure was 17 to 18 bp long and presented three mismatches usually at positions -4, -8 and -12 (Fig. 3B). The lower stem flanking sequences distal to the upper stem were single stranded.

Thermodynamic profiles and patterns of nucleotide substitutions

The average thermodynamic profile calculated from the predicted minimum free energy (MFE) structure of each species was nearly identical at the level of the upper stem and more variable for the lower stem of both microRNAs (Fig. 4A). A common feature of both the upper and lower stem was that the secondary structure was less stable (higher free energy value, dG) at the 5' side with an increase in stability in the central part and at the 3' side. The level of nucleotidic conservation across species, however, did not correlate with the dG values, indicating that the observed footprints could not be explained by a simple increase in the stability of the corresponding secondary structure (see e.g., *MIR168a*; Fig. 4A). On the contrary, the comparison of *MIR168a* and *MIR168b* thermodynamic profiles and the classification of their nucleotide substitutions with respect to base pairing indicated a clear positional effect concerning the lower stem: the central region was more variable than the 3–4 bp close to each end of both stems. In particular the nucleotidic stretch of 5–6 bp connecting upper and lower stems of both microRNAs (position -3, +3) were extremely conserved despite having an average free energy of -1.6 Kcal/mole, which is the average free energy of both stems.

The highest number of both structurally conservative (in yellow and ochre in Figure 4B) and non-conservative nucleotide substitutions (in blue and red in Figure 4B)

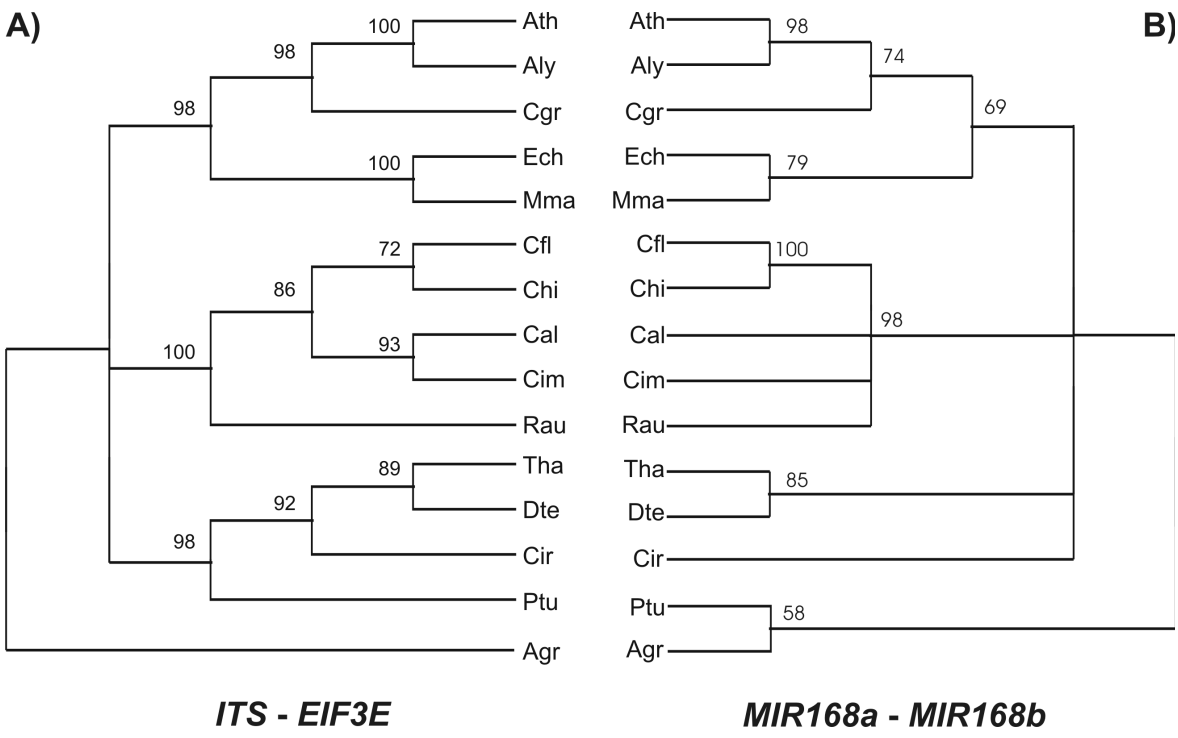


Figure 2
Phylogenetic reconstruction of *MIR168* in Brassicaeae. Phylogenetic reconstruction of *MIR168a* and *MIR168b* in the Brassicaceae family compared with a phylogenetic tree drawn using the *ITS* and *EIF3* markers. Values at the branch roots correspond to majority rule consensus bootstrap values ≥ 50%. A) *ITS-EIF3* phylogenetic tree; B) *MIR168a-MIR168b* phylogenetic tree.

was found in the central portion of *MIR168a* lower stem. This was in stark contrast with the whole upper stem and the neighboring 6 bp of the lower stem in miR168a, where no nucleotide substitutions were observed, indicating the effect of a strong purifying selection. On the contrary, an overall lower number of substitutions (mostly conservative) were spread all along the stem of *MIR168b*, with a clear depletion towards the ends of both upper and lower stems.

Expression pattern of MIR168a *and* MIR168b

The high conservation of *MIR168b* suggests that it could be expressed and functional, even though, up to now, no experimental evidence has been reported. The Arabidopsis intergenic region upstream of the mature miR168b is only approximately 500 bp long. Therefore, we used two genomic regions including the whole intergenic region plus 255 or 1038 bp upstream to functionally characterize the *MIR168b* promoter and ascertain if some regulatory elements may be present in the upstream gene. These two regions were used to drive the expression of a reporter *eGFP-uidA* fusion gene (*pMIR168b1::GFP-GUS* and *pMIR168b2::GFP-GUS*; Fig. 5B. See Methods) in stably transformed Arabidopsis transgenic lines. A construct encompassing the *MIR168a* promoter was used as a control (Fig. 5A). Both *pMIR168b1::GFP-GUS* and *pMIR168b2::GFP-GUS* constructs produced the same expression pattern (data not shown). This result indicates that the intergenic region used in the shortest construct contains all the regulatory information to drive *MIR168b* expression. Similarly to what was observed for *MIR168a*, the expression of *MIR168b* was localized in emerging leaves and in a region underneath the shoot apical meristem corresponding to leaf primordia (Fig. 5C). None of the *MIR168b* transgenic lines, in contrast to *MIR168a*, displayed expression in correspondence with vascular tissues.

Discussion

Since the first reports about the presence of microRNAs in plants [8] a number of miRNA families have been identified. While attention has been devoted mostly to their dis-

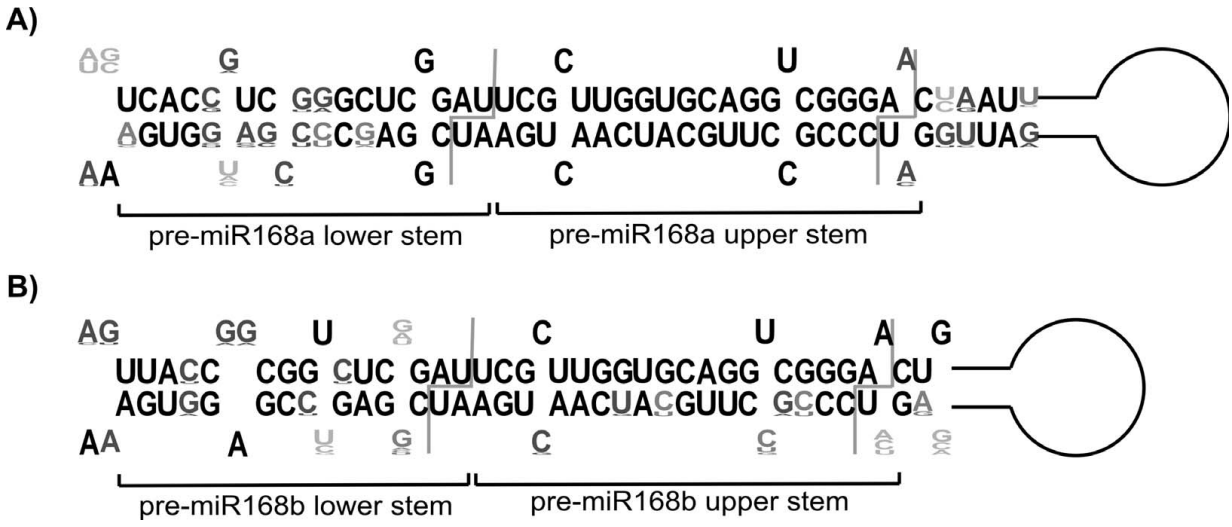


Figure 3
Stem-loop structure and conservation of the *pre-miR168* homologs. LOGO representation of the stem-loop structure of the pre-miR168 homologs in Brassicaceae species. The base composition is indicated at each position. Gray lines correspond to the pre-microRNA processing sites. A) pre-miR168a; B) pre-miR168b.

covery, both *in silico* and experimentally, relatively little is as yet known about plant microRNA evolution and biogenesis. In this study we applied a phylogenetic footprinting approach to the comparative study of the evolutionary patterns of two paralogous microRNA loci, *MIR168a* and *MIR168b*, in the Brassicaceae family. The presence of highly conserved phylogenetic footprints, in fact, is an indication of selective constraints acting on specific sequences [40]. If, as in the case of *MIR168*, the divergence time among genes can be demonstrated to be sufficiently high, parallel phylogenetic footprinting of paralogs provides a powerful tool to yield evolutionary evidence for the functionality of a locus as a whole or of its parts.

Evolution of MIR168 *in Brassicaceae*

Based on the analysis of synteny conservation and on the phylogenetic reconstruction of a set of closely linked homologs, we dated the origin of Arabidopsis *MIR168* paralogs to shortly before the divergence between the sister families Brassicaceae and Cleomaceae, about 41 million years ago [29]. Applying a genome walking method based on microsyntenic conservation, we were able to ascertain reliably the presence of and isolate both *MIR168a* and *MIR168b* paralogs in all analyzed species. The successful isolation of both *MIR168a* and *MIR168b* from the most basal crucifer, *Aethionema grandiflora*, provides demonstration that the origin of *MIR168* paralogs predates both Arabidopsis alpha polyploidization, which took place approximately 34 million years ago (Mya) [41,42], and Brassicaceae radiation which took place between 40 and 50 Mya [43]. The limited synteny conservation observed in Arabidopsis further suggests that the *MIR168a* and *MIR168b* paralogs escaped the extensive diploidization resulting in the maintenance of only one homeolog per locus in the surrounding regions.

Similarly to *MIR319a* [44], we identified phylogenetic footprints that corresponded to functionally relevant regions, such as the TSS and the mature miR and miR* sequences, that indicate a functional conservation of both *MIR168a* and *MIR168b* throughout the Brassicacea family. Additionally, in the present study a novel 9 nt highly conserved region has been identified immediately downstream of the lower stem. The palindromic structure of this phylogenetic footprint and its pattern nearly perfectly matched the consensus-binding site of *APG1*, the tobacco putative ortholog of *A. thaliana BME3*. This would suggest its function as a homodimeric transcription factor binding site [45]. The functional complementation with *MIR168a* promoter, however, indicates that this motif is not necessary for normal *MIR168* expression [26]. It may, instead, have a functional relevance for RNA processing or stability even while not being involved in any of the predicted pre-miR168 secondary structures. The lack of a significant over-representation downstream of other microRNA gene families in Arabidopsis indicates that this motif is not involved in a general mechanism of microRNA biogenesis or regulation. However, the occurrence of the same motif dowstream of *MIR403*, a microRNA predicted to target *AGO2* (another member of the *AGO* family) raises the

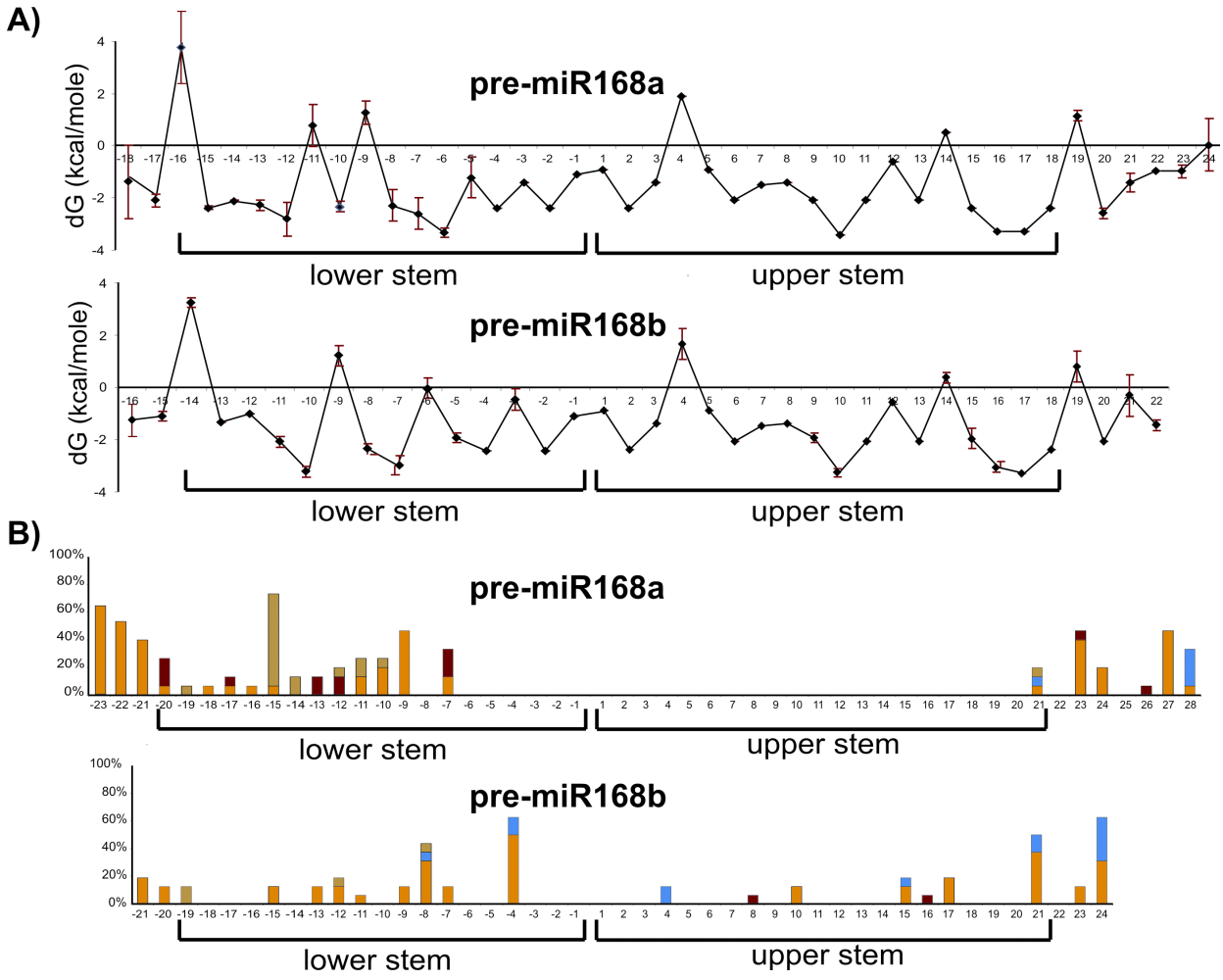


Figure 4

Thermodynamic stability and nucleotide substitution profiles of pre-miR168a and pre-miR168b. A) Thermodynamic stability profile of pre-miR168a and pre-miR168b in the Brassicaceae family. Free energy values are given in kcal/mole. Vertical bars: between-species variability calculated as double standard error. B) Distribution of nucleotide substitutions with respect to base pairing in the pre-miR168a and pre-miR168b secondary structures. Yellow: structurally conservative base substitution; ochre: base substitution comporting a change in length of a bulge loop; blue: base substitution comporting a change from unpaired to paired bases; red: base substitution comporting a change from paired to unpaired bases. The rate of nucleotide substitution is given in percentages.

interesting possibility that it may be specifically involved in the regulation of *AGO* genes by microRNAs. Further studies are, therefore, required to clarify the functional relevance of this phylogenetic footprint.

Role of the lower stem in miR168 biogenesis

Based on the combination of phylogenetic footprinting and secondary structure predictions, the only secondary structures conserved in *MIR168* during the approximately 40 million years of Brassicaceae evolution were the stem containing the miR-miR* pairing (upper stem) and its distal extension (lower stem). Recently it has been demonstrated that correct animal pri-miRNA processing depends on the length of the lower stem [18]. In agreement with this, our results indicate that the lower stem is particularly conserved in *MIR168*, with the difference that the phylogenetic footprint identified in plants (ranging from 17 to 19 base pairs) is significantly longer than the 11 base pair lower stem reported for animals [18]. In animals, the Drosha-Pasha (Microprocessor) complex required for pre-miR processing is responsible for conversion of pri-miRNA to pre-miRNA [46]. In plants, this function is car-

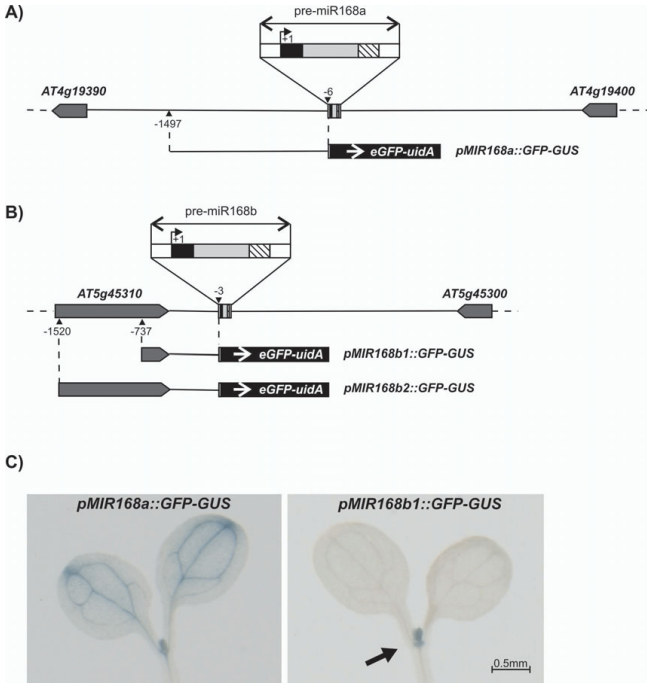


Figure 5

Expression pattern of *MIR168* paralogs in Arabidopsis. A) Genomic region encompassing *MIR168a*; B) genomic region encompassing *MIR168b*. Black box: mature miR168; dashed box: miR168*; white boxes: 20 bp sequences forming the basal stem; light gray box: miR168 loop region; dark gray boxes: nearest exons in the genes upstream and downstream of *MIR168*, arrows indicate gene orientation. Distances are drawn to scale, with the exception of pre-miR168 (to a larger scale for clarity); +1 is the first nucleotide of the mature miR168. The *pMIR168a::GFP-GUS*, *pMIR168b1::GFP-GUS* and *pMIR168b2::GFP-GUS* constructs are represented underneath the genomic regions. C) GUS-staining of Arabidopsis transformant lines carrying the *pMIR168a::GFP-GUS* and *pMIR168b1::GFP-GUS* constructs.

ried out by a functionally analogous complex involving DCL1, HYL1 and SE [47]. The observed difference in length of the lower stems may, therefore, indicate a general difference in the mechanisms of miRNA biogenesis in plants and animals.

The phylogenetic footprints identified in this study are consistent with two step pri-miRNA processing analogous to that described for *MIR163* in Arabidopsis [9]. The recent origin of *MIR163* and the extensive base complementarity of its inverted repeats [48] may indicate that the multi-step processing of this microRNA could be more an exception than the rule. Our finding that a clearly detectable selective pressure has been acting on *MIR168* lower stem throughout Brassicaceae radiation indicates that multi-step pri-miRNA processing is not peculiar to *MIR163* or to newly formed microRNAs. HYL1 has been recently shown to interact with DCL1 for the correct processing of *MIR163*. Assuming a common processing mechanism, it is possible that the highly conserved regions we identified in *MIR168* at the ends of both lower and upper stems may be the footprints of the DCL1/HYL1 complex [49]. The phylogenetic reconstruction carried out on concatenated *MIR168a* and *MIR168b* sequences indicates that *MIR168* evolution did not depart from that of the analyzed species. Interestingly, however, while a large difference in purifying selection is evident in *MIR168a*, the distribution of nucleotide substitutions turns out to be much more uniform in the case of *MIR168b*, as also reflected by their thermodynamic profiles. This may indicate that the lower stem has a function in fine-tuning the *pri-MIR168* precursor processing efficiency.

Function of MIR168 *paralogs in Arabidopsis*

The high conservation of *MIR168a* and *MIR168b* sequences, RNA predicted secondary structures and thermodynamic profiles observed in all the species we analyzed indicates that constant selective pressure has been acting on both loci throughout the Brassicaceae evolution. Interestingly, these results point to the fact that *MIR168b* has most likely been functionally conserved in all of the tested species. Former attempts to confirm *MIR168b* expression by RACE were not successful, possibly due to tissue specific expression [23,50]. In contrast to the extreme conservation observed in both *MIR168a* and *MIR319a* [44], *MIR168b* TSS identified in *A. thaliana* by RACE mapped to a phylogenetic footprint only partly conserved in the examined species, thus leaving open the possibility that the second footprint identified may function as a primary or alternative TSS in other species. This lower conservation indicates a lower selective pressure acting on the expression of *MIR168b* as compared with *MIR168a*, consistent with an accessory function of this locus [25]. However, the clear staining we observed in *A. thaliana* transformed with a *uidA* reporter gene driven by the whole intergenic regions of *MIR168b* and part of its upstream gene confirms *MIR168b* expression. Taken together, these results and the presence in the *MIR168b* stem-loop structure of the sequence information necessary for processing the mature microRNA [50], provide evidence for the functionality of this locus.

The similar but more circumscribed expression pattern of *MIR168b* as compared with *MIR168a* is consistent with either neo- or sub-functionalization of duplicated genes previously reported for other microRNA loci [51]-[52]. In light of the nearly overlapping expression patterns of *MIR168a* and *AGO1* [26], the difference in expression in the leaf vasculature observed between *MIR168* paralogs is most likely due to sub-functionalization of *MIR168b* than to neo-functionalization of *MIR168a*.

Conclusion

Phylogenetic footprinting is a powerful technique for the identification of regions that, being functionally relevant, have been maintained under selective constraint during evolution [53]. We used a comparative phylogenetic footprinting approach to identify the structural and functional constraints that shaped the evolution of *MIR168* paralogs in Brassicaceae. Previous studies in Arabidopsis demonstrated the functionality of *MIR168a* [25], but left open the possibility that *MIR168b* may be either non-functional or functionally redundant with respect to its paralog. Although their duplication happened at least 40 million years ago, we found evidence that both *MIR168* paralogs have been maintained throughout Brassicaceae evolution. The extremely high conservation of regions functionally relevant for microRNA expression and biogenesis, predicted secondary structure and thermodynamic profile also provide evolutionary evidence of functionality of both loci, as further supported by the expression of *MIR168b* in Arabidopsis. Interestingly, the expression pattern of *MIR168b* indicates partial sub-functionalization based on the expression patterns of both *MIR168a* and *AGO1*. The identification of a highly conserved *MIR168b*-specific footprint downstream of the TSS matching the binding sites for the *AGL1* and *AGL2* transcription factors [34,35]-[36], provides the indication for a first candidate motif possibly involved in the regulation of *MIR168b* at specific developmental stages.

The phylogenetic footprinting carried out on the *MIR168* paralogs finally points to the fact that the *MIR168* lower stem (the RNA-duplex structure adjacent to the miR-miR* stem) is significantly longer than animal lower stems and possibly indicates a multistep miR168 biogenesis process analogous to the one for miR163 maturation.

The application of phylogenetic footprinting to more microRNA and plant families holds the promise of furthering our understanding of the regulation of biogenesis, the function and evolution of these intriguing regulators of both animal and plant gene expression. The design of artificial microRNAs [54,55] and its application to both basic and applied research may also greatly benefit from a more detailed identification of the determinants for efficient miRNA biogenesis.

Methods

Plant material

Brassicaceae species for tissue collection were grown in the greenhouse from seeds collected in Trentino Alto Adige (Italy) from wild populations or purchased from Chiltern Seeds (Bortree Stile, Ulverston, Cumbria, LA12 7PB, England. Table 1).

Genomic isolation of MIR168 *loci in Brassicaceae species*

Genomic DNA was extracted from leaves using the CTAB method [56]. Intergenic regions encompassing *MIR168a* and *MIR168b* were obtained through gene to gene amplification by Long-Range PCR using Advantage® 2 Polymerase Mix (Clontech; Fig. 5A and 5B). Primers were designed either on conserved regions of the *A. thaliana* genes upstream and downstream of *MIR168a* and *MIR168b* or on the highly conserved sequences of the mature miR168 and miR168* (Additional File 3). For species where no PCR amplification was obtained, additional primers were designed on conserved sequences in the intergenic regions amplified from the other Brassicaceae species.

Amplification products were cloned in pGEM-T (Promega) or in pCR-XL-TOPO (Invitrogen) vectors. At least three clones corresponding to each product were sequenced bi-directionally to confirm their identity. *Arabidopsis lyrata* sequences were assembled from the NCBI Trace Archives http://www.ncbi.nlm.nih.gov/Traces/. GenBank accession numbers corresponding to the sequences used in this study are provided in Additional File 4. Multiple sequence alignments were performed with M-Coffee [57] and manually edited in Bioedit [58]. Additional alignments performed with Mulan [59] were used to identify the most conserved phylogenetic footprints by using a sliding window of 5 bp and a similarity cutoff of 90%. The TSS of *MIR168b* could not be detected by means of Mulan. The results of the RACE experiments (see below) were in this case used to identify the homologous regions from the different species and the corresponding phylogenetic footprint was obtained by manual editing of multiple sequence alignments performed with ClustalW [60].

Analysis of synteny conservation in poplar

The aminoacidic sequences corresponding to 20 Arabidopsis genes surrounding *MIR168a* and *MIR168b* (10 upstream and 10 downstream) were used for local BLASTP searches with an *e*-value cutoff of 1E-5 against the *Populus trichocarpa* genome annotation v1.1 (DoE Joint Genome Institute and Poplar Genome Consortium, http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.download.html. All poplar peptide homologs were used for a second BLASTP search against the Arabidopsis genome annotation v5.0 (TIGR, ftp://ftp.tigr.org/pub/data/a_thaliana/ath1/SEQUENCES/). Reciprocal Best Matches (RBM, [28]) were obtained as the gene pairs with the highest E-value scores in the two analyses. To detect recent segmental duplications, an additional BLASTP search was run against a joint database containing all Arabidopsis and poplar genes using all the queries from the former analyses. The hits in the genomic regions of interest were con-

sidered if their score was better than that of any other gene from the species used as query.

Phylogenetic reconstruction

Fast evolving nuclear loci (*ITS* [61] and *EIF3E* [37]) were used for phylogenetic reconstruction of the species used in this study. Primers used are listed in Additional File 3.

Multiple sequence alignments for the single genes obtained with M-Coffee [57] were manually refined using BioEdit [58]. PAUP* vers. 4.0 b10 [62] was used for phylogenetic analysis and tree-building using maximum likelihood (ML) with best substitution determined by Modeltest 3.7 [63]. Trees were calculated with swap = TBR, addition = random, hsearch replicates = 1000, trees hold at each step = 1, collapse = MaxBrLen, gaps were treated as missing. Bootstrapping was carried out with 100 re-sampling replicates, each performed with 100 heuristic search replicates. Phylogenetic reconstructions were carried out first on the single data partitions to assess the level of polymorphism and data congruence. Due to the low level of polymorphism in the single datasets, the partitions used for the final analyses were: 1) *ITS* + *EIF3E*, 2) *MIR168a* + *MIR168b*.

Phylogenetic reconstruction for the *At4g19410* peptide homologs present in both Arabidopsis and poplar genomes was carried out with Mega 4.0 [64], using the neighbor-joining method with a variable rate among aminoacidic sites (Gamma = 1.0) and 1000 bootstrap replicates. The cladogram representing the 50% majority-rule consensus tree was used to depict the lineage divergence and duplication events. Rates of synonymous substitution (Ks) were calculated with DnaSP v4.0 [65].

A. thaliana *whole genome motif search*

To analyze the representation of the conserved TCAGATCTG motif and of the *MIR168b*-specific footprint, the average length of the 24016 *A. thaliana* 3' and 22998 5' untranslated regions (UTRs) TAIR7 blastset was calculated (233 bp and 146 bp, respectively; ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/TAIR7_blastsets/). A second dataset (miRNA dataset) was obtained by extracting from the TIGR v5.0 pseudochromosomes the 233 bases downstream or the 146 bases upstream of the 184 Arabidopsis microRNA hairpins annotated in miRBase v.10.1 [66]. The presence of the TCAGATCTG motif (with a stringency of 1 mismatch) or of the two *MIR168b*-specific sub-motifs identified by the point of a 14 bp insertion in *Aethionema grandiflora* (stringency of 2 mismatches; Additional File 2B) in the miRNA and the TAIR7 3' and 5' UTR datasets was calculated with the EMBOSS fuzznuc application. A two-tailed G-test was used to test the goodness of fit for the distribution in the miRNA dataset compared with the distribution obtained from the whole genome TAIR7 UTR datasets. To check for over-representation of these motifs in specific groups of microRNAs, 94 microRNA superfamilies were defined based on classification of their targets. The number of microRNAs in each family with an occurrence of the motifs in the 233 bases downstream or the 146 bases upstream of the pre-microRNA (from now on indicated for brevity as a "hit") was further used to define 14 classes of superfamilies charcaterized by the same number of members and the same number of hits. A random permutation resampling approach was used to model the probability of each superfamily class to originate by chance in the whole complement of Arabidopsis microRNAs: a set of 1000000 random boolean strings, each 184 characters long and containing a number of "1" corresponding to the number of microRNA genes with at least one occurrence of each motif, were generated with the Mersenne Twister algorithm [67]. The probability of random occurrence of each superfamily class was given by the frequency of boolean strings matching exactly the number of hits for that class in a number of randomly selected positions corresponding to the number of its members. A Bonferroni correction was applied to keep into account multiple testing of classes.

The analysis of similarity of the conserved footprints to known binding sites was carried out by means of AthaMap database [32] and of the MultiTF program [68].

Secondary structure prediction and thermodynamic profiles

The predicted secondary structures were generated using the RNAstructure program [38]. The LOGO representation of these structures was obtained with the WebLogo software [69].

The species-specific thermodynamic stability profiles of the predicted secondary structures were calculated for pre-miR168a or pre-miR168b according to the nearest neighbour method [39], and summarized in a single profile by averaging the free energy values at each position.

Expression analysis of **MIR168a** *and* **MIR168b**

The intergenic regions upstream of miR168a and miR168b were used to drive the expression of an enhanced green fluorescent protein-beta glucuronidase (*eGFP-uidA*) fusion reporter construct (pKGWFS7; [70]). The *MIR168a* promoter region encompassed 1491 bp from -1497 to -6 upstream of the mature microRNA (Fig. 5A). For *MIR168b* two regions upstream of the mature miR, from -1520 to -3 and from -737 to -3 (including 255 and 1038 bp of the upstream gene coding sequences, respectively) were used to prepare two constructs (*pMIR168b1::GFP-GUS* and *pMIR168b2::GFP-GUS*; Fig. 5B). 4-week-old Arabidopsis plants were transformed by

floral dip [71]. 15 transformed plants from each of 13 T2 lines were selected on MS medium and subjected to GUS staining [72]. Mapping of *MIR168b* TSS was carried out with the GeneRacer™ Kit (Invitrogen). Gene-specific primers are listed in Additional File 3.

Authors' contributions

SG carried out homolog isolation, thermodynamic profile calculation and drafted the manuscript. ML isolated phylogenetic markers and participated in data analysis, made expression constructs and participated in manuscript drafting. SM and AF carried out plant transformation. ES, MG and EB participated in homolog isolation and sequencing. JW helped in phylogenetic reconstruction. HS participated in the design of the study and manuscript drafting. CV conceived and coordinated the study, took part in data analysis and drafted the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Arabidopsis-poplar* MIR168 *syntenic information. *Arabidopsis and poplar homologous gene pairs and BLASTP RBM pairs present in the* MIR168 *syntenic regions.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-62-S1.pdf]

Additional file 2

Sequence alignments. *Alignment of regions containing the predicted* MIR168a *and* MIR168b *TSS, of pre-miR168a and pre-miR168b and of the region containing the conserved 9 bp motif.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-62-S2.pdf]

Additional file 3

Primers. *List of primers used for amplification of* MIR168 *homologs from Brassicaceae.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-62-S3.pdf]

Additional file 4

GenBank accession numbers. *List of GenBank accession numbers corresponding to the sequences obtained during this study or downloaded from public databases.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-62-S4.pdf]

Acknowledgements

The authors are thankful to DH Mathews for sharing an advanced update of the RNAstructure software (4.4) before official release and to O. Savolainen, F. Salamini and A. Si-Ammour for comments on the manuscript. The authors wish to thank R. Zorer for help with GUS-staining pictures, and L. Martinelli for space in the growth room. We also wish to thank B. Neuffer and J.K. Zhu for kindly providing *B. grandiflora* and *T. halophila* seeds and F. Prosser for information about the location of natural populations of Brassicaeae. We wish to thank two anonymous Referees which significantly contributed to improve the manuscript through their insightful observations.

References

1. Jones-Rhoades MW, Bartel DP, Bartel B: **MicroRNAS and their regulatory roles in plants.** *Annu Rev Plant Biol* 2006, **57:**19-53.
2. Mallory AC, Vaucheret H: **Functions of microRNAs and related small RNAs in plants.** *Nat Genet* 2006, **38(Suppl):**S31-36.
3. Phillips JR, Dalmay T, Bartels D: **The role of small RNAs in abiotic stress.** *FEBS Lett* 2007, **581(19):**3592-3597.
4. Sunkar R, Chinnusamy V, Zhu J, Zhu JK: **Small RNAs as big players in plant abiotic stress responses and nutrient deprivation.** *Trends Plant Sci* 2007, **12(7):**301-309.
5. Chen X: **MicroRNA biogenesis and function in plants.** *FEBS Lett* 2005, **579(26):**5923-5931.
6. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN: **MicroRNA genes are transcribed by RNA polymerase II.** *Embo J* 2004, **23(20):**4051-4060.
7. Park W, Li J, Song R, Messing J, Chen X: **CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana.** *Curr Biol* 2002, **12(17):**1484-1495.
8. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP: **MicroRNAs in plants.** *Genes Dev* 2002, **16(13):**1616-1626.
9. Kurihara Y, Watanabe Y: **Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions.** *Proc Natl Acad Sci USA* 2004, **101(34):**12753-12758.
10. Baumberger N, Baulcombe DC: **Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs.** *Proc Natl Acad Sci USA* 2005, **102(33):**11928-11933.
11. Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS: **Nuclear processing and export of microRNAs in Arabidopsis.** *Proc Natl Acad Sci USA* 2005, **102(10):**3691-3696.
12. Floyd SK, Bowman JL: **Gene regulation: ancient microRNA target sequences in plants.** *Nature* 2004, **428(6982):**485-486.
13. Arazi T, Talmor-Neiman M, Stav R, Riese M, Huijser P, Baulcombe DC: **Cloning and characterization of micro-RNAs from moss.** *Plant J* 2005, **43(6):**837-848.
14. Axtell MJ, Bartel DP: **Antiquity of microRNAs and their targets in land plants.** *Plant Cell* 2005, **17(6):**1658-1673.
15. Fattash I, Voss B, Reski R, Hess WR, Frank W: **Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution.** *BMC Plant Biol* 2007, **7:**13.
16. Zhang B, Pan X, Cobb GP, Anderson TA: **Plant microRNA: a small regulatory molecule with big impact.** *Dev Biol* 2006, **289(1):**3-16.
17. Parizotto EA, Dunoyer P, Rahm N, Himber C, Voinnet O: **In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA.** *Genes Dev* 2004, **18(18):**2237-2242.
18. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN: **Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex.** *Cell* 2006, **125(5):**887-901.
19. Yamaguchi-Shinozaki K, Shinozaki K: **Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses.** *Annu Rev Plant Biol* 2006, **57:**781-803.
20. Ehrenreich IM, Purugganan MD: **Sequence variation of MicroRNAs and their binding sites in Arabidopsis.** *Plant Physiol* 2008, **146(4):**1974-1982.
21. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP: **Prediction of plant microRNA targets.** *Cell* 2002, **110(4):**513-520.
22. Xie Z, Kasschau KD, Carrington JC: **Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation.** *Curr Biol* 2003, **13(9):**784-789.
23. Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, Carrington JC: **Expression of Arabidopsis MIRNA genes.** *Plant Physiol* 2005, **138(4):**2145-2154.

24. Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA: **Conservation and divergence of plant microRNA genes.** *Plant J* 2006, **46(2):**243-259.
25. Vaucheret H, Vazquez F, Crete P, Bartel DP: **The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development.** *Genes Dev* 2004, **18(10):**1187-1197.
26. Vaucheret H, Mallory AC, Bartel DP: **AGO1 homeostasis entails coexpression of MIR168 and AGO1 and preferential stabilization of miR168 by AGO1.** *Mol Cell* 2006, **22(1):**129-136.
27. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, *et al.*: **The genome of black cottonwood, Populus trichocarpa (Torr. & Gray).** *Science* 2006, **313(5793):**1596-1604.
28. Li L, Stoeckert CJ Jr, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes.** *Genome Res* 2003, **13(9):**2178-2189.
29. Schranz ME, Mitchell-Olds T: **Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae.** *Plant Cell* 2006, **18(5):**1152-1165.
30. Wikstrom N, Savolainen V, Chase MW: **Evolution of the angiosperms: calibrating the family tree.** *Proc Biol Sci* 2001, **268(1482):**2211-2220.
31. Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, Seki M, Shinozaki K, Abe T: **Identification of plant promoter constituents by analysis of local distribution of short sequences.** *BMC Genomics* 2007, **8:**67.
32. Steffens NO, Galuschka C, Schindler M, Bulow L, Hehl R: **AthaMap: an online resource for in silico transcription factor binding sites in the Arabidopsis thaliana genome.** *Nucleic Acids Res* 2004:D368-372.
33. Sugimoto K, Takeda S, Hirochika H: **Transcriptional activation mediated by binding of a plant GATA-type zinc finger protein AGP1 to the AG-motif (AGATCCAA) of the wound-inducible Myb gene NtMyb2.** *Plant J* 2003, **36(4):**550-564.
34. Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF: **B and C floral organ identity functions require SEPALLATA MADS-box genes.** *Nature* 2000, **405(6783):**200-203.
35. Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF: **Assessing the redundancy of MADS-box genes during carpel and ovule development.** *Nature* 2003, **424(6944):**85-88.
36. Huang H, Tudor M, Su T, Zhang Y, Hu Y, Ma H: **DNA binding properties of two Arabidopsis MADS domain proteins: binding consensus and dimer formation.** *Plant Cell* 1996, **8(1):**81-94.
37. Li M, Wunder J, Bissoli G, Scarponi E, Gazzani S, Barbaro E, Saedler H, Varotto C: **Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species.** *Cladistics* 2008, **24(5):**727-745.
38. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH: **Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure.** *Proc Natl Acad Sci USA* 2004, **101(19):**7287-7292.
39. Mathews DH, Sabina J, Zuker M, Turner DH: **Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.** *J Mol Biol* 1999, **288(5):**911-940.
40. Duret L, Bucher P: **Searching for regulatory elements in human noncoding sequences.** *Curr Opin Struct Biol* 1997, **7(3):**399-406.
41. Blanc G, Hokamp K, Wolfe KH: **A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome.** *Genome Res* 2003, **13(2):**137-144.
42. Bowers JE, Chapman BA, Rong J, Paterson AH: **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events.** *Nature* 2003, **422(6930):**433-438.
43. Bonnet E, Wuyts J, Rouze P, Peer Y Van de: **Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences.** *Bioinformatics* 2004, **20(17):**2911-2917.
44. Warthmann N, Das S, Lanz C, Weigel D: **Comparative analysis of the MIR319a microRNA locus in Arabidopsis and related Brassicaceae.** *Mol Biol Evol* 2008, **25(5):**892-902.
45. Liu PP, Koizuka N, Martin RC, Nonogaki H: **The BME3 (Blue Micropylar End 3) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination.** *Plant J* 2005, **44(6):**960-971.
46. Yeom KH, Lee Y, Han J, Suh MR, Kim VN: **Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing.** *Nucleic Acids Res* 2006, **34(16):**4622-4629.
47. Song L, Han MH, Lesicka J, Fedoroff N: **Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body.** *Proc Natl Acad Sci USA* 2007, **104(13):**5437-5442.
48. Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC: **Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana.** *Nat Genet* 2004, **36(12):**1282-1290.
49. Kurihara Y, Takashi Y, Watanabe Y: **The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis.** *Rna* 2006, **12(2):**206-212.
50. Jiang D, Yin C, Yu A, Zhou X, Liang W, Yuan Z, Xu Y, Yu Q, Wen T, Zhang D: **Duplication and expression analysis of multicopy miRNA gene family members in Arabidopsis and rice.** *Cell Res* 2006, **16(5):**507-518.
51. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, *et al.*: **High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes.** *PLoS ONE* 2007, **2(2):**e219.
52. Maher C, Stein L, Ware D: **Evolution of Arabidopsis microRNA families through duplication events.** *Genome Research* 2007, **16(4):**510-519.
53. Gumucio DL, Shelton DA, Zhu W, Millinoff D, Gray T, Bock JH, Slightom JL, Goodman M: **Evolutionary strategies for the elucidation of cis and trans factors that regulate the developmental switching programs of the beta-like globin genes.** *Mol Phylogenet Evol* 1996, **5(1):**18-32.
54. Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D: **Specific effects of microRNAs on the plant transcriptome.** *Dev Cell* 2005, **8(4):**517-527.
55. Alvarez JP, Pekker I, Goldshmidt A, Blum E, Amsellem Z, Eshed Y: **Endogenous and synthetic microRNAs stimulate simultaneous, efficient, and localized regulation of multiple targets in diverse species.** *Plant Cell* 2006, **18(5):**1134-1151.
56. Doyle JJ, Doyle JL: **A rapid DNA isolation procedure for small quantities of fresh leaf tissue.** *Phytochemistry Bulletin* 1987, **19:**11-15.
57. Moretti S, Armougom F, Wallace IM, Higgins DG, Jongeneel CV, Notredame C: **The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods.** *Nucleic Acids Res* 2007:W645-648.
58. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl Acids Symp Ser* 1999, **41:**95-98.
59. Ovcharenko I, Loots GG, Giardine BM, Hou M, Ma J, Hardison RC, Stubbs L, Miller W: **Mulan: multiple-sequence local alignment and visualization for studying function and evolution.** *Genome Res* 2005, **15(1):**184-194.
60. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22(22):**4673-4680.
61. Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ: **The ITS region of nuclear ribosomal DNA – a valuable source of evidence on angiosperm phylogeny.** *Ann Mo Bot Gard* 1995, **82:**247-277.
62. Swofford DL: **PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.** Sinauer Associates, Sunderland, Massachusetts; 2003.
63. Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution.** *Bioinformatics* 1998, **14(9):**817-818.
64. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24(8):**1596-1599.

65. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R: **DnaSP, DNA polymorphism analyses by the coalescent and other methods.** *Bioinformatics* 2003, **19(18):**2496-2497.
66. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res* 2006:D140-144.
67. Matsumoto M, Nishimura T: **Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator.** *ACM Trans Model Comput Simul* 1998, **8(1):**3-30.
68. Loots GG, Ovcharenko I: **Dcode.org anthology of comparative genomic tools.** *Nucleic Acids Res* 2005:W56-64.
69. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14(6):**1188-1190.
70. Karimi M, Inze D, Depicker A: **GATEWAY vectors for Agrobacterium-mediated plant transformation.** *Trends Plant Sci* 2002, **7(5):**193-195.
71. Clough SJ, Bent AF: **Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana.** *Plant J* 1998, **16(6):**735-743.
72. Jefferson RA, Kavanagh TA, Bevan MW: **GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants.** *Embo J* 1987, **6(13):**3901-3907.