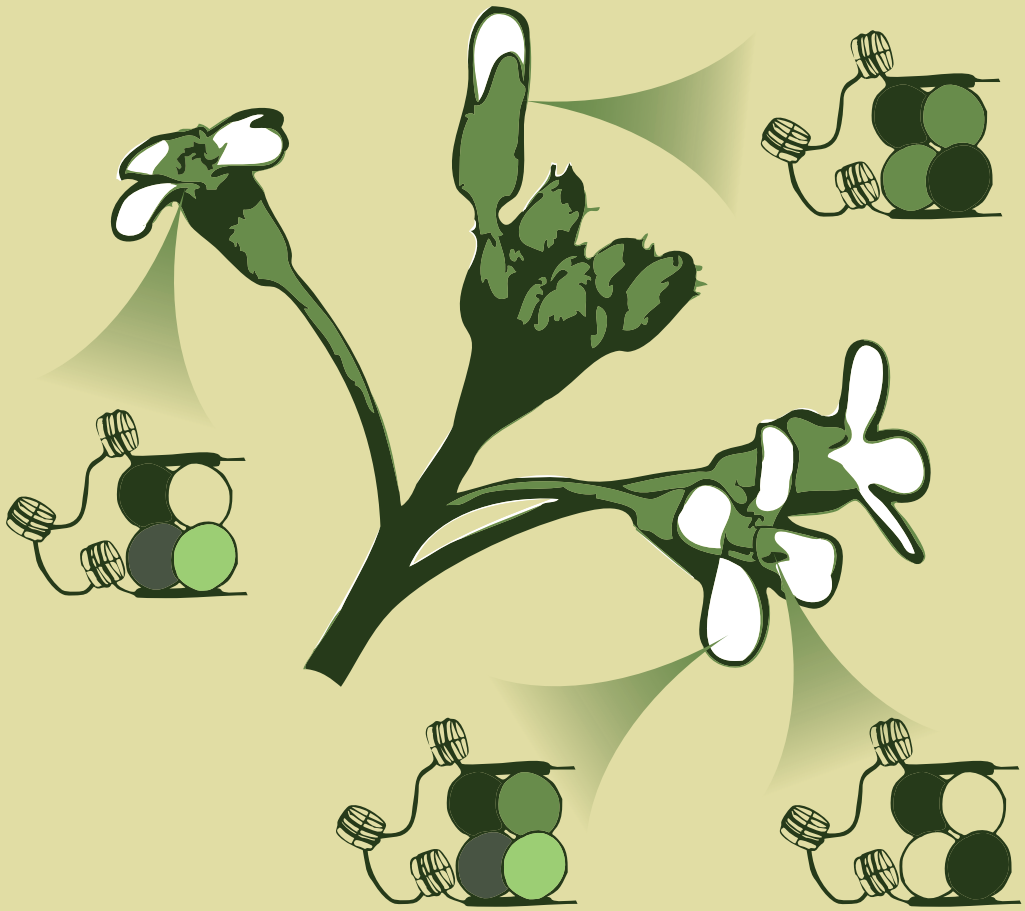
MADS interactomics

Towards understanding the molecular mechanisms of plant MADS-domain transcription factor function



Cezary Smaczniak

MADS interactomics

Towards understanding the molecular mechanisms of plant MADS-domain transcription factor function

Cezary Smaczniak

Thesis Committee

Promotor
Prof. dr. ir. G.C. Angenent
Personal chair at the Laboratory of Molecular Biology
Wageningen University

Co-promotor
Dr. K. Kaufmann
Assistant Professor, Laboratory of Molecular Biology
Wageningen University

Other members
Prof. dr. ir. B.J.G. Scheres, Wageningen University
Prof. dr. B. Davies, University of Leeds, United Kingdom
Dr. C.C. Carles, Joseph Fourier University, Grenoble, France
Dr. ir. J.J.M. Vervoort, Wageningen University

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences.

MADS interactomics

Towards understanding the molecular mechanisms of plant MADS-domain transcription factor function

Cezary Dominik Smaczniak

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 16 January 2013
at 1.30 p.m. in the Aula.

Cezary Dominik Smaczniak

MADS interactomics. Towards understanding the molecular mechanisms of plant MADS-domain transcription factor function.

178 pages

Thesis, Wageningen University, Wageningen, NL (2013).
With references, with summaries in Dutch, English and Polish.

ISBN 978-94-6173-452-5

Contents

Chapter 1

Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies

Cezary Smaczniak[1,2,3], Richard G. H. Immink[2,3], Gerco C. Angenent[2,3], Kerstin Kaufmann[1,2]

[1] Laboratory of Molecular Biology, Wageningen University, 6708PB Wageningen, The Netherlands
[2] Business Unit Bioscience, Plant Research International, 6708PB Wageningen, The Netherlands
[3] Centre for BioSystems Genomics, 6708PB Wageningen, The Netherlands

Development **139**, 3081-98 (2012)

Abstract

Members of the MADS-box transcription factor family play essential roles in almost every developmental process in plants. Many MADS-box genes have conserved functions across the flowering plants, but some have acquired novel functions in specific species during evolution. The analyses of MADS-domain protein interactions and target genes have provided new insights into their molecular functions. Here, we review recent findings on MADS-box gene functions in *Arabidopsis* and discuss the evolutionary history and functional diversification of this gene family in plants. We also discuss possible mechanisms of action of MADS-domain proteins based on their interactions with chromatin-associated factors and other transcriptional regulators.

Introduction

MADS-domain transcription factors comprise one of the best studied gene families in plants and members of this family play prominent roles in plant development. Two decades ago, the first MADS-box genes *AGAMOUS* (*AG*) from *Arabidopsis thaliana* (Yanofsky et al., 1990) and *DEFICIENS* (*DEF*) from *Antirrhinum majus* (Schwarz-Sommer et al., 1990) were discovered as regulators of floral organ identity. The sequence of the ~60 amino acid DNA-binding domains within these proteins showed striking similarities to that of the previously characterized proteins serum response factor (SRF) in *Homo sapiens* (Norman et al., 1988) and Minichromosome maintenance 1 (Mcm1) in *Saccharomyces cerevisiae* (Passmore et al., 1988). This shared and conserved domain was named the MADS domain (for MCM1, AG, DEF and SRF) and is present in all MADS-domain transcription factor family members (Schwarz-Sommer et al., 1990). Structural analysis of animal and yeast MADS domains showed that the N-terminal and central parts of the MADS domain make contacts with the DNA, while the C-terminal part of this domain contributes mainly to protein dimerization, resulting in a DNA-binding protein dimer consisting of two interacting MADS monomers (e.g. Pellegrini et al., 1995; Huang et al., 2000). Over the past 22 years, many MADS-box gene functions were uncovered in the model species *Arabidopsis thaliana* and in other flowering plants. Important model plant species for MADS-box gene research include snapdragon (*Antirrhinum majus*) (reviewed by Schwarz-Sommer et al., 2003), tomato (*Solanum lycopersicum*), petunia (*Petunia hybrida*) (Gerats and Vandenbussche, 2005), gerbera (*Gerbera hybrida*) (Teeri et al., 2006) and rice (*Oryza sativa*) (reviewed by Yoshida and Nagato, 2011).

Initially, MADS-box genes were found to be major players in floral organ specification, but more recent studies revealed functions for MADS-box genes in the morphogenesis of almost all organs and throughout the plant life cycle, from embryo to gametophyte development. The MADS-box gene family in higher plants is

significantly larger than that found in animals or fungi, with more than 100 genes in representative flowering plant genomes (De Bodt et al., 2005). This large family arose by a number of duplication events, which allowed divergence of functions of individual paralogs (see Glossary, Box 1).

Box 1. Glossary.

Angiosperms. Flowering plants that produce: seeds from ovules contained in ovaries after double fertilization by pollen; and endosperm (a nutritive tissue) containing a seed surrounded by a fruit.

Apical meristem. A meristem located at the tip of a plant shoot (SAM) or root (RAM).

CArG-box. The consensus MADS-domain binding motif with the DNA sequence: CC[A/T]$_6$GG.

Ecotype. A genetically distinct variety or population of a species that is adapted to a particular set of environmental conditions.

Floral meristem. A meristem that produces floral organs: sepals, petals, stamens and carpels.

Gymnosperms. Seed-bearing plants with ovules that are not contained in ovaries. Gymnosperms produce unenclosed ('naked') seeds.

Homeotic genes. Genes that control the transformation of one organ type into another.

Inflorescence meristem. A shoot meristem that produces flowers. In *Arabidopsis*, an example of a monopodial plant, inflorescence meristems (IMs) grow continuously and initiate flowers laterally. In tomato, a sympodial plant, IMs terminate in flowers and growth continues from new axillary IMs that repeat this process to generate compound inflorescences. In grasses, IMs produce lateral meristems with more specialized IM identities, reflecting the complex architecture of the grass inflorescence.

Meristem. A tissue of undifferentiated plant cells (analogous to stem cells) typically located at regions where growth takes place.

Neofunctionalization. The process by which a homologous gene develops a function that differs from that of the ancestral gene.

Orthologs. Homologous genes in different species that originated from a single ancestral gene through a speciation process. Owing to frequent gene duplication, which is often linked with polyploidization in plants, orthologs in a strict sense can only be found in very closely related species. A more correct, but less well known, term would be 'orthogroup': the set of genes from extant species that descended from a single gene in the species' last common ancestor (Wapinski et al., 2007).

Paralogs. Homologous genes that originated from an ancestral gene through gene duplication.

Subfunctionalization. The process by which multiple functions of the ancestral gene are divided between homologous genes.

In this review, we provide an overview of the developmental functions of MADS-box genes in flowering plants, with a main focus on *Arabidopsis*. We also summarize the roles of MADS-box genes in other plant species. Owing to the vast array of functions performed by MADS-box genes, and hence the large body of literature that is devoted to this field of research, a comprehensive review of all known studies of MADS-box genes would not be possible, but we hope that the examples discussed below illustrate different aspects of the evolution of MADS-box gene functions, their conserved roles, and their contribution to the origin of morphological novelties.

Type I and type II MADS-domain proteins

The MADS-box gene family can be divided into two lineages, type I and type II, based on their protein domain structure (**Figure 1**). Genes from the type I lineage are a heterogeneous group, having only the ~180 bp DNA sequence encoding the MADS domain in common (De Bodt et al., 2003; Kofuji et al., 2003; Parenicová et al., 2003). They can be further classified into 3 subclasses: Mα, Mβ and Mγ (**Figure 1A**). Type I genes were discovered only after the completion of the *Arabidopsis* genome sequence (Alvarez-Buylla et al., 2000; The Arabidopsis Genome Initiative, 2000). Although the type I MADS-box genes outnumber the type II genes, no gene functions were assigned to type I genes until relatively recently (reviewed by Masiero et al., 2011).

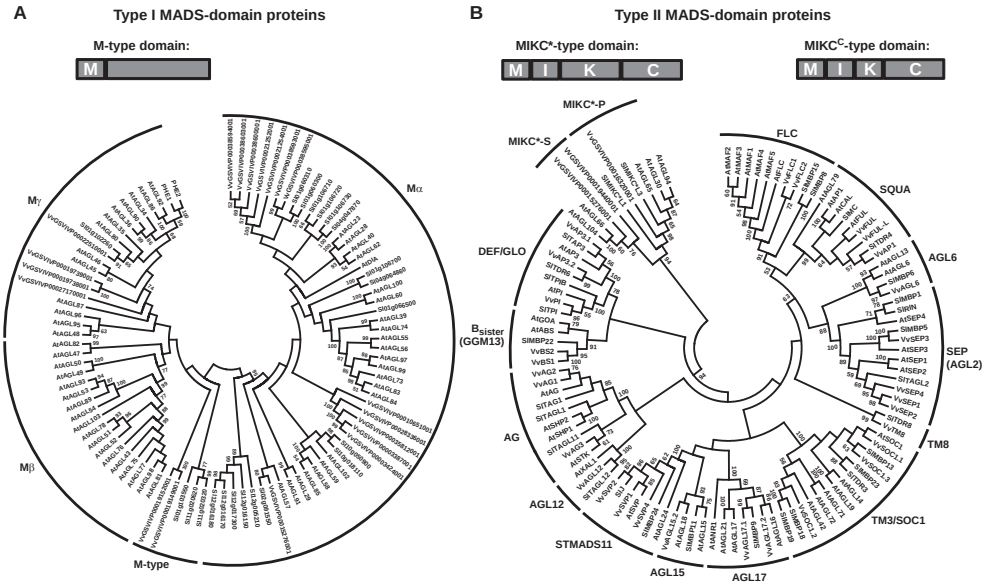


Figure 1. Domain structure and classification of MADS-domain proteins. Phylogenetic analyses and the domain structure of selected representatives of **(A)** type I and **(B)** type II MADS-box transcription factors from thale cress (*Arabidopsis thaliana*, *At*), grape (*Vitis vinifera*, *Vv*) and tomato (*Solanum lycopersicum*, *Sl*). Trees were built after codon alignment by MUSCLE (Edgar, 2004) using the neighbor-joining method with a 1000 replicate bootstrap analysis and visualized in a topology-only mode. Phylogenetic analyses were conducted in MEGA5 (Tamura et al., 2011). Type I MADS-box transcription factors possess one conserved domain, the DNA-binding MADS domain (M), and a long, variable C-terminal domain. Plant type II MADS-box transcription factors have four domains: the DNA-binding MADS, the intervening (I), the keratin-like (K) and the C-terminal (C) domains. MIKC*-type proteins are usually longer than MIKC^C-type proteins, probably owing to a longer K domain (Kwantes et al., 2012).

The type II lineage contains the well-studied floral homeotic genes (see **Glossary**, **Box 1**) as well as other genes involved in various developmental processes

(e.g. embryogenesis, flowering time and fruit development). Plant type II MADS-domain proteins have a modular domain structure, which is referred to as the MIKC structure; they contain an N-terminally located DNA-binding MADS domain, followed by the I (intervening) and K (keratin-like) regions, which are essential for dimerization and higher-order complex formation, and finally a highly variable C-terminal domain, which may have roles in protein complex formation and transcriptional regulation (reviewed by Kaufmann et al., 2005b). Based on differences in their domain structure, MIKC-type MADS-box genes have been further classified into (canonical) MIKC^C-type and MIKC*-type genes (Henschel et al., 2002) (**Figure 1B**). The latter are characterized by an altered protein domain structure, possibly linked to the duplication of exons encoding a subregion of the K-domain (Kwantes et al., 2012). Moreover, we can divide MIKC^C-type MADS-box genes into several distinctive subfamilies based on their phylogeny (**Figure 1B**). Most subfamilies of MIKC^C-type genes appear to have originated in ancestral seed plants and have been named after their first identified founding members (Becker and Theissen, 2003). Proteins of the different subfamilies are often characterized by distinct sequence motifs in their C-terminal domains, which further diversified during evolution by frameshift mutations (see Vandenbussche et al., 2003a). At least for some MIKC-type proteins, the C-terminal motifs appear to be dispensable for basic protein function (Piwarzyk et al., 2007; Benlloch et al., 2009).

Members of the different MIKC^C-type subfamilies often have related or even conserved functions in different flowering plant species. For example, the specification of stamens and carpels in the flower is exerted by genes of the AGAMOUS (AG) clade in different angiosperm species. In a similar fashion, members of the DEFICIENS (DEF) and GLOBOSA (GLO) subfamilies control stamen and petal identity, and members of the SQUAMOSA (SQUA) and SEPALLATA [SEP or AGAMOUS-LIKE 2 (AGL2)] have (partly) conserved roles in floral meristem (see **Glossary**, **Box 1**) and organ specification in various angiosperms. Members of other MIKC^C-type subfamilies, such as the TOMATO MADSBOX 3 (TM3), FLOWERING LOCUS C (FLC) and SOLANUM TUBEROSUM MADS-BOX 11 (STMADS11) clades, act predominantly in floral transition. AGL12 and AGL17 subfamily members appear to act mostly in root development (although they also influence floral transition). Intriguingly, many MIKC^C-type genes act in more than one developmental process or developmental stage.

MADS-box gene functions in the *Arabidopsis thaliana*

The functional characterization of *Arabidopsis thaliana* MADS-box genes started with their discovery in the early 1990s. To date, functions for nearly half of these genes have been described (**Table 1**). In addition to genetic studies, genome-

wide expression and interaction studies have shed light on the potential roles of MADS-domain proteins in plant development. Below, we provide an overview of MADS-box gene functions in the *Arabidopsis thaliana* life cycle (summarized in **Figure 2**), highlighting some of the recent studies and advances.

Gametophyte, embryo and seed development

The plant life cycle culminates in the generation of male and female haploid gametes (sperm cells and embryo sac, respectively) by meiosis. The gametes are then fused during the fertilization process to generate a diploid zygote. In *Arabidopsis* and many other flowering plant species a second sperm nucleus fuses with two nuclei of the central cell in the embryo sac to produce the extra-embryonic triploid endosperm. Embryonic development results in a developmentally arrested embryo in the mature seed in which the major body axis is established. As we highlight below, MADS-domain proteins are involved in several stages of gametophytic and embryonic development.

Genetic studies have revealed functions for several type I MADS-box genes in female gametogenesis and in seed development (**Figure 2** and **Table 1**) (reviewed by Masiero et al., 2011). For example, the Mγ protein AGL80 and the Mα protein DIANA (DIA; AGL61) form a functional protein dimer and control the differentiation of the central cell (Portereiko et al., 2006; Bemer et al., 2008; Steffen et al., 2008). *AGL80* is also expressed during endosperm development. *AGL62*, a close paralog of *DIA*, suppresses premature endosperm cellularization (Kang et al., 2008) and encodes a protein that can also interact with AGL80 (Kang et al., 2008), although the relevance of this interaction is not well understood. The overlapping type II MADS-domain proteins, at least some type I MADS-domain proteins act together in heteromeric protein complexes. A large-scale yeast two-hybrid protein interaction screen revealed multiple interactions between type I MADS-domain proteins, mostly between members of different subclades (de Folter et al., 2005). A large scale expression analysis showed that most (38 out of 61) type I MADS-box genes are active in the female gametophyte and seed development processes (Bemer et al., 2010), and some of them exhibit highly specific expression patterns in particular cells (Bemer et al., 2010; Wuest et al., 2010). However, for the majority of these genes, no direct function has been attributed so far, probably owing to genetic redundancy.

Several type I MADS-box genes are epigenetically repressed by the action of a PRC2-type polycomb group (PcG) complex during seed development and other stages of plant development (Zhang et al., 2007; Dreni et al., 2011). Examples are *AGL23*, which is an Mα-type MADS-box gene that has a role in embryo sac development (Colombo et al., 2008), and *PHERES1* (*PHE1*; *AGL37*) (Kohler et al., 2003). *PHE1* provided one of the first examples of imprinting in plants: the expression of the

Table 1. MADS-box gene functions in development of *Arabidopsis thaliana*.

Gene	Symbol	Phylogenetic Group (Subfamily)	Functions	References
AGAMOUS-LIKE 65, 66, 104	*AGL65, 66, 104*	MIKC*	Pollen maturation and tube growth.	(Adamczyk and Fernandez, 2009)
AGAMOUS	*AG*	MIKC^C (AG)	Homeotic C-class gene; carpel and stamen specification	(Yanofsky et al., 1990)
SHATTERPROOF 1, 2	*SHP1, 2*	MIKC^C (AG)	Carpel, ovule and fruit development. Dehiscence. Periodic lateral root formation	(Liljegren et al., 2000; Moreno-Risueno et al., 2010)
SEEDSTICK	*STK*	MIKC^C (AG)	Carpel and ovule development. Periodic lateral root formation	(Pinyopich et al., 2003; Moreno-Risueno et al., 2010)
XAANTAL 1	*XAL1*	MIKC^C (AGL12)	Root development - cell-cycle regulation. Transition to flowering (activator).	(Tapia-Lopez et al., 2008)
AGAMOUS-LIKE 15	*AGL15*	MIKC^C (AGL15)	*Embryogenesis. Transition to flowering (repressor) with AGL18. *Sepal and petal longevity. *Fruit maturation.	(Heck et al., 1995; Fernandez et al., 2000; Harding et al., 2003)
AGAMOUS-LIKE 18	*AGL18*	MIKC^C (AGL15)	Transition to flowering (repressor) with AGL15.	(Adamczyk et al., 2007)
AGAMOUS-LIKE 16	*AGL16*	MIKC^C (AGL17)	*Number and distribution of stomata	(Kutter et al., 2007)
AGAMOUS-LIKE 17	*AGL17*	MIKC^C (AGL17)	*Transition to flowering (activator).	(Han et al., 2008)
ARABIDOPSIS NITRATE REGULATED 1	*ANR1*	MIKC^C (AGL17)	Root development; nutrient response.	(Zhang and Forde, 1998)
AGAMOUS-LIKE 6	*AGL6*	MIKC^C (AGL6)	Transition to flowering (activator). *Lateral organ development.	(Koo et al., 2010; Yoo et al., 2011)
ARABIDOPSIS BSISTER	*ABS*	MIKC^C (GGM13)	Seed pigmentation and endothelium development	(Nesi et al., 2002; Kaufmann et al., 2005a; de Folter et al., 2006)
GORDITA	*GOA*	MIKC^C (GGM13)	Fruit development	(Prasad et al., 2010)
APETALA 3	*AP3*	MIKC^C (DEF/GLO)	Homeotic B-class gene; petal and stamen specification	(Jack et al., 1992)
PISTILLATA	*PI*	MIKC^C (DEF/GLO)	Homeotic B-class gene; petal and stamen specification	(Goto and Meyerowitz, 1994)
FLOWERING LOCUS C	*FLC*	MIKC^C (FLC)	Transition to flowering (repressor). *Germination. *Juvenile-to-adult transition. *Initiation of flowering. *Flower organ development.	(Michaels and Amasino, 1999; Chiang et al., 2009; Deng et al., 2011)
MADS AFFECTING FLOWERING 1-4	*MAF1-4*	MIKC^C (FLC)	*Transition to flowering (repressors).	(Ratcliffe et al., 2001; Ratcliffe et al., 2003)
MADS AFFECTING FLOWERING 5	*MAF5*	MIKC^C (FLC)	*Transition to flowering (activator).	(Ratcliffe et al., 2003)

Table 1 Continued.

Gene	Symbol	Phylogenetic Group (Subfamily)	Functions	References
SEPALLATA 1-4	*SEP1-4*	MIKC^C (AGL2)	Homeotic E-class genes; sepal, petal, stamen and carpel specification	(Mandel and Yanofsky, 1998; Pelaz et al., 2000; Ditta et al., 2004)
AGAMOUS-LIKE 19	*AGL19*	MIKC^C (TM3/SOC1)	Transition to flowering (activator).	(Schonrock et al., 2006)
AGAMOUS-LIKE 42 (FOREVER YOUNG FLOWER)	*AGL42 (FYF)*	MIKC^C (TM3/SOC1)	Transition to flowering (activator). *Flower organ senescence and abscission. *Root development.	(Nawy et al., 2005; Chen et al., 2011; Dorca-Fornell et al., 2011)
AGAMOUS-LIKE 71, 72	*AGL71, 72*	MIKC^C (TM3/SOC1)	Transition to flowering (activators) with AGL42.	(Dorca-Fornell et al., 2011)
SUPPRESSOR OF OVEREXPRESSION OF CO 1	*SOC1*	MIKC^C (TM3/SOC1)	Transition to flowering (activator). Periodic lateral root formation.	(Lee et al., 2000; Moreno-Risueno et al., 2010)
APETALA 1	*AP1*	MIKC^C (SQUA)	Meristem identity specification. Homeotic A-class gene.	(Mandel et al., 1992; Weigel et al., 1992; Ferrandiz et al., 2000a)
CAULIFLOWER	*CAL*	MIKC^C (SQUA)	Meristem identity specification.	(Kempin et al., 1995; Ferrandiz et al., 2000a)
FRUITFULL	*FUL*	MIKC^C (SQUA)	Meristem identity specification. Annual life cycle regulator, with SOC1. Fruit development. Cauline leaf growth.	(Gu et al., 1998; Ferrandiz et al., 2000a; Ferrandiz et al., 2000b; Melzer et al., 2008)
AGAMOUS-LIKE 24	*AGL24*	MIKC^C (STMADS11)	Transition to flowering (activator).	(Michaels et al., 2003)
SHORT VEGETATIVE PHASE	*SVP*	MIKC^C (STMADS11)	Transition to flowering (repressor).	(Hartmann et al., 2000)
AGAMOUS-LIKE 23	*AGL23*	Mα	Embryo sac development.	(Colombo et al., 2008)
AGAMOUS-LIKE 28	*AGL28*	Mα	*Transition to flowering (activator).	(Yoo et al., 2006)
AGAMOUS-LIKE 61 (DIANA)	*AGL61 (DIA)*	Mα	Central cell and endosperm development	(Bemer et al., 2008; Steffen et al., 2008)
AGAMOUS-LIKE 62	*AGL62*	Mα	Central cell development	(Kang et al., 2008)
AGAMOUS-LIKE 80	*AGL80*	Mγ	Central cell and endosperm development.	(Portereiko et al., 2006)
PHERES 1	*PHE1*	Mγ	*Seed development.	(Kohler et al., 2003; Kohler et al., 2005)

Subfamily names are according to Becker and Theissen, 2003.
* - function that is inferred based on other than mutant phenotype analysis.

maternal allele of *PHE1* is silenced by the PcG complex, whereas the paternal copy is active in embryo and endosperm, resulting in a parent-of-origin-dependent expression of *PHE1* in seeds (Kohler et al., 2005). Expression of *PHE1* is also regulated by DNA

(de)methylation (Makarevich et al., 2008; Hsieh et al., 2009; Villar et al., 2009). The dual epigenetic regulation of *AGL36* provides another example of complex control of type I MADS-box gene expression in seed development (Shirzadi et al., 2011). The downregulation of *PHE1*, *PHE2*, *AGL35*, *AGL36*, *AGL40*, *AGL62*, and *AGL90* coincides with the transition of endosperm from syncytial to cellularized stage, and this appears to be crucial for endosperm differentiation (Kang et al., 2008; Walia et al., 2009). The dosage-sensitive PRC2-mediated repression of these type I genes contributes to postzygotic compatibility and reproductive isolation between species (Walia et al., 2009).

Whereas type I MADS-box genes predominantly regulate female gametophyte and seed development, MIKC*-type genes were found to control development of male gametophytes (pollen). Combinations of double and triple mutants of *agl65*, *agl66* and *agl104* MADS-box genes give rise to several pollen-affected phenotypes with disturbed viability, delayed germination and aberrant pollen tube growth (Verelst et al., 2007a; Adamczyk and Fernandez, 2009). Expression and interaction data confirmed that these MIKC*-type gene products form a protein interaction and regulatory network controlling pollen maturation (Verelst et al., 2007a; Adamczyk and Fernandez, 2009). Moreover, in depth gene expression analysis in such double and triple mutants showed that these MIKC*-type MADS-complexes regulate transcriptome dynamics during pollen development and revealed the extent of their functional redundancy (Verelst et al., 2007b). In summary, these findings highlight the importance of protein multimerization within the MADS-box family during gametophytic and embryo development.

Despite the fact that many MIKC^{C}-type MADS-box genes show detectable expression during embryo development (Lehti-Shiu et al., 2005), few roles have been attributed to them in this developmental process. One of the first MADS-box genes shown to play a potential role in embryogenesis was the MIKC^{C}-type gene *AGL15* (Heck et al., 1995; Perry et al., 1999). Although single *agl15* mutant plants do not show any obvious embryonic phenotype, overexpression of *AGL15* promotes the production of secondary embryos (Harding et al., 2003). The identification of AGL15 target genes revealed that it directly binds loci of B3 domain transcription factor genes, which are known regulators of embryogenesis (Zheng et al., 2009). In addition to this potential role in embryo development, *AGL15* represses floral transition together with its close paralog *AGL18* (Adamczyk et al., 2007; Zheng et al., 2009).

Phase transitions in sporophytic development

In *Arabidopsis* and other plant species, major developmental transitions occur during postembryonic growth: the change from the juvenile to the vegetative phase, and later to the reproductive phase. The juvenile-to-adult transition is characterized

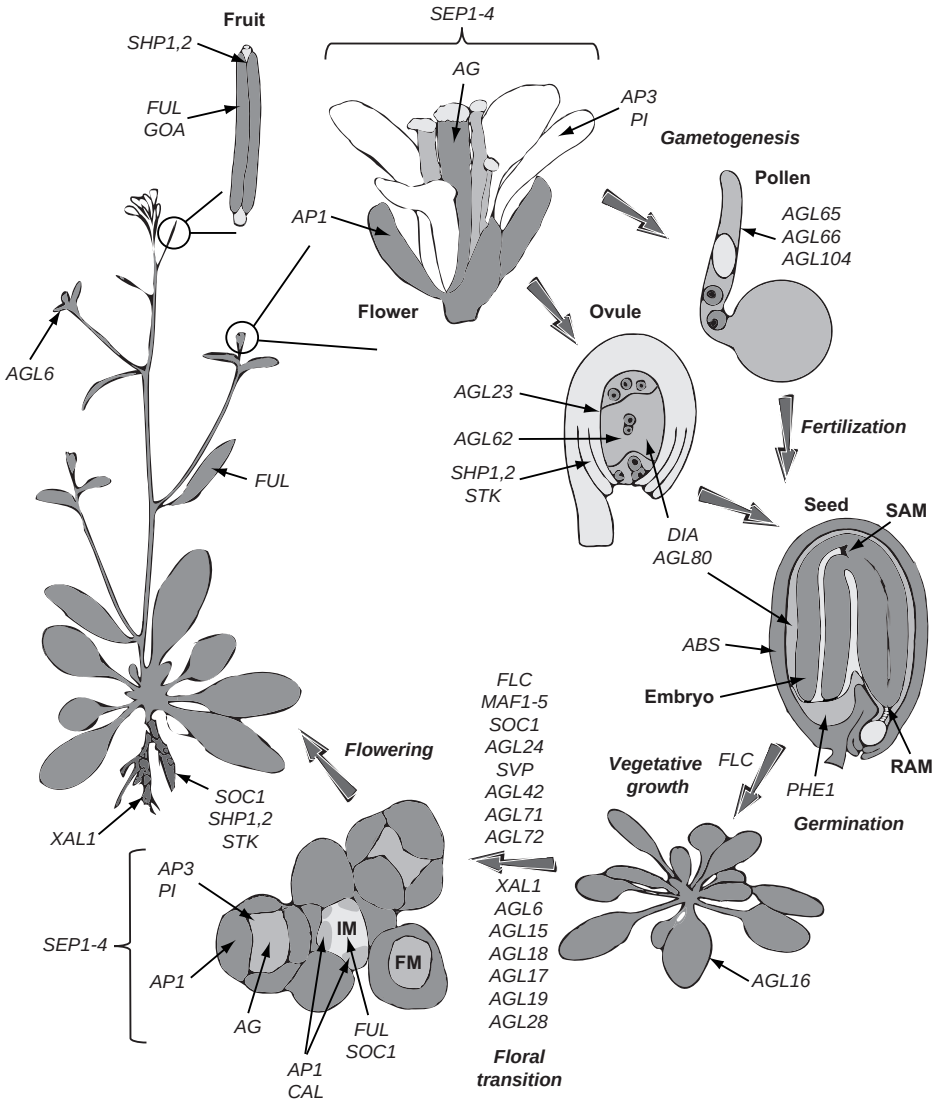


Figure 2. Functions of MADS-box genes throughout the life cycle of *Arabidopsis thaliana*. *Arabidopsis* progresses through several major phase changes during its life cycle and MADS-box genes play distinct roles in the various developmental phases and transitions. Reproductive development starts with the generation of male and female haploid gametes (gametogenesis) and, after double fertilization, this results in a developmentally arrested embryo that possesses a root apical meristem (RAM) and a shoot apical meristem (SAM), enclosed within a seed. Under favorable conditions, seeds germinate and young plants go through the vegetative phase of development in which leaves are formed and plants gain size and mass. Finally, the plant is ready to flower and the floral transition stage results in the conversion of vegetative meristems into inflorescence meristems (IMs) and floral meristems (FMs) that produce floral organs. Subsequently, gametes are formed within the inner flower organs, thus completing the cycle. The MADS-box genes that are involved in each of the various stages of development are indicated.

mainly by changes in the morphology and epidermal patterning of leaves in *Arabidopsis*. The vegetative-to-reproductive transition results in the conversion of the vegetative apical meristem (see **Glossary**, **Box 1**) into an inflorescence meristem (IM; see **Glossary**, **Box 1**), which then produces flowers and cauline leaves. Developmental transitions are regulated by external and internal cues, such as light, plant age and temperature (Blazquez, 2000; Poethig, 2003). The different signaling cascades that respond to these cues are integrated by transcriptional master regulators, many of which are MIKC^{C}-type MADS-box transcription factors. These factors can act as repressors or activators of the transition, and integrate the input from temperature, day-length, autonomous and hormonal pathways.

An important repressor of the floral transition is *FLC*, the expression of which is controlled by vernalization (Michaels and Amasino, 1999). During prolonged cold exposure, *FLC* expression is downregulated by epigenetic chromatin regulators and possibly by long non-coding RNAs, allowing the plant to flower in spring in winter-annual accessions of *Arabidopsis* (reviewed by Kim and Sung, 2012). FLC interacts with another MIKC^{C}-type floral repressor, SHORT VEGETATIVE PHASE (SVP) (Li et al., 2008). FLC and SVP repress the expression of the mobile floral inducer ('florigen') *FLOWERING LOCUS T* (*FT*) and other genes that initiate floral transition, in a partly tissue-specific fashion (Searle et al., 2006; Li et al., 2008; Jang et al., 2009). Recently, data from chromatin immunoprecipitation followed by hybridization to tiling arrays (ChIP-CHIP) revealed that SVP also directly activates other repressors of floral transition, including members of the *APETALA2* (*AP2*) transcription factor family (Tao et al., 2012), which in turn also repress *FT*. A similar genome-wide target gene identification approach indicated that *FLC* is involved in other developmental processes in addition to floral repression, including the juvenile-to-adult transition and floral organ development (Deng et al., 2011). *FLC* also has a role in in temperature-dependent germination during seed development (Chiang et al., 2009).

One major target of repression by FLC is the MIKC^{C}-type transcription factor *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), which is an activator of floral transition at the shoot apex. *SOC1* integrates external (e.g. light) and internal signals (Lee et al., 2000; Samach et al., 2000; Seo et al., 2009) and acts in a positive feedback loop with *AGL24* (Liu et al., 2008), yet another important MIKC^{C}-type factor that positively regulates flowering in *Arabidopsis* (Michaels et al., 2003). SOC1 and AGL24 appear to work in a larger molecular complex and transmit the flowering signals onto *LEAFY* (*LFY*) (Lee et al., 2008), which is a non-MADS regulator of floral meristem identity that links floral induction with flower development (Weigel et al., 1992). Additionally, SOC1 represses the precocious expression of floral homeotic B-, C- and E-class genes (see **Box 2**) in IMs and early

floral meristems in a redundant manner with AGL24 and SVP, respectively (Gregis et al., 2009; Liu et al., 2009; Torti et al., 2012). *SOC1* also interacts with *FRUITFULL* (*FUL*) and together they play a role in establishing the annual life habit of *Arabidopsis* (Melzer et al., 2008). Recently, it was revealed that the floral activator SOC1 and the floral repressor SVP act in an opposing fashion on a partially overlapping set of direct target genes during floral transition (Tao et al., 2012).

A number of other MIKC^C-type genes, for example other TM3 clade members in addition to *SOC1* (**Table 1**) (Dorca-Fornell et al., 2011), have been shown to regulate the floral transition in *Arabidopsis*. We conclude that flowering time is determined by the interplay between multiple MADS-box genes, whereby master regulators such us the flowering repressor *FLC* and the flowering activator *SOC1* act in concert with other non-MADS key regulators like the *FT-FD* complex and *LFY* to integrate and process external and internal flowering signals (reviewed by Pose et al., 2012).

Flower and fruit development

The floral transition results in the formation of IMs, which generate floral meristems at their flanks that in turn produce floral organs (sepals, petals, stamens and carpels). Meristems are specified by the action of meristem identity genes, which interact in complex regulatory networks with multiple feedback and feed forward loops (Kaufmann et al., 2010b). Whereas *SOC1* and *AGL24* have been referred to as IM identity genes, the partially redundantly acting MIKC^C-type genes *AP1* and *CAULIFLOWER* (*CAL*) specify floral meristem identity (Kempin et al., 1995). It has been shown that most of the early AP1 target genes are downregulated by AP1, suggesting that this protein acts mainly as a transcriptional repressor during floral meristem initiation (Kaufmann et al., 2010c). During the early stages of flower development, AP1 can interact with SVP and this complex may initially repress homeotic gene activity in early floral meristems (Gregis et al., 2009). In addition, AP1 activates (together with LFY) the expression of (other) floral homeotic genes and other genes involved in floral patterning, at least at later developmental stages (Ng and Yanofsky, 2001; Kaufmann et al., 2010c; Winter et al., 2011).

The identities of different types of floral organs are specified by homeotic genes, nearly all of which encode MIKC^C-type proteins. Fundamental models have been proposed to explain the genetic and molecular interactions of these floral master regulators (see **Box 2**). Homeotic genes were classified into functional classes A to E based on their characteristic mutant phenotypes (Coen and Meyerowitz, 1991; Colombo et al., 1995; Theissen, 2001). The homeotic A-function has received critical attention in recent years. The A-class gene *APETALA1* (*AP1*) has been proposed to have a more general role in establishing floral meristem fate, which more accurately

explains the phenotype of most *ap1* mutant alleles in *Arabidopsis* and those of orthologous genes in other plant species (Causier et al., 2010). It has also been proposed that the second traditional 'A-class' gene *AP2*, the only non-MADS-box transcription factor in the ABCDE model, acts as a cadastral gene, which becomes restricted in its expression by microRNA172; the miR172/AP2 module coordinates the specification of perianth versus reproductive organs (Wollmann et al., 2010). The E-class proteins, which comprise the four largely redundantly acting SEP subfamily members, have a special role as mediators of higher-order complex formation among floral MADS-domain proteins (Honma and Goto, 2001). Homeotic MADS-box genes are initially expressed in patterned fashion in floral meristems and maintain expression during floral organ differentiation (Urbanus et al., 2009; for review, see Krizek and Fletcher, 2005). They control the expression of many other genes at the different stages, a number of them directly (reviewed by Ito, 2011). The D-class genes *SHATTERPROOF 1* and *2* (*SHP1*,*2*) and *SEEDSTICK* (*STK*) specify ovule identity and differentiation (Favaro et al., 2003; Pinyopich et al., 2003; Matias-Hernandez et al., 2010), in part by regulating the expression of REM family transcription factors (Matias-Henandez et al., 2010). D-class proteins interact in larger complexes with E-class proteins and the homeobox transcription factor BELL1 (Favaro et al., 2003; Brambilla et al., 2007).

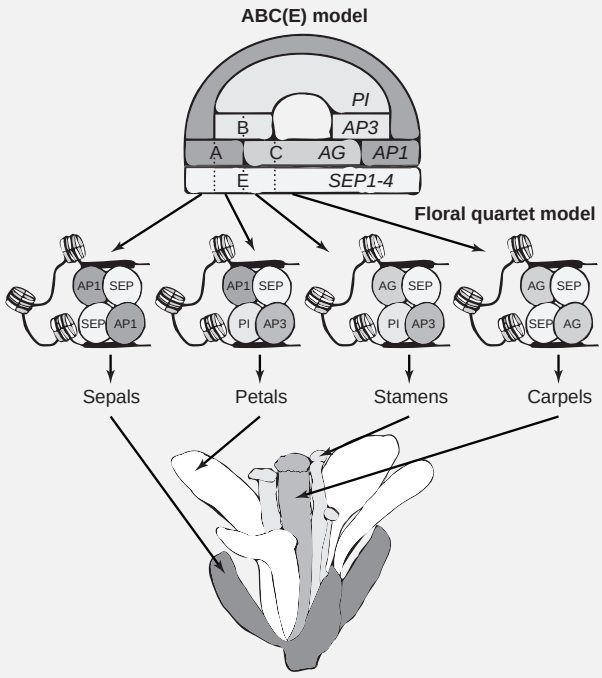
Fruit differentiation is controlled by the antagonistically acting *SHP1*,*2* and *FUL* genes, which are expressed in the valve margins and in the valves, respectively (Ferrandiz et al., 2000b; Colombo et al., 2010). The B_{sister} clade gene *GORDITA* (*GOA*), which has a divergent protein sequence, regulates fruit size in *Arabidopsis* by repressing cell expansion (Prasad et al., 2010). A close paralog of *GOA*, the more conserved B_{sister} gene *ARABIDOPSIS BSISTER* (*ABS*, *TT16*), controls endothelium development and (thereby) seed maturation (Nesi et al., 2002; Kaufmann et al., 2005a; de Folter et al., 2006; Mizzotti et al., 2012). The interaction of ABS with AG clade proteins is mediated by SEP proteins, suggesting roles for tetrameric $MIKC^C$-type protein complexes in processes beyond floral organ specification (Kaufmann et al., 2005a; Mizzotti et al., 2012).

Root and leaf morphogenesis

Although $MIKC^C$-type MADS-box genes are best known for their roles in floral transition and flower development, several of them have additional or specific functions during root morphogenesis. *ARABIDOPSIS NITRATE REGULATED1* (*ANR1*) has a function in nutrient response in roots and controls lateral root elongation in response to nitrate (Zhang and Forde, 1998; Gan et al., 2005). Other members of the AGL17 clade (e.g. *AGL16*, *AGL17* and *AGL21*) are also expressed predominantly in roots (Burgeff et al., 2002). *AGL16* and *AGL21* are regulated by

nitrogen, similar to *ANR1*, and AGL21 has recently been shown to interact with an endosome-associated protein that promotes intercellular movement (Gan et al., 2005; Koizumi et al., 2011). Besides its potential role in root morphogenesis, *AGL17* also affects floral transition (Han et al., 2008).

Box 2. ABC and floral quartet models of floral organ specification.



As for the majority of angiosperm flowers, the *Arabidopsis thaliana* flower is structured into four concentric whorls of floral organs. The four organ types are sepals (outermost whorl, whorl 1), petals (whorl 2), stamens (whorl 3) and carpels (whorl 4) (Haughn and Somerville, 1988). In the classic ABC model, which is based on homeotic mutant phenotypes in *Arabidopsis thaliana* and *Antirrhinum majus*, three classes of genes (A, B and C) are essential to guide the specification and formation of floral organs (Coen and Meyerowitz, 1991; see also Haughn and Somerville, 1988): A-class genes specify sepal identity, A-class and B-class genes together determine petals, B-class genes and the C-class gene specify stamens, and the C-class gene determines carpel identity. In *Arabidopsis*, the A-class genes are *APETALA1* (*AP1*) and *AP2*, B-class genes are *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), and the C-class gene is *AGAMOUS* (*AG*). Based on the overexpression phenotypes of the AG clade gene *FLORAL BINDING PROTEIN 11* (*FBP11*) in petunia, an additional homeotic gene class, the D class, was proposed to specify ovule identity (Colombo et al., 1995), and, in *Arabidopsis*, ovule identity is specified by the related AG subfamily member *SEEDSTICK* (*STK*) together with *SHATTERPROOF 1* and *2* (*SHP1,2*) (Pinyopich et al., 2003). Identification of the redundantly functioning SEPALLATA genes (*SEP1-4*), which are essential for the development of all flower whorls (Pelaz et al., 2000; Ditta et al., 2004), led to the extension of the ABC model to include these E-class genes (Theissen, 2001). The homeotic A-function has been under debate in recent years (see text).

Except for *AP2*, all floral homeotic genes encode MADS-domain transcription factors. In line with the observed combinatorial higher-order complex formation of MADS-domain proteins (Honma and Goto, 2001), the floral quartet model was postulated to explain the molecular mechanism of action underlying ABCDE protein function in floral organ specification (Theissen, 2001; Theissen and Saedler, 2001). The organ-specific combinatorial quaternary MADS-domain protein complexes are proposed to control differentiation and outgrowth of the distinct floral organs in the four concentric whorls.

XAANTAL1 (*XAL1*; *AGL12*) controls auxin-dependent cell-cycle regulation affecting root growth and also has an influence on flowering time (Tapia-Lopez et al.,

2008). *SOC1*, as well as the AG-clade genes *SHP1*,*2* and *STK*, which have well-described roles in reproductive transition and carpel development, have recently been shown to act in periodic lateral root formation (Moreno-Risueno et al., 2010). Other TM3/SOC1 clade genes that control floral transition in the shoot, are also expressed in the root (Nawy et al., 2005), but the biological relevance of this is not yet known.

As with their functions in roots, the roles of MADS-box genes in leaf development are largely unexplored. One example of a functionally characterized gene is the microRNA-regulated *AGL16*, which controls stomata initiation in leaves and other organs (Kutter et al., 2007). More studies are needed, however, to unveil whether other MADS-box genes that are expressed in leaves play roles in leaf morphogenesis.

Examples of MIKC^C-type MADS-box gene functions in other plant species

The key functions of MIKC^C-type MADS-box transcription factors in a variety of developmental processes in plants suggest possible roles of these proteins in the evolution of morphologies, life history strategies and reproductive mechanisms (see **Table 2** for examples). MIKC^C-type genes are thus major research targets in evolutionary developmental biology (evo-devo) studies as well as in crop plant biotechnology and domestication research. The availability of transcriptome datasets and/or genome sequences led to a more comprehensive identification and characterization of MIKC^C-type genes in different plant species, such as tomato (Hileman et al., 2006) and grapevine (*Vitis vinifera*) (Diaz-Riquelme et al., 2009), or of MADS-box-genes in general in species such as rice (Arora et al., 2007), poplar (*Populus trichocarpa*) (Leseberg et al., 2006) and cucumber (*Cucumis sativus*) (Hu and Liu, 2012). Because of the tremendous amount of research carried out on MIKC^C-type genes in various species, we highlight here only some of the recent findings. We focus on examples where the function or regulation of MIKC^C-type genes deviates from their orthologs in *Arabidopsis* and might thus have an impact on evolution.

Flower development

A major model system in evo-devo research is the angiosperm (see **Glossary**, **Box 1**) flower. While the basic types of floral organs are largely conserved, the number and morphology of floral organs are highly diverse, reflecting diversity in reproductive strategies (Soltis et al., 2002). Next to *Arabidopsis*, the roles of MIKC^C-type genes in flower development have been extensively studied in the eudicot species such as snapdragon, petunia and tomato as well as in monocots such as rice and the orchid *Phalaenopsis*. Among the upcoming model species are pea (*Medicago sativa*) (Hecht et al., 2005) and basal eudicots such as California poppy (*Eschscholzia californica*) (Zahn et al., 2010). The ability to analyze gene functions in a plant species depends on the

availability of tools, such as the ability to transform the plant or amenability for virus-induced gene silencing (VIGS) (Becker and Lange, 2010), and of genome and/or transcriptome resources.

Table 2. Evolution of MIKC^C-type MADS-box gene functions in flowering plants.

Subfamily	Functions in *Arabidopsis*	(Additional) functions in other plant lineages	References
AG	Floral homeotic C and D functions.	Lineage-specific subfunctionalization of the homeotic C function. Fruit development, e.g. tomato vs. *Arabidopsis*.	(Causier et al., 2005; Airoldi et al., 2010)
AP3 **PI**	Floral homeotic B function.	Tepal diversification in orchids. Variable roles in specification of petaloid organs.	(Mondragon-Palomino and Theissen, 2008; Chang et al., 2010)
STMADS11	Control of floral transition. Repression of precocious homeotic gene expression.	Inflated calyx syndrome in *Physalis*. Floral bud dormancy in *Prunus*. Repression of prophyll development in *Antirrhinum*. Flower abscission zone development in tomato.	(Mao et al., 2000; Masiero et al., 2004; He and Saedler, 2005; Z. Li et al., 2009)
AGL2	Floral homeotic E function.	Inflorescence meristem determinacy in *Gerbera*. Tomato fruit ripening.	(Vrebalov et al., 2002; Uimari et al., 2004)
SQUA	Floral meristem and organ identity specification. Floral transition. Fruit development.	Potato axillary bud formation. Potential role in *Vitis* tendril development. Variable roles in fruit development, sepal size and floral abscission in tomato. Variable roles in floral transition.	(Rosin et al., 2003; Calonje et al., 2004; Nakano et al., 2012)
FLC	Repressor of floral transition. Seed germination.	Potential role in floral bud dormancy. Perennial life history in *Arabis alpina*.	(Du et al., 2008; Wang et al., 2009; Zhang et al., 2009)

This table exemplifies MIKC^C-type gene subfamilies for which gene functions have been studied in different angiosperm species. Subfamily names are according to Becker and Theissen, 2003.

Some of the core functions of MIKC^C-type genes (e.g. in floral organ identity specification) appear to be largely conserved across flowering plants. For example, the mutant phenotypes of B-, C-, D- and E-class homeotic genes in grasses such as rice and maize (*Zea mays*), revealed basic conservation of the (A)BCE model, although it is not always readily apparent based on single-mutant phenotypes owing to the presence of multiple, largely functionally redundant paralogs, for example of C-class genes in rice (e.g. Dreni et al., 2011; for a detailed review on floral MIKC^C-type genes in grasses, see Ciaffi et al., 2011). The A class is the most debated and apparently least evolutionarily conserved homeotic function (Causier et al., 2010). Recent analysis of the function of SQUA subfamily genes from basal eudicots suggests that the 'A-function' evolved via subfunctionalization after gene duplication(s) at the base of core eudicots from a more broad action of SQUA subfamily members in floral meristem

specification, floral organ specification and fruit development (see Pabon-Mora et al., 2012 and references therein). Interestingly, the E-function appears to be not only exerted by genes from the SEP subfamily, but also from the closely related AGL6 subfamily at least in some flowering plant species, such as petunia (Vandenbussche et al., 2003b; Rijpkema et al., 2009), rice (Ohmori et al., 2009; Cui et al., 2010; Gao et al., 2010; Li et al., 2011) and maize (B. E. Thompson et al., 2009). This provides an indication that partial functional redundancy of members from different subfamilies may have persisted over long evolutionary time-scales. Future research needs to reveal how this apparent redundancy is reflected in the molecular action of the different genes.

Independent MIKC^{C}-type gene duplication events in the different flowering plant lineages can be associated with lineage-specific subfunctionalization (see **Glossary**, **Box 1**) or to a lesser extent, neofunctionalization (see **Glossary**, **Box 1**) of individual paralogs. The process of plant-lineage specific subfunctionalization after gene duplication is also exemplified by functionally equivalent paralogous homeotic C-function genes *AG* from *Arabidopsis* and *PLENA* (*PLE*) from *Antirrhinum* (Bradley et al., 1993). Their respective orthologs (see **Glossary**, **Box 1**), *FARINELLI* (*FAR*) in *Antirrhinum* (Davies et al., 1999) and *SHP1*,2 in *Arabidopsis* (Liljegren et al., 2000) have undergone independent subfunctionalization (Causier et al., 2005; Airoldi et al., 2010). Plant lineage-specific functional diversification of AG clade genes is also reflected in the evolution of their *cis*-regulatory regions (Causier et al., 2009; Moyroud et al., 2011).

A crucial aspect in the patterning of the floral meristem and regulation of homeotic gene expression is the restriction of C-class expression to the inner two floral whorls in the floral meristem and during organ development. Many factors regulating *AG* expression in *Arabidopsis* at the transcriptional level have been characterized (reviewed by Kaufmann et al., 2010b). It was shown that C-class repression in the outer whorls is mediated by mechanisms that differ somewhat in different eudicot species: in *Arabidopsis*, *AG* expression is among others regulated by the miRNA172/AP2 module, whereas in *Petunia* and *Antirrhinum*, a miRNA169/NF-YA module has a primary role in restricting the expression of the C-class genes *pMADS3* and *PLE*, respectively, to the inner floral whorls (Cartolano et al., 2007). In contrast to the *Arabidopsis* miR172, miR169 (which is encoded by the *BLIND* locus in *Petunia* and *FISTULATA* in *Antirrhinum*) has a repressive role in C-gene regulation, by repressing the activity of NF-YA genes that in turn activate C-class gene expression. A broad expression of miR169 is thought to translate into a threshold activation of C-class gene expression that induces a positive autoregulatory feedback. Conserved DNA-binding sites for NF-YA factors are also found in the *Arabidopsis AG* regulatory

intron, although the role of NF-YA genes in regulating *AG* expression is still not well understood (Hong et al., 2003).

Regulatory and protein-protein interactions among homeotic MADS-domain factors have also undergone changes during evolution. For example, the class-B floral homeotic genes encode closely related DEF-like and GLO-like MADS-domain transcription factors, which originated by a gene duplication event prior to the origin of angiosperms (reviewed by Becker and Theissen, 2003). DEF- and GLO-like proteins bind to DNA only as heterodimers in a number of flowering plant species especially core eudicots, but not as homodimers. Heterodimerization is therefore also required for a positive autoregulatory loop that is important for class-B homeotic gene function. The finding that these proteins have the ability to homodimerize in some flowering plant species and in gymnosperms led to the hypothesis that obligate heterodimerization of DEF- and GLO-like proteins arose from homodimerization (several times independently) during flowering plant evolution (Winter et al., 2002). Autoregulatory circuits of B-class proteins also partially diverged following more recent gene duplication events and differential gene loss (Lee and Irish, 2011), for example in Solanaceae (Rijpkema et al., 2006; Geuten and Irish, 2010) and the basal eudicot opium poppy (*Papaver somniferum*) (Drea et al., 2007).

Changes in homeotic gene expression in the different floral whorls have suggested a role for homeosis in the evolution of flower morphologies (reviewed by Hintz et al., 2006). Heterotopic expression of B-class genes in first whorl floral organs has been implicated in the formation of formation of petaloid tepals instead of sepals in tulips (Kanno et al., 2003), as proposed in the 'shifting boundaries' model (Van Tunen and Angenent, 1993). B-class gene duplications followed by functional divergence have also been implicated in the formation of different tepal types in the orchids (e.g. Chang et al., 2010; Mondragon-Palomino and Theissen, 2011). However, the evolution of petal-like sepals may not always involve shifts in B-class gene expression (Landis et al., 2012). In basal angiosperms, B-class genes in particular show broader expression in floral organs compared to more derived flowering plant lineages (Kim et al., 2005), which has been suggested to be linked with the gradual morphological intergradations often observed between adjacent floral organs in basal angiosperms [see the 'fading boundaries model' (Buzgo et al., 2004)]. It should be noted that it will be important in the future to complement comparative gene expression studies in evo-devo research with analysis of mutants of the respective genes in the studied species, because we know, for example from *Arabidopsis* that mRNA expression does not always reflect protein expression/function in certain organs or tissues, for instance, the B-class factor *AP3* is post-transcriptionally regulated (Jack et al., 1994).

Because of their role in the specification of male and female reproductive organs, B- and C-class MADS-box genes have also been implicated in the evolution of unisexual flowers. While the mechanisms underlying sex determination in dioecious plants are highly variable, in some species, such as *Thalictrum dioicum* and *Spinacia oleracea*, sex determination evolved by changes in regulation of B- and C-class gene expression (Di Stilio et al., 2005; Sather et al., 2010; reviewed by Diggle et al., 2011). Also the presence of B-class gene loci on X chromosomes in *Silene* species suggests a role in the evolution of unisexual flowers (Cegan et al., 2010).

Inflorescence architecture and transfer of functions

Changes in plant morphologies have been linked to the heterotopic expression of normally vegetatively expressed MIKC^C-type genes in flowers, or of floral homeotic MIKC^C-type genes outside the flower. For example, the study of petaloid bracts in the dove tree (*Davidia involucrata*) shows that petal identity can be partially transferred to organs outside the flower, such as bracts surrounding a contracted inflorescence with reduced flowers (Vekemans et al., 2012). In *Gerbera*, the *SEP1* ortholog *GERBERA REGULATOR OF CAPITULUM DEVELOPMENT2* (*GRCD2*) functions in inflorescence determinacy (Uimari et al., 2004) and controls inflorescence architecture (Teeri et al., 2006). SEP subfamily members also control the development of grass-specific spikelet meristems and thereby inflorescence development in grasses (Malcomber and Kellogg, 2004; Cui et al., 2010; Gao et al., 2010; Kobayashi et al., 2010). Another example of an MIKC^C-type factor with a role in controlling inflorescence architecture is the *VEG1* gene, which is an AGL79-like gene (SQUA subfamily) that controls secondary inflorescence meristem identity to generate a compound inflorescence in pea (Berbel et al., 2012).

Whereas some floral MADS-box genes have adapted novel roles outside the flower, others have frequently been recruited in evolution to functions in floral organ development. *INCOMPOSITA* (*INCO*), a member of the STMADS11 subfamily, whose members in *Arabidopsis* mostly control floral transition, represses the development of prophylls (extra flower organs) and therefore regulates floral architecture in *Antirrhinum* (Masiero et al., 2004). *MPF2*, another member of the STMADS11 subfamily in *Physalis floridana* (Solanaceae), has been shown to control the inflated-calyx syndrome, which is a morphological novelty in which sepals resume growth after pollination in order to protect the mature fruit (He and Saedler, 2005, 2007). Furthermore, gene duplication of *MPF2*-like genes followed by functional diversification at regulatory and protein levels can be linked to the complex evolution of sepal morphologies in Solanaceae (Khan et al., 2009).

Fruit development

Beyond their roles in floral organ specification, MIKC^C-type genes have also been recruited to control the development of various fruit morphologies and seed dispersal mechanisms in flowering plants, and therefore have also likely played a role during crop plant domestication. For example, *SHP1*,*2* (from the AG subfamily) in *Arabidopsis* specify the replum in the silique. By contrast, their tomato ortholog *TAGL1* controls fleshy fruit expansion and the ripening process (Itkin et al., 2009; Vrebalov et al., 2009; Gimenez et al., 2010).

Remarkably, members of the same subfamilies have been recruited to function in very different fruit types, for example in *Arabidopsis* (silique), *Solanum* and *Vaccinium* ('berry'), as well as in *Fragaria* (strawberry, which is botanically not a berry and is derived from the receptacle of the flower) or *Malus* (apple, a 'pome') (e.g. Cevik et al., 2010; Jaakola et al., 2010; Seymour et al., 2011). The strawberry *SEP1*,*2* ortholog *FaMADS9* (Vrebalov et al. 2002) has an important function in receptacle (and thereby fruit) development, and it also controls ripening programs during later stages of development (Seymour et al., 2011). Besides its role in flower development, the tomato *SEP1*,*2* ortholog *TM29* also functions in fruit development, since its downregulation results in the generation of parthenocarpic fruits (Ampomah-Dwamena et al., 2002). However, *TM29* is not reported to affect fruit ripening. By contrast, the tomato *SEP4* ortholog *RIPENING INHIBITOR* (*RIN*) is a key regulator of fruit ripening and controls climacteric respiration and ethylene biosynthesis (e.g. Vrebalov et al., 2002; Fujisawa et al., 2011; Martel et al., 2011).

The tomato *AP1* ortholog *MACROCALYX* (*MC*) (SQUA subfamily), a regulator of sepal size and inflorescence determinacy (Vrebalov et al., 2002), controls development of the pedicel abscission zone and thereby seed dispersal. The MC protein interacts with JOINTLESS (J), a member of the STMADS11 subfamily and a regulator of fruit abscission, to form a functionally active transcription factor complex (Nakano et al., 2012). The multiple roles of SQUA subfamily members in floral transition, axillary meristem growth, perianth identity and fruit development are already evident in the basal eudicot species California poppy and opium poppy (Pabon-Mora et al., 2012).

Transition to flowering

The evolution of MADS-box gene subfamilies that control the vegetative-to-floral transition appears to be highly dynamic and linked to the enormous complexity of life-history strategies in flowering plants ranging from ephemeral annuals to long-lived trees. An example is the STMADS11 subfamily, whose members evolved novel functions in reproductive transition alongside acquiring roles in flower and fruit development. An example is the series of tandem duplications in peach (*Prunus*

persica) that led to six DORMANCY-ASSOCIATED MADS-BOX (DAM) genes that are associated with floral bud dormancy, and thereby seasonal flowering, in this species (Jimenez et al., 2009; Z. Li et al., 2009). Also, *BpMADS4*, a member of the SQUA subfamily and ortholog of the uncharacterized *Arabidopsis AGL79*, has a role in the initiation of inflorescence development and the transition from vegetative to reproductive development in the silver birch tree (*Betula pendula*) (Elo et al., 2007).

Another subfamily of MIKC^{C}-type genes with a highly dynamic evolution is the FLC subfamily. *FLC*-like genes have been mainly identified as vernalization-controlled floral repressors in *Arabidopsis*, *Brassica* and in sugar beet (*Beta vulgaris*) (Michaels and Amasino, 1999; Tadege et al., 2001; Schranz et al., 2002; Reeves et al., 2007). Natural variation in *FLC* gene activity is associated with flowering time variation and differential vernalization response among ecotypes (see **Glossary**, **Box 1**) of *Arabidopsis* and related species (Schranz et al., 2002; Nah and Jeffrey Chen, 2010; Salome et al., 2011). Evolutionarily diverged regulation of *FLC* orthologs has been linked with the perennial life habit, such as *PERPETUAL FLOWERING 1* (*PEP1*) in *Arabis alpina* (Wang et al., 2009) and has also been observed in species with floral bud dormancy, for example *PtFLC* in trifoliate orange (*Poncirus trifoliate*) (Zhang et al., 2009) and *TrMADS3* in Rosaceae (*Taihangia rupestris*) (Du et al., 2008).

The origin and early evolution of major plant MADS-box gene lineages

Type I and type II MADS-box genes have been identified in all major land plant lineages, from bryophytes to flowering plants (Gramzow and Theissen, 2010). Importantly, the number and functional diversity of MADS-box genes increased considerably during land plant evolution, and is linked to the elaboration of plant body plans and life history strategies (Becker and Theissen, 2003; Kaufmann et al., 2005b; Kramer and Hall, 2005).

Land plants evolved from multicellular charophycean algae ~500 million years ago. The colonization of land was associated with the elaboration of the sporophytic (diploid) phase in the plant life cycle. MIKC-type MADS-box genes are found in land plants and charophycean algae, but not in other, more primitive, algae (Tanabe et al., 2005). Expression studies in charophycean algae suggest an ancestral role of MIKC-type MADS-box genes in haploid reproductive cell differentiation in the gametophytic phase (Tanabe et al., 2005). Prior to the origin of the most primitive extant land plants, the bryophytes, a gene duplication event led to the origin of MIKC^{C}-type and MIKC*-type genes (Henschel et al., 2002). In the moss *Physcomitrella patens*, MIKC^{C}-type genes function in the gametophyte as well as in specific tissues of the sporophyte, whereas MIKC*-type genes are specifically expressed in the gametophyte (Singer et al., 2007; Kwantes et al., 2012). This gametophytic expression appears to be highly conserved across land plant evolution and might reflect an ancestral, conserved role of

MIKC*-type genes in gametophyte development (Verelst et al., 2007a; Zobell et al., 2010; Kwantes et al., 2012).

MIKC^C-type MADS-box genes: the key to the origin of seeds and flowers?

The enigmatic origin and success of seed plants, and more recently of flowering plants (angiosperms) is one of the biggest evolutionary mysteries. Seed plants now constitute more than 90% of all land plant species, and by far the greatest diversity is seen in angiosperms, which comprise 250,000-400,000 species. Key to the success of seed plants was a major elaboration of reproductive organ morphologies, most markedly the origin of the seed and, in angiosperms, the origin of the bisexual flower. In addition, the elaboration of floral transition and plant architecture can be considered as major evolutionary innovations.

Extant seed plants, which comprise flowering plants and gymnosperms (see **Glossary**, **Box 1**), evolved from a most recent common ancestor ~300 million years ago. Many subfamilies of MIKC^C-type genes appear to have originated in ancestral seed plants (Becker and Theissen, 2003), and gene expression analyses suggest that the functions of some subfamilies might be conserved between angiosperms and gymnosperms. Examples are the homeotic AG (C/D class) and DEF/GLO (B class) subfamilies, as well as the B_{sister} subfamily (e.g. Tandre et al., 1995; Becker et al., 2002; reviewed by Becker and Theissen, 2003). Their important functions and conserved expression suggest roles in the origin and evolution of seed plant reproductive structures.

The seed represents a special type of heterospory in which the female gametophyte is protected by integuments that, after fertilization, allow the developing embryo to be retained and nourished on the mother plant. Interestingly, whereas B-class genes show conserved expression in male reproductive organs (and angiosperm petals), B_{sister} genes exhibit conserved expression in the evolutionarily most conserved parts of the ovule (Becker et al., 2002). The contrasting expression of B- and B_{sister}-class genes has led to the hypothesis that the origin of these subfamilies played an important role in the evolution of male and female reproductive structures in seed plants (Becker et al., 2002). B_{sister} genes control endothelium formation and later aspects of seed development in *Arabidopsis* (Nesi et al., 2002; Mizzotti et al., 2012), *Petunia* (de Folter et al., 2006) and rice (Yin and Xue, 2012), which supports a role for this subfamily in the evolution of the seed.

Another major innovation in seed plant evolution was the origin of the angiosperm flower, characterized by synorganization of female and male reproductive organs (Bateman et al., 2006). Given their important role in floral meristem formation, the SQUA and SEP subfamilies, which are only found in flowering plants (Becker and Theissen, 2003), could be the key to the origin of flowers. In addition,

concerted gene duplications linked to rounds of whole-genome duplications in different MIKC^C-type subfamilies prior to the origin of extant flowering plants, and at the base of core eudicots, may have contributed to the evolution of the floral bauplan (see Zahn et al., 2005; Shan et al., 2009). Genome sequences from extant gymnosperms are likely to reveal the full complement of MIKC^C-type genes outside flowering plants in the near future, and thereby shed light on the origin and early diversification of these genes in seed plant evolution.

Molecular mechanisms of action of MADS-domain proteins

Despite the wealth of information about the biological functions of plant MADS-domain proteins from genetic studies, we still do not fully understand their molecular mode of action. In the early 1990s it was shown that, in analogy to mammalian MADS-domain proteins, plant MADS proteins bind their consensus DNA binding site (the CArG-box, see **Glossary**, **Box 1**) as dimers (Schwarz-Sommer et al., 1992). Around this time, the yeast two-hybrid system was introduced as a method with which to study protein-protein interactions and, a few years later, evidence was provided for multiple interactions between *Antirrhinum* floral homeotic MIKC^C-type MADS-domain proteins (Davies et al., 1996). These initial studies were followed by large-scale MADS-domain protein interaction screenings in a variety of species, which provided information about MADS-domain protein dimerization potential (Immink et al., 2003; de Folter et al., 2005; Leseberg et al., 2008; Liu et al., 2010; Ruokolainen et al., 2010).

The next breakthrough in our understanding of MADS-domain protein function came from the finding that MIKC^C-type proteins can assemble into higher-order complexes (Egea-Cortines et al., 1999; Honma and Goto, 2001), which led to the postulation of the 'floral quartet' model (see **Box 2**). According to this model, a tetrameric protein complex consisting of two dimers binds to a target DNA sequence containing two CArG-boxes and thereby generates a DNA loop between the two binding sites (Theissen, 2001; Theissen and Saedler, 2001). Although the presence of two CArG-boxes may provide stability through cooperative DNA binding, heterotetrameric homeotic protein complexes can also bind to DNA sequences containing only one CArG-box, which may or may not contain additional 'weak affinity' binding sites (Melzer and Theissen, 2009; Smaczniak et al., 2012b). Members of the SEP subfamily play an important role as mediators of higher-order complex formation (Immink et al., 2009), and also at least some proteins from other subfamilies can also mediate higher-order complex formation (Egea-Cortines et al., 1999; Ciannamea et al., 2006). The K-domain in particular plays a role in the formation of higher-order complexes of MIKC^C-type proteins (Egea-Cortines et al., 1999; Honma and Goto, 2001; Yang and Jack, 2004; Melzer and Theissen, 2009),

and in some cases it contributes also to heterodimerization (Y. Yang et al., 2003). The K domain probably forms three amphipathic α-helices that may assemble into coiled-coil structures (reviewed by Kaufmann et al., 2005b). Large-scale yeast-based screenings showed that various *Arabidopsis*, tomato and *Gerbera* MIKC^{C}-type MADS-domain proteins have the capacity to multimerize (Leseberg et al., 2008; Immink et al., 2009; Ruokolainen et al., 2010), and ternary complexes consisting of type I proteins could also be identified (Immink et al., 2009). Floral homeotic B- and C-class MADS-domain proteins from the gymnosperm *Gnetum gnemon* have the ability to form higher-order protein complexes (Y. Q. Wang et al., 2010), suggesting that the requirement for angiosperm-specific SEP proteins in mediating higher-order complex formation among floral homeotic proteins is a derived state that evolved due to differential loss of the ability of B+C-class proteins to multimerize. Multimerization expands the number of potential and unique MADS protein transcription factor units and might be a key molecular mechanism in providing DNA-binding specificity. The latter hypothesis is supported by *in vitro* binding assays that show stabilized binding of DNA sequences containing two CArG-box elements by quaternary MADS domain protein complexes (Egea-Cortines et al., 1999; Melzer and Theissen, 2009; Smaczniak et al., 2012b).

Recent technological progress, such as sensitive mass spectrometry analysis, has allowed the isolation of MADS-domain protein complexes from plant tissues. A recent pioneering study (Smaczniak et al., 2012b) unveiled the composition of homeotic protein complexes on which the 'floral quartet' model is based. In addition to the expected identification of MADS domain protein interaction partners, corepressors, chromatin remodeling factors, and transcription factors from other families were identified as interaction partners. The identification of transcription factors from other families in the isolated complexes points towards a role for these transcription factor interactions in target gene selection. Previously, evidence was provided for the assembly of MADS protein complexes that includes the SEUSS and LEUNIG transcriptional corepressors (Sridhar et al., 2006). Physical interactions had also been reported between SVP, SOC1 and AGL24 with chromatin-associated factors that mediate gene repression. These factors include the polycomb PRC1 analog TERMINAL FLOWER 2 [TFL2; LIKE HETEROCHROMATIN PROTEIN 1 (LHP1)] are the SIN3 histone-deacetylase complex component SAP18 (Liu et al., 2009), and the interaction with these factors are proposed to play a role in compacting the chromatin at bound loci and thereby in transcriptional repression. These interactions presumably prevent premature activation of floral homeotic genes in inflorescence and early floral meristems (Liu et al., 2009). This repression may be overcome by interactions of AP1 and other floral homeotic proteins and chromatin remodelers. This hypothesis (**Figure 3**) is exemplified by the finding that SEP3

physically interacts with the SWI2/SNF2 ATPases BRAHMA (BRM) and SPLAYED (SYD), providing complexes that overcome polycomb-mediated repression of *AP3* and *AG* during early floral meristem development (Smaczniak et al., 2012b; Wu et al., 2012). The direct activation of the C2H2-type zinc-finger gene *KNUCKLES* (*KNU*) by AG has also been shown to be associated with release from repressive H3K27me3 chromatin states, and therefore provides another example for an interplay between MADS-box transcription factors and epigenetic regulators (Sun et al., 2009). In fact, a number of MIKC^C-type MADS-box genes are targets of polycomb-mediated repression, as indicated by the deposition of repressive H3K27me3 marks and ectopic activation in polycomb mutants (Goodrich et al., 1997; Turck et al., 2007; Zhang et al., 2007). This suggest that overcoming or enforcing repressive chromatin states may be an important mode of action in regulatory networks that are formed by MIKC^C-type proteins during developmental transition.

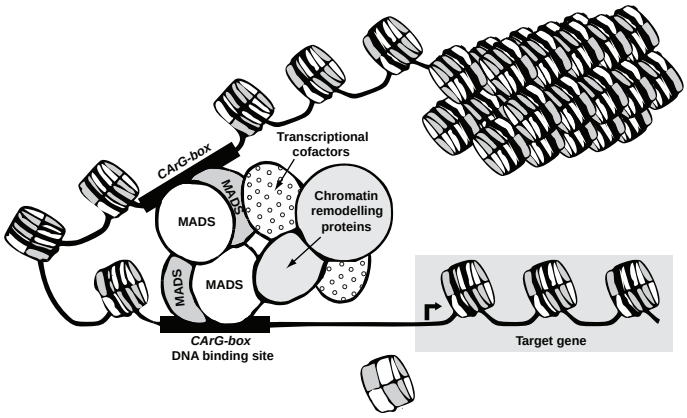


Figure 3. Model for the action of MADS-domain protein complexes. Shown is a model of MADS-domain protein complex formation and a hypothesized mechanism of regulatory action. In this model, MADS-domain proteins form quaternary complexes according to the 'floral quartet' model and interact with two DNA binding sites (CArG-boxes) in close proximity, resulting in DNA looping. Subsequently, MADS-domain proteins recruit transcriptional cofactors, which mediate transcriptional regulation and may influence target gene specificity, as well as chromatin remodeling proteins which relax the chromatin structure at the target gene transcription start site allowing for the initiation of transcription. Depending on the selection of transcriptional cofactors and chromatin remodeling factors, the complex may also play a role as a transcriptional repressor.

A combination of genome-wide expression analysis and ChIP followed by deep sequencing (ChIP-Seq) or hybridization to microarrays (ChIP-CHIP), has revealed genes, and hence biological processes, that are directly controlled by MADS-domain transcription factors. These experiments showed that the MIKC^C-type proteins bind hundreds to thousands of loci. Analysis of the target gene sets for the floral repressor FLC (Deng et al., 2011) and the homeotic proteins SEP3 and AP1 (Kaufmann et al., 2009; Kaufmann et al., 2010c) revealed a large number of genes involved in transcriptional and cellular signaling, for example hormonal regulation. Among the FLC targets, various genes involved in abscisic acid (ABA) signaling were identified,

which could be related to the role of FLC in temperature-dependent germination (Chiang et al., 2009). Among the potential direct SEP3 target genes, auxin response genes attracted attention and could be related to the role of SEP3 in floral organ outgrowth and morphogenesis (Kaufmann et al., 2009). The current data suggest that floral homeotic MADS-domain proteins directly regulate the expression of a variety of genes that are important for growth, shape and structure of different organs, indicating that floral MADS domain proteins not only specify organ identity at the onset of organ primordial initiation, but are also involved in subsequent differentiation processes (reviewed by Dornelas et al., 2011; Ito, 2011).

The data also reveal complex regulatory interactions among MADS family members, and the existence of a large number of positive and negative (auto)regulatory loops. Negative feedback loops are required for developmental phase switches, and have been hypothesized to be important for MADS-box gene function during the transition from vegetative to reproductive growth (Yu et al., 2004; de Folter et al., 2005), while feed-forward loops are important for robust and balanced expression of target genes. The non-MADS transcription factor LFY is, for example, involved in activation of the MADS-box gene *SEP3*, and in turn, both LFY and SEP3 are essential for the activation of the MADS-box genes *PI*, *AP3* and *AG* (reviewed by Wagner, 2009). Positive (auto)regulatory loops involving two partners, for example, can facilitate a stable upregulation and maintenance of gene expression, as is the case for the B-type MADS-box genes (Schwarz-Sommer et al., 1992; Lenser et al., 2009) and for *AGL24* and *SOC1* (Liu et al, 2008).

The spatiotemporal activity of MADS-domain proteins is not only regulated at the transcriptional level, and a few examples of posttranslational modifications affecting MADS-domain protein function have been described. Wang and colleagues (Y. Wang et al., 2010) demonstrated the phosphorylation-dependent prolyl *cis/trans* isomerization of AGL24 and SOC1, and showed that this modification affects the stability of AGL24 in the nucleus. Furthermore, transport of (at least some) MADS proteins from the cytoplasm to the nucleus appears to be regulated (see He and Saedler, 2007) and for some type II and type I MADS-domain proteins dimerization was shown to be essential for translocation to the nucleus (e.g. McGonigle et al., 1996; Bemer et al., 2008). Additionally, intercellular transport could be shown for a few selected MADS domain proteins from different species (Perbal et al., 1996; Sieburth et al., 1998; Urbanus et al., 2010), providing an additional mechanism for spatial control of their activity.

Conclusions

In the past 20 years, a tremendous knowledge of plant MADS-domain transcription factors has been generated. We have also obtained a better understanding

of previously overlooked lineages of MADS-box genes, such as the type I and MIKC*-type genes. MADS-box genes have been shown to play roles in a variety of developmental processes and a surprising number of them have more than one function in seemingly unrelated processes. Future research should address the issue of how such apparently different functions of the same MADS-box gene, for example in the shoot and in the root, relate to each other. This could also help us to understand the evolutionary mechanisms by which MADS-box genes are recruited to new functions in other species.

Functional redundancy might have hampered the assignment of functions to some genes, but we also need a better understanding of what 'redundancy' really means, for example by characterizing molecular phenotypes and analyzing natural variation in gene regulatory networks in more depth. This holds for the exploration of type I as well as for type II genes. The recent finding that AG clade MIKC^{C}-type genes have a role in lateral root initiation in addition to their well-known function in reproductive development also emphasizes that we might need to employ more systematic and comprehensive approaches in the characterization of mutant phenotypes. MIKC^{C}-type genes, in particular, are involved in evolutionarily highly dynamic developmental processes, such as control of flowering time. Analyzing the natural variation in regulatory networks formed by MIKC^{C}-type genes is therefore likely to provide new insights into the dynamics and significance of specific regulatory interactions, and this approach might unveil gene functions that are not obvious from the analysis of only one specific ecotype. A classic example in this respect is the finding that *FLC* is dependent on the *FRIGIDA* locus, of which different alleles are present in the ecotypes with strongly varying flowering times (Johanson et al., 2000).

Although recent studies have revealed functions of some type I and MIKC*-type genes, most remain to be characterized, especially in species other than *Arabidopsis*. The current data suggest that these genes are important regulators of gametophytic and embryo development in plants. Therefore, understanding the evolution of these MADS-box gene functions may also help us to gain more insights into essential aspects of plant reproductive processes.

Recent results have also provided insights into the molecular mechanisms by which plant MADS-domain transcription factors recognize and control the expression of their target genes. MIKC-type proteins, and possibly also type I MADS-domain proteins, form complex protein interaction networks. But how do MADS-domain proteins obtain their functional specificity? The first genome-wide DNA-binding studies of MADS-domain proteins (Kaufmann et al., 2009; Zheng et al., 2009; Kaufmann et al., 2010c; Deng et al., 2011) revealed a large number of binding sites and potential direct target genes. Even proteins that act at different developmental stages show at least some overlap in DNA binding sites. This could indicate that these

factors control overlapping sets of target genes and achieve their regulatory specificity by whether they activate or repress expression. Target gene activity would then be controlled by different MADS-domain factors that compete for common binding sites. It is also conceivable that common target genes might be responsible for general cellular processes, whereas the distinct target genes might be specific for a particular biological or developmental process. Understanding the specificity of target gene regulation by MADS-domain proteins will be a challenge for future research. The consequences of DNA binding for spatial promoter organization, including the formation of DNA loops, also need to be considered here.

MADS-domain proteins form complex intrafamily interaction and regulatory networks. MADS-box gene expression appears to be regulated at many levels: transcriptionally, post-transcriptionally and post-translationally (e.g. protein localization). Advanced proteomics and *in vivo* imaging approaches can be used to systematically study the regulation of MADS-box transcription factor activities *in planta*. In addition, the modeling of MADS-box regulatory networks can provide novel insights (Espinosa-Soto et al., 2004; van Mourik et al., 2010), but will require more quantitative *in vivo* data in the future.

Finally, a number of studies have shown that many MADS-box genes have roles in more than one organ or developmental stage. How can the same factor have different functions in different developmental contexts? And how can apparently conserved proteins control diverse organ morphologies, such as flower development? In order to address these questions, we need to understand the developmental and evolutionary dynamics of regulatory networks formed by MADS-domain transcription factors. This will provide insight into the recruitment of MADS-domain proteins during the origin of morphological innovations and, thereby, help us to understand the morphological diversity of flowering plants.

Acknowledgements

We apologize to authors whose publications could not be cited owing to space constraints; given the large body of MADS-box gene research in plants, we were restricted to examples illustrating aspects of MADS-box gene function and evolution, focusing on recent publications.

Funding

C.S. received funding from the Netherlands Proteomics Centre (NPC); R.G.H.I., G.C.A. and C.S. received funding from the Centre for BioSystems Genomics (CBSG); K.K. received funding from a Netherlands Organization for Scientific Research (NWO) VIDI project and a HORIZON Breakthrough project;

and R.G.H.I. and K.K. received funding from the European Research Area in Plant Genomics (ERA-PG) BLOOMNET project.

Aim and outline of the thesis

The aim of the research described in this thesis was to study the physical interactions of MADS-domain transcription factors and their functional consequences. Protein-protein and protein-DNA interactions form the molecular basis of gene regulatory networks. By studying these interactions, complex gene and protein regulatory networks were unraveled, shedding light on the molecular mechanisms controlling flower development in *Arabidopsis*. The results open new avenues for future research.

The main research questions that are addressed in this thesis:

1. Are protein complexes as suggested in the 'floral quartet' model formed in the flower?
2. Are there other specific interaction partners of MADS-domain proteins that mediate the transcriptional regulation during flower development?
3. What determines the specificity of protein-DNA interactions of MADS-domain transcription factor complexes?
4. What are the molecular mechanisms by which MADS-domain proteins act in *Arabidopsis*?

Chapter 1 reviews functions of MADS-domain transcription factors in flowering plants, with a main focus on *Arabidopsis*, where functions for nearly half of the MADS-box gene family members have already been described. The functional evolution of MADS-box genes, which may contribute to morphological diversification in plants, is illustrated. Furthermore, a hypothetical model of MADS-domain protein action that combines higher-order protein complex formation and active chromatin remodeling by large transcriptional machineries is suggested.

Chapter 2 describes the *in vivo* composition of MADS-domain protein complexes that are active in *Arabidopsis* flower development. By applying a targeted proteomics approach, the MADS-domain protein interactome is unraveled. These characterized interactions shed light on the combinatorial mode of action of MADS-domain transcription factors and strongly supports a mechanistic link between MADS-domain proteins and chromatin remodeling factors.

Chapter 3 reviews recent advances in proteomics approaches used to study cellular signaling and developmental processes in plants. The emerging roles of the characterization of whole proteomes as well as the description of entire cellular signaling cascades and transcriptional regulatory pathways in plants by new proteomics techniques are illustrated.

Chapter 4 comprehensively describes the protein immunoprecipitation protocol that was used in Chapter 2 to identify *in vivo* MADS-domain protein complexes. The main characteristics of this method are the use of fluorophore-tagged, single step affinity purification of protein complexes and label-free mass spectrometry-based protein quantification to distinguish true complex partners from non-specifically precipitated proteins.

Chapter 5 and **6** aim to characterize molecular mechanisms of DNA sequence recognition by MADS-domain transcription factors. These chapters address the intriguing questions, whether various MADS-domain protein complexes possess different DNA-binding specificities and which are the molecular features of different DNA-binding specificities of MADS-domain transcription factors.

The thesis is concluded in **Chapter 7**, where the main findings are discussed and future perspectives in research on plant MADS-domain proteins are given.

Chapter 2

Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development

Cezary Smaczniak[1,2,3], Richard G.H. Immink[2,3], Jose M. Muiño[1,2], Robert Blanvillain[4,5,6,7], Marco Busscher[2], Jacqueline Busscher-Lange[2], Q.D. (Peter) Dinh[2,3], Shujing Liu[8], Adrie H. Westphal[9], Sjef Boeren[9], François Parcy[4,5,6,7], Lin Xu[8], Cristel C. Carles[4,5,6,7], Gerco C. Angenent[2,3], Kerstin Kaufmann[1,2]

[1] Laboratory of Molecular Biology, Wageningen University, 6708PB Wageningen, The Netherlands
[2] Business Unit Bioscience, Plant Research International, 6708PB Wageningen, The Netherlands
[3] Centre for BioSystems Genomics, 6708PB Wageningen, The Netherlands
[4] Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5168, F-38054 Grenoble, France
[5] Université Joseph Fourier-Grenoble I, Unité Mixte de Recherche 5168, F-38041 Grenoble, France
[6] Institut de Recherches en Technologies et Sciences pour le Vivant, Laboratoire Physiologie Cellulaire et Végétale, F-38054 Grenoble, France
[7] Institut National de la Recherche Agronomique, F-38054 Grenoble, France
[8] National Laboratory of Plant Molecular Genetics, Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China
[9] Laboratory of Biochemistry, Wageningen University, 6703HA Wageningen, The Netherlands

Proc. Natl. Acad. Sci. USA **109**, 1560-1565 (2012)

Abstract

Floral organs are specified by the combinatorial action of MADS-domain transcription factors, yet the mechanisms by which MADS-domain proteins activate or repress the expression of their target genes and the nature of their cofactors are still largely unknown. Here, we show using affinity purification and mass spectrometry that five major floral homeotic MADS-domain proteins (AP1, AP3, PI, AG and SEP3) interact in floral tissues as proposed in the 'floral quartet' model. *In vitro* studies confirmed a flexible composition of MADS-domain protein complexes depending on relative protein concentrations and DNA sequence. *In situ* bimolecular fluorescent complementation assays demonstrate that MADS-domain proteins interact during meristematic stages of flower development. By applying a targeted proteomics approach we were able to establish a MADS-domain protein interactome that strongly supports a mechanistic link between MADS-domain proteins and chromatin remodeling factors. Furthermore, members of other transcription factor families were identified as interaction partners of floral MADS-domain proteins suggesting various specific combinatorial modes of action.

Introduction

Flower development is one of the best understood developmental processes in plants. According to the classic ABC model (Coen and Meyerowitz, 1991), floral organs in the model plant species *Arabidopsis* are specified by the combinatorial activity of three functional gene classes. The A class genes represented by *APETALA1* (*AP1*) and *APETALA2* (*AP2*) specify sepal identity, and together with B class genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), they determine the identity of petals. The C class gene *AGAMOUS* (*AG*) alone determines carpel identity and together with B class genes it specifies stamen identity. The ABC model was extended to the ABCE model, in which E class genes [*SEPALLATA1-4* (*SEP1-4*)] are required for the specification of all four types of floral organs (Pelaz et al., 2000; Ditta et al., 2004). Based on genetic and yeast *n*-hybrid protein interaction data it was later proposed in the 'floral quartet' model that floral organs are specified by combinatorial protein interactions of ABCE-class MADS-domain transcription factors, which are thought to assemble into organ-specific quaternary protein complexes that bind to two CArG-boxes, DNA consensus sequence CC[A/T]$_6$GG), in regulatory sequences of target genes (Honma and Goto, 2001; Theissen and Saedler, 2001). E-class proteins have a special role in this model as major mediators of higher-order complex formation. Although interactions that were predicted in this model were further supported by additional *in vitro* DNA-binding assays and protoplast FRET-FLIM experiments

(Immink et al., 2009; Melzer and Theissen, 2009; Melzer et al., 2009), formation and composition of these complexes in endogenous tissues remained unknown.

Heterologous interaction studies in yeast and genetic data suggest recruitment of transcriptional coregulators such as SEUSS (SEU) and LEUNIG (LUG) by floral MADS-domain proteins (Sridhar et al., 2006). Ovule-specific MADS-domain protein complexes were found to form higher-order interactions with BELL1 (BEL1), a member of the homeobox family of transcription factors, in a yeast-based screen (Brambilla et al., 2007). Also, interactions between other plant MADS-domain proteins and proteins which are functionally analogous to polycomb group (PcG) proteins, as well as putative components of histone-deacetylase complexes, have been reported, suggesting that these types of interactions play a role in the activity of the transcriptional regulatory complexes (Hill et al., 2008; Liu et al., 2009; Kaufmann et al., 2010b). Unraveling the *in planta* interactome of floral homeotic MADS-domain proteins could, therefore, advance our understanding of the mechanism and specificity underlying target gene regulation by these proteins.

In this study, we identified MADS-domain protein complexes by immunoprecipitation followed by mass spectrometry (MS) and label-free quantification. Our results indicate that MADS-domain proteins interact not only with each other but also with non-MADS transcriptional regulators. Chromatin remodeling and modifying factors represent the most prominent group among these interactors.

Results

Floral homeotic MADS-domain proteins interact in floral tissues

To analyze interactions of floral homeotic MADS-domain proteins in floral tissues, we made use of transgenic plant lines that express the MADS-domain proteins AP1, AG, AP3 and SEP3 from their native promoters linked to GREEN FLUORESCENT PROTEIN (GFP) as C-terminal fusions (de Folter et al., 2007; Urbanus et al., 2009). Protein complexes were isolated by immunoprecipitation using anti-GFP antibodies and characterized by liquid chromatography (LC)-MS/MS followed by label-free protein quantification analysis. This approach allowed us to identify proteins that were enriched in IP samples compared to control samples and provide an approximation of their relative abundance. Our results confirmed all major protein interactions proposed in the 'floral quartet' model (**Figure 1A**). We identified the class B floral homeotic proteins AP3 and PI as major interaction partners of each other and found them in similar abundance in the IP samples. Also putative higher-order complex partners, such as SEP3 (E-class), AP1 (A-class), and AG (C-class) were identified as interaction partners of AP3 and PI. SEP3, which acts as a major mediator of higher-order complex formation (Immink et al., 2009), appears to be the most

abundant interaction partner of class B proteins. SEP3 was also identified as interaction partner of AP1 and AG, whereas its paralog SEP4 was only detected as interaction partner of AP1 and FRUITFULL (FUL), supporting its predominant role in MADS-domain protein complexes that act during floral initiation and sepal development (Ditta et al., 2004).

Using SEP3-GFP as bait, fruit- and ovule-specific MADS-domain proteins, namely SHATTERPROOF1,2 (SHP1,2) and SEEDSTICK (STK), were identified in addition to the ABC floral homeotic protein classes. This supports the proposed role of higher-order MADS-domain protein complexes in ovule identity specification, referred to as 'D-class' function (Colombo et al., 1995). Stamen and carpel complex partners, such as AG and B-class proteins, were more strongly represented than AP1 when using SEP3 as bait. This could reflect the abundance of certain complexes in the inflorescence tissues that were sampled, where the largest relative amount of tissue corresponds to later stages of floral organ differentiation. In the AG-GFP IP, an almost equal amount of AP3/PI and AG proteins were enriched, although one should expect less AP3/PI interacting with AG because of the formation of the carpel identity complexes. This could reflect differences in complex stability, tissue sampling, efficiency of elution from the bait protein in the IP or estimation of protein levels.

Using AP1-GFP as bait, we also identified a lowly abundant interaction with SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), which is a major regulator of floral transition (Samach et al., 2000). In addition, SOC1 was identified as major interaction partner of FUL, supporting the existence of a FUL/SOC1 protein complex acting in floral transition. Using FUL-GFP as bait, we also identified several floral homeotic proteins, in particular AP1 and SEP proteins, as FUL interaction partners. In contrast to endogenous *FUL*, some expression of the FUL-GFP transgene has been observed in stage 1 and 2 floral buds, and later in whorl 2 and 3 (Urbanus et al., 2009), which might explain the observed interactions of FUL and floral MADS-domain proteins. Remarkably, using AP1 or FUL as bait, most tagged protein appeared not to be present in a heteromeric complex, because interaction partners are far less abundant than the bait protein (**Figure 1A**). This could reflect presence of these proteins in a homodimeric or monomeric form, or a low stability of heteromeric complexes for these proteins during the biochemical isolation procedure. Although AP1 and FUL fusions to GFP can complement the respective mutant phenotypes (Wu et al., 2003; Urbanus et al., 2009), it remains possible that the level of transgenic AP1- and FUL-GFP is elevated or stabilized compared with that of endogenous protein. This could potentially result in an overrepresentation of these proteins relative to their interaction partners in these transgenic lines.

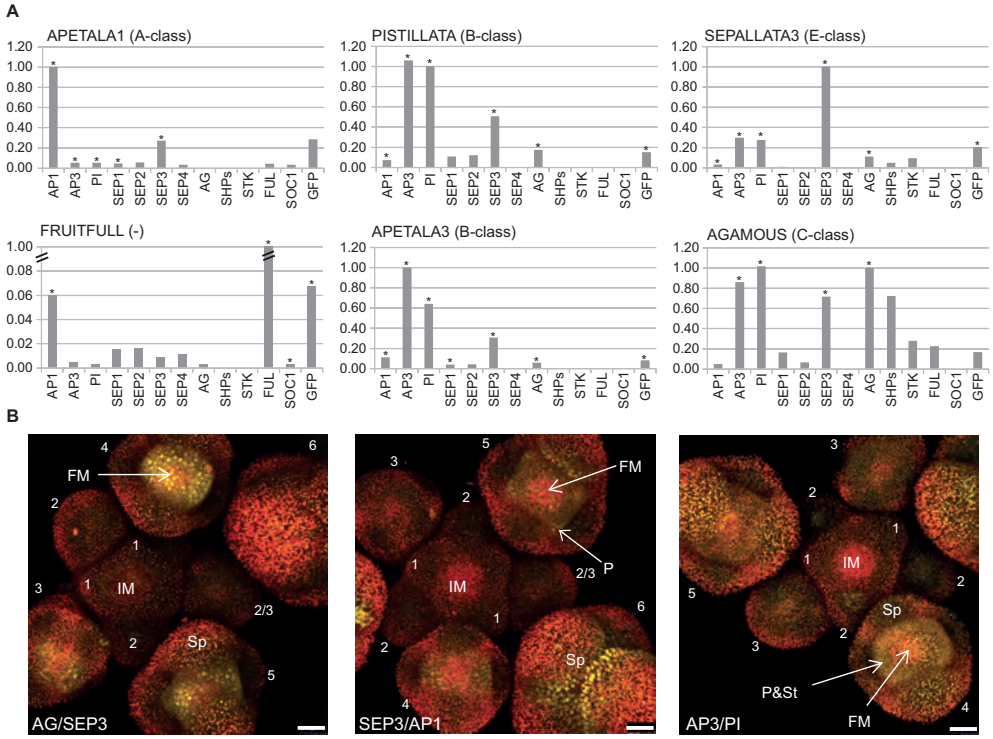


Figure 1. *In planta* MADS-domain protein interactions. (A) Average MADS-domain protein abundance ratios between the IP samples and the control samples scaled to the ratio of the bait protein. Ratios calculated based on 4–5 most abundant and unique peptides of a particular protein identified by LC-MS/MS are marked with an asterisk. Ratios calculated based on three or fewer identified peptides were not marked. **(B)** 3D maximum projections of in situ BiFC data using MADS-domain proteins expressed from their own promoters, confirming the interactions between MADS-domain proteins in floral meristems. Left: pAG:AG-eYFP/ N + pSEP3:SEP3-eYFP/C. The yellow spots are characteristic of the nuclear localized interaction signal. The signal is positioned in the FM center where stamens and carpels will arise. Center: pSEP3:SEP3-eYFP/N + pAP1:AP1-eYFP/C. Most YFP signal is located in sepal tips and at the edges of the FM from where petals will be formed. Right: pAP3:AP3- eYFP/N + pPI:PI-eYFP/C. Weak YFP signal is found in the meristematic domain giving rise to petal and stamens, which is characteristic for PI and AP3 protein expression patterns (**Figure S1 J** and **K**). 1–6, flower bud stages; FM, flower meristem; IM, inflorescence meristem; P, petal initiation site; Sp, sepal; St, stamen initiation site. Scale bars, 25 µm.

To obtain detailed spatial information on *in planta* interactions of MADS-domain proteins, we applied the bimolecular fluorescence complementation (BiFC) assay (Walter et al., 2004) using MADS-domain proteins expressed from their endogenous promoter and fused to either the N-terminal or C-terminal half of enhanced YELLOW FLUORESCENT PROTEIN (eYFP). Using this method we confirmed the interactions of SEP3 and AG, SEP3 and AP1, and AP3 and PI in floral meristems (**Figure 1B** and **Figure S1 A-I**). The interactions were mainly detected at

stages of meristem development when floral organ identities are initially specified. While AG/SEP3 and AP1/SEP3 heterodimers showed preferentially nuclear localization, the AP3/PI heterodimer shows an even distribution throughout the whole cell.

Formation of quaternary MADS-domain protein complexes on the DNA

It is still not well understood how heteromeric higher-order MADS-domain protein complexes assemble and associate with their target DNA. To date, only DNA-binding homotetrameric and quartet-like complexes consisting of a SEP3 homodimer and AP3/PI heterodimers have been reconstituted *in vitro* (Melzer and Theissen, 2009; Melzer et al., 2009).

We identified a regulatory region in the *SEP3* promoter that was bound *in planta* by several MADS-domain proteins such as AP1, SEP3, FUL and AG (**Figure 2A** and **Figure S2 A** and **B**) (Kaufmann et al., 2009; Kaufmann et al., 2010c), and chose this region to study the DNA-binding of higher-order MADS-domain protein complexes. The distal *SEP3* promoter region between -2.6 to -3.1 kb containing these MADS transcription factor binding sites is required for the positive autoregulation of *SEP3* in an inducible system and triggers enhancement of expression in floral tissues (**Figure S2 C-G**), and is also bound by AP1 during early floral meristematic stages (Kaufmann et al., 2010c).

Two pairs of CArG-boxes (pairs named '2' and '3') were identified to be located closest to the site of maximum ChIP enrichment (**Figure 2A**), of which CArG-box pair '3' showed the strongest binding of MADS-domain proteins *in vitro* (**Figure 2B** and **Figure S3A**). We choose fragment '3' containing a CArG-box pair (CArG 3 and CArG 3') for further analysis. AG, SEP3, and AP1 proteins bind as homodimers to this sequence, as does the AP3/PI heterodimer (**Figure 2B**). When the SEP3 protein was incubated with either AG or AP1, we observed the predominant formation of DNA-binding heteromeric higher-order complexes, which were abolished when using a truncated SEP3 protein (SEP3ΔC) missing the C-terminus and the last α-helix of the K-domain that is involved in higher-order complex formation (**Figure 2B**). Weak bands corresponding to higher-order complexes were visible in the presence of either SEP3, SEP3ΔC or AG protein (marked with asterisks in **Figure 2B**), which could arise from two MADS dimers binding separately to two DNA-binding sites on this probe. Next, we analyzed the DNA-binding of heteromeric higher-order complexes consisting of SEP3, AP3 and PI together with either AP1 (petal specification) or AG (stamen specification). We noticed that two bands were present in the shift corresponding to tetrameric complexes, indicating that at least two different higher-order complexes can potentially be formed on this DNA sequence in the presence of four different MADS-domain proteins (**Figure 2C**). The composition of these

complexes was analyzed by protein titration experiments. The results show that SEP3 is required for both complexes to form. AG is present in the upper complex, whilst the

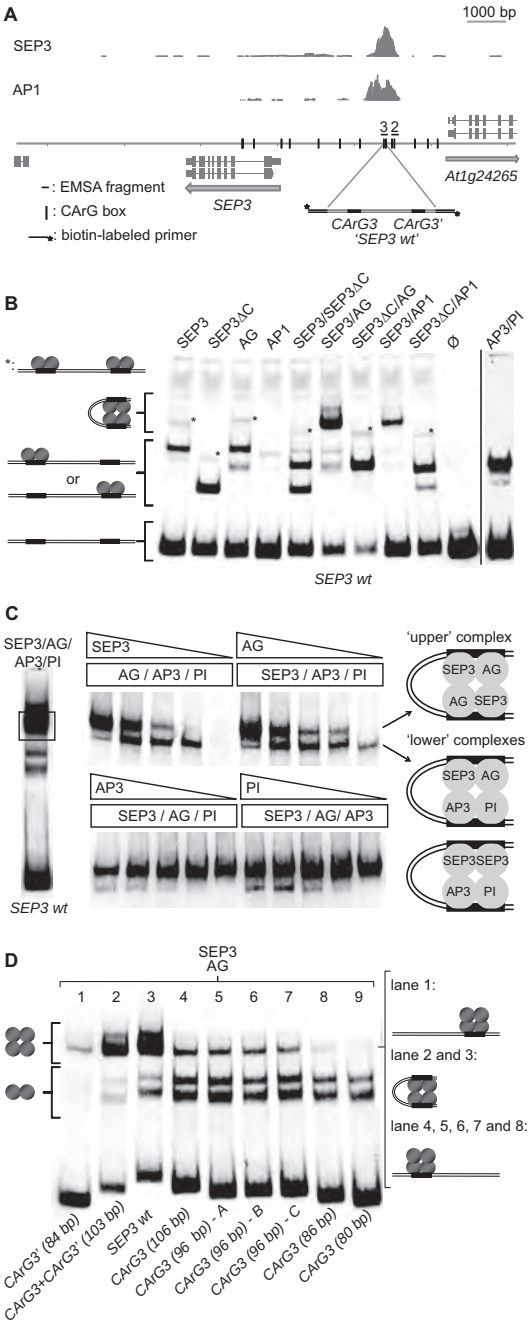


Figure 2. Assembly of the MADS-domain protein complexes in a distal region of the *SEP3* promoter. (A) Graphic representation of the *SEP3* locus with a 4.1-kb promoter region and the SEP3 and AP1 ChIP-SEQ profiles. Fragments used in the EMSA experiments were flanked with the biotin primers used for amplification and detection. Vertical lines in the sequence map indicate position of the CArG-boxes. **(B)** EMSA of the different MADS-domain protein complexes with the *SEP3* wild-type promoter fragment and possible model representations of formed protein-DNA complexes. **(C)** Left: EMSA of the SEP3/AG/AP3/PI protein mix with the '*SEP3 wt*' DNA fragment containing two CArG-boxes. Center: EMSAs where the concentration of a single protein component was gradually reduced from approximately equimolar amounts to 0. Only the part of the gel containing the slow migrating complexes (rectangle in the left EMSA) is shown. Right: Model representation of the higher-order protein complexes formed in the presence of SEP3, AG, AP3, and PI binding to the *SEP3* promoter fragment *in vitro*. **(D)** EMSA of the SEP3/AG protein mix with the truncated versions of the *SEP3* wild-type DNA fragments. The *SEP3 wt* fragment was shortened from both 3′ and 5′ ends and contains either a single or double binding site. CArG3 (96 bp) – A, CArG3 (96 bp) – B, and CArG3 (96 bp) – C are different, randomized versions of the 3′-end flanking region of the CArG3 fragment.

lower band may correspond to two complexes, one with and one without the AG protein. AP3 and PI proteins are present in the lower complexes because decrease of their concentrations affects only these complexes. When all proteins are present in similar concentrations, the stronger DNA-binding SEP3/AG/SEP3/AG and the weaker SEP3/AG/AP3/PI and/or SEP3/SEP3/AP3/PI complexes are

formed (**Figure 2C**). Furthermore, we also observed formation of higher-order complexes consisting of SEP3 and AP1, as well as SEP3, AP1, AP3 and PI on this *SEP3* promoter element (**Figure S3B**). These results suggest that MADS-domain protein complexes with different composition can coexist within a cell, and may compete for interaction partners and DNA-binding sites.

To evaluate the roles of the two CArG-boxes in recruiting higher-order complexes, we generated DNA probes where the sequence was gradually shortened. We found that the presence of only one CArG-box was sufficient to recruit heteromeric higher-order MADS-domain protein complexes; however a minimum length of DNA sequence is required (in this case ~85 bp) (**Figure 2D**). This result indicates that additional non-sequence specific DNA contacts stabilize binding of higher-order complexes to the DNA in the presence of only one CArG-box, which supports and extends a previous finding (Melzer et al., 2009).

MADS-domain proteins act together with nucleosome remodelers and other transcriptional regulators

Gel filtration experiments performed on nuclear protein extract demonstrated that SEP3 is part of a large protein complexes of around 670 kDa, which is far beyond the molecular weight of a MADS heterotetramer (**Figure 3A**). Therefore, we analyzed which non-MADS proteins were enriched in the nuclear MADS immunoprecipitates by LC-MS/MS and label-free quantification (**Datasets S1** and **S2**). Among the proteins that were consistently enriched in the IP datasets of all MADS-domain proteins we found several classes of nucleosome-remodeling factors, as well as RELATIVE OF EARLY FLOWERING 6 (REF6), recently characterized as histone H3K27 demethylase (Lu et al., 2011) (**Table 1**). This suggests that MADS-domain proteins can recruit or redirect the basic chromatin remodeling machinery to modulate the promoter structure of their target genes. Selected interactions were confirmed by reciprocal complex isolation and co-immunoprecipitation (**Figure S4**). The notion that MADS-domain transcription factors recruit the nucleosome remodeling machinery to target gene promoters via more flexible, but in some cases less stable interactions is supported by the finding that interactions of PI with CHROMATIN REMODELING 4 (CHR4) and CHR11/17 are stabilized by the presence of DNA (**Figure S4**).

We also identified previously characterized interaction partners of MADS-domain proteins, the transcriptional coregulator SEU, as well as its interaction partner LEUNIG-HOMOLOG (LUH) (Sridhar et al., 2006) (**Table 1**). Next to basic transcriptional regulators, we identified members of several other transcription factor families as potential MADS interaction partners. AUXIN-RESPONSE FACTOR 2 (ARF2) and SQUAMOSA PROMOTER BINDING PROTEIN LIKE 8 (SPL8)

were among the proteins that were enriched in AP1 (and AG) IP samples (**Table 1**). Analysis of the ChIP-SEQ data of AP1 identified the enrichment of the ARF binding motif (**Figure S5**), which is also enriched in SEP3 ChIP-SEQ peaks (Kaufmann et al., 2009). In addition, we found that the DNA-binding motif of SPL8 (Birkenbihl et al., 2005) was enriched in the AP1 and SEP3 ChIP-SEQ peaks, suggesting that they assemble into complexes that bind to nearby sites in the same genomic region (**Figure S5**).

Table 1. List of potential interaction partners enriched in the MADS-GFP IP experiments.

	AG-GFP IP		**AP3-GFP IP**		**PI-GFP IP**		**SEP3-GFP IP**		**AP1-GFP IP**	
Protein name	Log2 Ratio	Peptide number	Log2 Ratio	Peptide number	Log2 Ratio	Peptide number	Log2 Ratio	Peptide number	Log2 Ratio	Peptide number
Nucleosome associated factors										
PKL	-	-	-	-	-	-	**2.67**	**4**	-	-
CHR4	**3.12**	**8**	**1.78**	**5**	**2.27**	**5**	**3.47**	**7**	**3.82**	**14**
SYD	0.71	2	-	-	-	-	1.17	2	**3.1**	**5**
BRM	0.65	2	0.17	2	-	-	1.05	3	**2.51**	**4**
CHR11	**2.32**	**19**	**2.09**	**17**	**1.79**	**17**	**2.32**	**19**	**3.38**	**25**
CHR17	**2.8**	**17**	1.38	19	**2.71**	**16**	**2.29**	**19**	**3.67**	**24**
INO80	1.06	2	-	-	-	-	-	-	**3.43**	**7**
REF6	**2.34**	**4**	0.87	3	-	-	**3.22**	**2**	**3.55**	**5**
General transcriptional coregulators										
LUH	-	-	-	-	-	-	**2.93**	**2**	**5.62**	**6**
SEU	0.28	2	-	-	-	-	-	-	**1.61**	**2**
Transcription factors										
KNAT3	-	-	-	-	-	-	-	-	**4.29**	**2**
BLH1	-	-	-	-	-	-	-	-	**2.34**	**3**
BLR	-	-	-	-	-	-	-	-	**1.69**	**3**
ARF2	-	-	-	-	-	-	-	-	**2.28**	**3**
SPL8	**3.04**	**3**	-	-	-	-	0.78	2	**1.69**	**2**

All protein enrichment values (log2 ratio) that showed significant differences at False Discovery Rate (FDR) 0.01, except for the AP3-GFP IP, where the FDR threshold was 0.05 because of the higher variability within samples and controls, are bolded. For the results of the detailed statistical analysis with the Student's *t*-test *P values*, see **Dataset S2**.

Also the homeodomain transcription factors BELLRINGER (BLR), KNOTTED-LIKE 3 (KNAT3) and BEL1-LIKE HOMEODOMAIN 1 (BLH1) were identified as complex partners of AP1. Because interactions between BELL-like and KNOTTED-like proteins have been found in yeast-two-hybrid (Y2H) experiments (Hackbusch et al., 2005), our data suggest the formation of larger complexes consisting of MADS and homeodomain transcription factors. Targeted yeast-3-hybrid (Y3H) experiments with a selected set of MADS-domain protein dimers revealed that mainly KNAT3, and to a lesser extent BLR and BLH1, is found as a direct interaction partner of floral MADS dimers AP1/SEP4 and AP1/SEP3 (**Figure S6A**).

Based on genetic data, BLR was previously shown to regulate meristem maintenance, as well as internode, flower, and fruit development (Bao et al., 2004; Lal et al., 2011). Together with the closely related factor POUND-FOOLISH (PNF), it controls floral evocation by regulation of *LEAFY* (*LFY*), *AP1* and other factors (Kanrar et al., 2008). BLR also represses *AG* in floral and inflorescence meristems, acting synergistically with the general corepressors *LUG* and *SEU* (Bao et al., 2004). Because of the related functions of BLR and AP1 and their coexpression in floral meristems, we used targeted ChIP of BLR on selected AP1 binding sites to test whether AP1 and BLR may regulate flower initiation by binding to common sites in the genome, possibly as part of a protein complex. Indeed, we found that BLR and AP1 binding sites overlap in the regulatory regions of several genes that control floral transition and meristem specification such as the *LFY*, *AP1*, *AP2* and TARGET OF EARLY ACTIVATION TAGGED (EAT) 1 (TOE1) [at least threefold enrichment of BLR-GFP ChIP in 7 out of 11 tested AP1-bound regions (Kaufmann et al., 2010c)] (**Figure S6C**). We also confirmed the interaction of BLR and AP1 by protein complex isolation experiments using BLR as a bait (**Figure S6B**).

Next we analyzed the expression patterns of plants expressing promoter:gene-GFP fusions of several potential MADS interactors. All showed expression in developing flower meristems or at later stages of flower differentiation (**Figure S6D**). The nucleosome remodelers BRAHMA (BRM) and CHR17, as well as REF6 and the other chromatin-associated proteins are broadly expressed throughout floral meristems, suggesting that they achieve their functional specificity through recruitment to target gene promoters by transcription factors, such as MADS-domain proteins.

Biological roles of interactions between MADS-domain proteins and chromatin-associated factors

We identified the H3K27me3 demethylase REF6, as well as nucleosome remodelers, as protein complex partners of floral MADS-domain proteins, suggesting that MADS-domain proteins regulate transcription by modulating chromatin structure and accessibility. We therefore tested local H3K27me3 distribution at DNA regions bound by MADS-domain proteins, using the *SEP3* genomic locus as an example (**Figure 3B**). We studied the H3K27me3 distribution at the *SEP3* promoter and genomic loci before and after induction of the AP1-GR fusion protein in *ap1 cal* background. *SEP3* is one of the earliest genes directly activated by AP1, first weakly 8 hours after AP1 induction, and more strongly after 2 days (Wellmer et al., 2006; Kaufmann et al., 2010c). Surprisingly, no change in H3K27me3 status associated with gene activation is detectable within the first *SEP3* intron (**Figure S7**), whereas in contrast, we observed a clear reduction in the level of H3K27me3 in the distal

enhancer element and less pronounced in the proximal promoter (**Figure 3C** and **D**). These results suggest that AP1-mediated activation of *SEP3* is (initially) associated with removal of H3K27me3 in the *SEP3* promoter. Because *SEP3* is also a target of the H3K27me3 demethylase REF6 (Lu et al., 2011), which is an interaction partner of AP1, it is tempting to speculate that AP1 can redirect or enhance REF6 activity at the *SEP3* promoter.

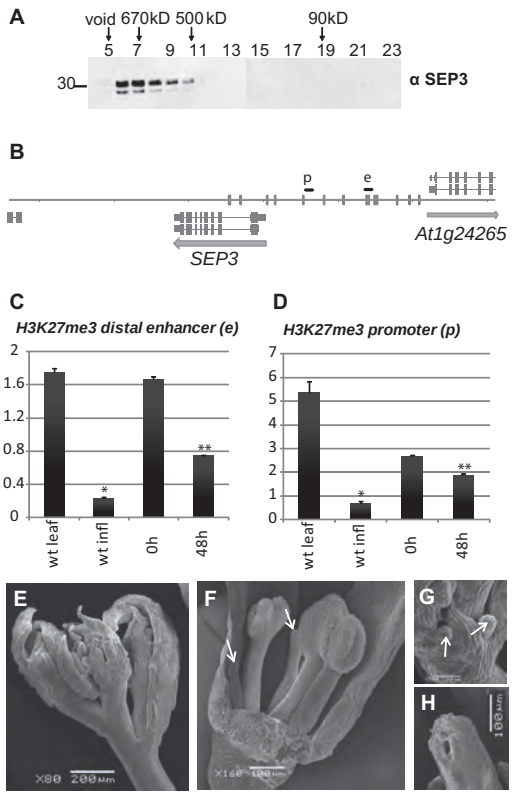


Figure 3. Interactions between MADS-domain transcription factors and other transcriptional regulators. (A) Gel filtration reveals that SEP3 is present in large nuclear complexes. **(B)** *SEP3* promoter and genomic locus representation with the quantitative PCR fragments in the distal enhancer site (e) and weaker proximal promoter site (p). Fragments were designed according to ChIP-SEQ profiles of AP1 and SEP3 (see **Figure 2A**). Vertical bars indicate CArG-box sequences. **(C)** Enrichment analysis of H3K27me3 at the MADS binding site in the distal SEP3 enhancer (e). ChIP was analyzed by quantitative PCR; material was obtained from inflorescence tissue of 35S:AP1-GR *ap1 cal* before (0 h) or 48 h after dexamethasone treatment and then subjected to ChIP with antibodies specific to H3K27me3. Results are presented as fold enrichment of input chromatin. Graphs represent average values from triplicates. Error bars represent SE of the mean. Asterisks indicate values that are significantly different from wild-type leaves (*) or from untreated 35S:AP1- GR *ap1 cal* plants (**) ($P < 0.05$ using Student t test). **(D)** Enrichment analysis of H3K27me3 in the proximal SEP3 promoter (p). For both C and D, H3K27me3 signal is reduced 48 h after AP1 induction compared with signal in 35S:AP1-GR *ap1 cal* uninduced tissues. **(E–H)** Scanning electron microscopy (SEM) pictures of chr11 chr17 double mutant inflorescences. **(E)** Overview of an inflorescence showing aberrations in floral organ development. **(F)** Close-up of a dissected *chr11 chr17* flower (sepal in front was removed) with malformed stamens and petals replaced by pin-like structures (see arrow). **(G)** Close-up of a developing *chr11 chr17* flower showing outgrowth of pin-like structures that replace the petals. **(H)** Incompletely closed carpel.

The functions of SWI/SNF-type chromatin remodelers BRM and SPLAYED (SYD), as well as the CHD-type remodeler PICKLE (PKL), in the regulation of flower and carpel development have been characterized previously (Eshed et al., 1999; Wagner and Meyerowitz, 2002; Hurtado et al., 2006; Bezhani et al., 2007; Aichinger et al., 2009). In contrast, no flower-specific functions of the ISWI-type nucleosome remodelers CHR11 and CHR17 have been described so far. We, therefore,

investigated flower phenotypes of *chr11 chr17* double mutants and found pleiotropic phenotypic alterations: sepals were abnormally curled and longer comparing to other organs (**Figure 3E**), petals and stamens were replaced by pin-like structures or were significantly reduced in size, and carpels did not fuse completely (**Figure 3 F-H**). These floral morphogenetic defects correlate with a function for CHR11 and CHR17 in the MADS complexes.

Discussion

Specificity of DNA binding and mechanisms of gene regulation by transcription factors can depend on recruitment of cofactors to specific regulatory DNA sequences. Here, we showed that a well-known class of transcription factors, the MADS-domain proteins interact not only with each other, as proposed in the 'floral quartet' model (Theissen and Saedler, 2001), but form large complexes with other types of transcriptional regulators *in planta*, shedding light on mechanisms by which MADS-domain proteins regulate the transcription of their target genes.

According to the current model, SEP proteins are major mediators of higher-order complex formation of MADS-domain proteins. Our complex isolation results suggest some functional diversification within the SEP subfamily, which is partly supported by genetic data (Ditta et al., 2004) and the results of yeast *n*-hybrid assays (Ditta et al., 2004; Immink et al., 2009). The A-class gene *AP1* does not only specify the identity of the outer two floral whorls but also plays a role in the switch from inflorescence to floral meristem identity, in a partially redundant fashion with the two related genes *CAULIFLOWER* (*CAL*) and *FUL* (Ferrandiz et al., 2000a). The presence of SOC1 and FUL in the AP1 IP may reflect the role of AP1 in *Arabidopsis* floral meristem specification. *AP1* and *SOC1* are only transiently coexpressed around stage 2-3 of flower development (Samach et al., 2000). AP1 has also been shown to repress *SOC1* in the two outer floral whorls (Liu et al., 2007). This supports a role for heterodimers formed by antagonistically acting MADS-domain proteins in the transition from inflorescence to floral meristem identity as has been suggested previously (de Folter et al., 2005). Several other MADS-domain proteins, such as SHORT VEGETATIVE PHASE (SVP), are also binding partners of AP1 according to Y2H studies (de Folter et al., 2005), but they were not detected in our AP1-GFP IP experiments, perhaps because of the very low abundance of these proteins in the native inflorescence tissues that were used in our analysis and their limited overlap in expression with AP1.

Based on our *in vitro* EMSA studies, we propose that different heteromeric MADS-domain protein complexes can coexist within the nucleus and may compete for partly overlapping sets of DNA-binding sites. The observation that one CArG-box is sufficient to recruit a heteromeric higher-order MADS-domain protein complex

suggests a mechanism by which these protein complexes might be recruited to DNA target sites *in vivo*: a preformed higher-order MADS-domain protein complex may first bind to a single, accessible CArG-box in a target gene promoter. Then, upon bending of the DNA (West et al., 1998), the second heterodimer present in the complex may bind to another CArG-box in the vicinity, which stabilizes the binding of the MADS-domain proteins to DNA. This would suggest a more 'active' role of MADS-domain protein complexes in creating DNA loops in native promoters.

While the quaternary complexes that we reconstituted *in vitro* may represent 'core' complexes, we found that floral MADS-domain proteins are part of large complexes or structures *in planta*. In addition to MADS-domain proteins, we also identified members of other transcription factor families as potential components of MADS-domain protein complexes. This suggests that MADS-domain proteins may also act in a combinatorial fashion with non-MADS transcriptional regulators. Most prominent were members of the homeobox transcription factor family. Homeobox transcription factors form complex, intra-family interaction networks (Hackbusch et al., 2005). Therefore, the interaction between MADS-domain protein complexes and individual homeodomain proteins may recruit other members of the family to target gene promoters. Future experiments using more specific plant material for complex isolation might result in a more sensitive detection of additional interactions between MADS-domain proteins and non-MADS transcription factors that may cooperate in the regulation of subsets of target genes.

In the complex isolation experiments, we confirmed a previously identified interaction between AP1 and the transcriptional corepressor SEU (Sridhar et al., 2006). We also identified the SEU interaction partner LUH, which acts in a partially redundant manner with LUG (Sitaraman et al., 2008). In addition, we identified several types of ATP-dependent nucleosome remodelers and their interaction partners in complexes of MADS-domain proteins, possibly as part of larger complexes that are stabilized in the presence of DNA. Chromatin-associated proteins were particularly abundant in the AP1 IP. This could reflect an interaction of AP1 with other proteins to reorganize chromatin structure in target gene promoters during the switch from inflorescence to floral meristem identity. One possible role for the interaction between AP1 and nucleosome remodelers could be in the activation of other floral homeotic genes, because, for example the SWI/SNF-type chromatin remodeler BRM and PKL have previously been shown to play a role in this process (Hurtado et al., 2006; Aichinger et al., 2009). The presence of SYD in the AP1 IP suggests that it can interact with AP1 in activation of *LFY* and supports the theory of mechanistic control of MADS-domain proteins target genes by modification of the chromatin states (Wagner and Meyerowitz, 2002). The interaction of AP1 and other floral MADS-domain proteins with the H3K27me3 demethylase REF6 also suggests a role in the

modification of specific chromatin states, specifically in antagonizing PcG mediated transcriptional repression. This is further supported by the finding of specific reduction of H3K27me3 around MADS DNA-binding sites in the *SEP3* promoter upon AP1 induction. The defects in flower development that are observed in mutants of chromatin remodelers support the finding that chromatin-associated factors act together with MADS-domain transcription factors to control flower initiation and differentiation. Examples are phenotypes of *brm* (Hurtado et al., 2006), *pkl* (Eshed et al., 1999) and *chr11 chr17* mutants, as well as phenotypes of overexpression of the H3K27me3 demethylase REF6 and the *ref6 curly leaf* (*clf*) double mutant (Lu et al., 2011). The timed activation of the *KNUCKLES* (*KNU*) gene by AG via modification of chromatin states may be another example of interaction between MADS-domain transcription factors and chromatin remodelers (Samach et al., 2000).

To summarize, our results show that MADS-domain proteins associate with other transcription factors and chromatin-associated proteins into larger structures. Future experiments need to reveal the roles of specific complexes in the selection of target genes and thereby specification of distinct floral organ identities. They also need to reveal how common interactions between DNA sequence specific transcription factors and the nucleosome remodeling machinery are in plants.

Materials and Methods

High-resolution LC-MS/MS of protein immunoprecipitates and quantitative data analysis with the MaxQuant software were essentially described before (Hubner and Mann, 2011). Chromatin immunoprecipitation experiments were in general performed as described previously (Kaufmann et al., 2010a). Detailed experimental and data analysis procedures are provided in SI Materials and Methods.

Acknowledgments

We thank Giuseppa Morabito for help with the generation of the BiFC constructs and Frank Wellmer for the 35S:AP1-GR *ap1 cal* seeds. The authors were supported by grants from the Netherlands Proteomics Centre and Centre for BioSystems Genomics (to C.S.); the European Research Area Network on Plant Genomics Project 'The meristematic regulatory network controlling the floral transition' (BLOOM-NET) (to K.K. and R.G.H.I.); Netherlands Genomics Initiative Horizon Grant 93519020 (to J.M.M.); Netherlands Organization for Scientific Research VIDI grant (to K.K.) the French National Agency Young Researcher grant for the ChromFlow project (to C.C.C. and R.B.); and Centre National de la Recherche Scientifique Higher Education Chair Program (to C.C.C.).

Supporting Information (SI)

SI Materials and Methods

SI Materials and Methods are available upon request or at http://www.pnas.org/ website.

SI Figures

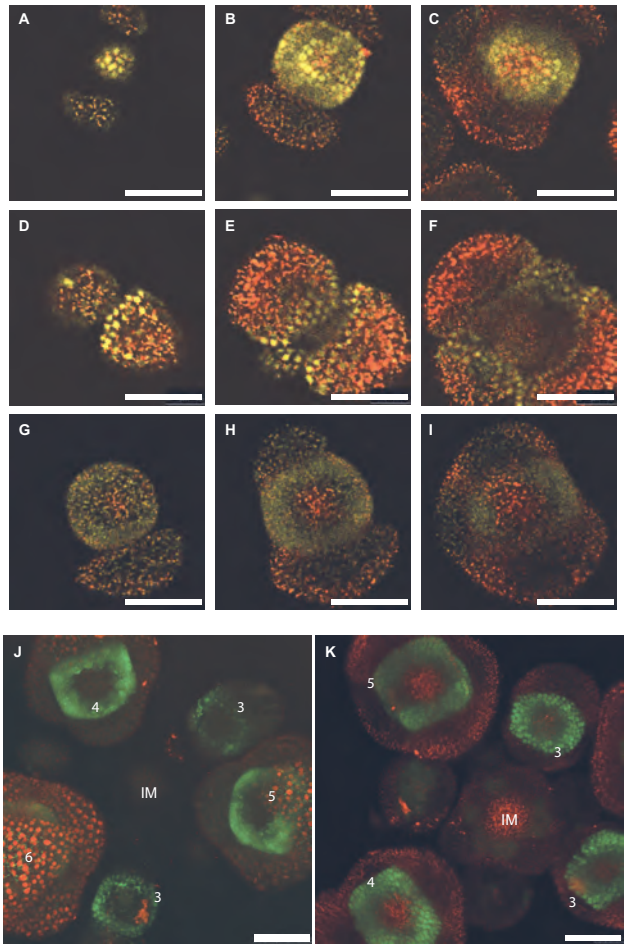


Figure S1. Confocal microscopy. (A-I) The magnified in situ BiFC data of MADS-domain proteins expressed from their own promoters in young *Arabidopsis* flower buds. **(A-C)** pAG: AG-eYFP/N + pSEP3:SEP3-eYFP/C. **(D-F)** pSEP3:SEP3-eYFP/N + pAP1:AP1-eYFP/C. **(G-I)** pAP3:AP3-eYFP/N + pPI:PI-eYFP/C. **(A, D,** and **G)** Individual layers taken from the upper part of the Z-stack. **(B, E,** and **H)** Individual layers taken from the middle part of the Z-stack. **(C, F,** and **I)** Individual layers taken from the bottom of the Z-stack. **(J-K)** GFP localization of pAP3:AP3-GFP (0.9-kb promoter) **(J)** and pPI:PI-GFP (1.4-kb promoter) **(K)** in floral meristems of different stages. GFP signal is indicated in green, chloroplast and other 'background' signal are indicated in red. IM, inflorescence meristem; numbers indicate floral stages according to Smyth et al., 1990. (Scale bars, 50 μm.)

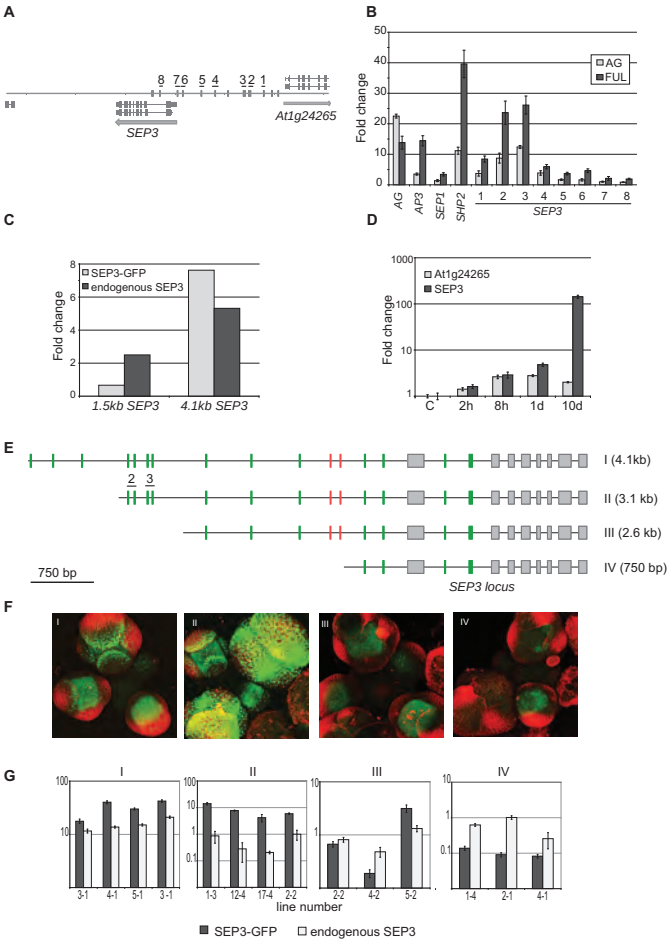


Figure S2. Regulatory elements in the *SEP3* promoter. (A) Structure of *SEP3* promoter and genomic locus with CArG-boxes indicated as vertical bars. Numbers indicate primer pairs used for ChIP-qPCR in B. **(B)** Results of ChIP qPCR of AG-GFP and FUL-GFP at the *SEP3* promoter/first intron. Also, binding of these two proteins to sites in other homeotic gene loci was detected. All of those sites are also bound by SEP3 according to ChIP-SEQ experiments, and the binding pattern of AG and FUL in the SEP3 promoter is similar to that of SEP3 and AP1 (see **Figure 2A**). **(C)** Autoactivation of SEP3 by a SEP3-GR fusion protein in seedlings requires elements in the distal SEP3 promoter. Expression of a transgene expressing SEP3 from a 1.5-kb endogenous promoter fragment cannot be induced by SEP3-GR, in contrast to a transgene with the 4.1-kb promoter (1-d induction). **(D)** Not only SEP3, but also a neighboring locus potentially sharing regulatory elements in their promoters can be activated by induction of SEP3-GR in seedlings. Experimental conditions are as described before (Kaufmann et al., 2009) **(E–G)** SEP3 promoter deletion studies by confocal analysis and qPCR. Reporter constructs were generated as C-terminal fusions to GFP. Confocal image analysis indicates that a 750-bp construct is not sufficient to recover the full meristematic spatiotemporal expression. A minimum of 1.5-kb promoter largely recovers the spatiotemporal expression pattern of endogenous SEP3 (Urbanus et al., 2009). The constructs with 4.1- and 3.1-kb promoters show enhanced expression of the SEP3-GFP transgene in the endogenous SEP3 expression domain in floral meristems. Three to four representative transgenic T1 lines were chosen for qPCR studies. Error bars indicate the SE of the mean.

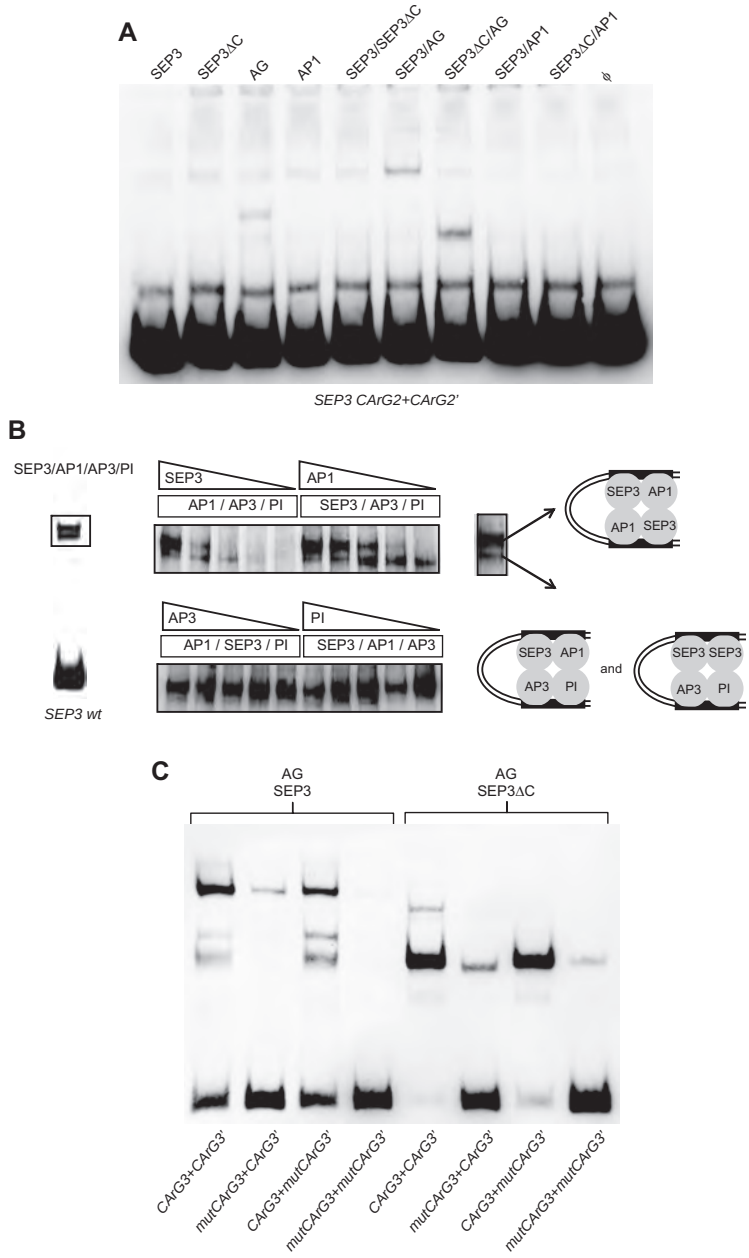
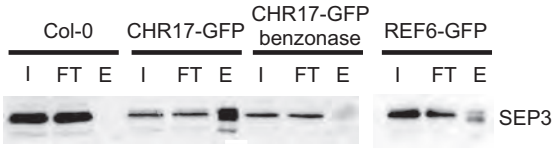


Figure S3. Gel retardation assays. (A) Analysis of binding of homo- and heterodimers to the second CArG-box pair present in the distal ChIP-SEQ peak in the *SEP3* promoter. Only binding of AG homodimer, SEP3/AG quaternary complex, and SEP3ΔC/AG dimer was detectable. **(B)** Higher-order complexes that are formed at the '*SEP3 wt*' fragment (*CArG3 and 3'*; see **Figure 2**) in the presence of AP1, SEP3, AP3, and PI. Titration experiments were performed in the same way as in **Figure 2**. **(C)** Formation of quaternary SEP3/AG and dimeric SEP3ΔC/AG complexes at mutated versions of the '*SEP3 wt*' fragment.

A



B

	CHR17-GFP IP		REF6-GFP IP	
Protein name	Log2Ratio	Peptide number	Log2Ratio	Peptide number
Floral MADS-domain proteins				
AP1	3.28	3	3.79	3
AP3	6.40	7	7.41	7
PI	5.24	3	5.96	4
SEP3	3.85	4	5.34	6
AG	-	-	-	-
Nucleosome associated factors				
CHR11	3.70	45	3.81	19
CHR17	7.46	65	3.43	20
REF6	4.60	6	9.80	47
PKL	3.86	4	-	-
CHR4	6.13	13	-	-
SYD	-	-	-	-
BRM	1.34	3	-	-
INO80	5.59	9	2.83	3

C

	PI-GFP IP		PI-GFP IP (Benzonase/EtBr)	
Protein name	Log2Ratio	Peptide number	Log2Ratio	Peptide number
PKL	-	-	-	-
CHR4	2.27	5	-0.18	3
SYD	-	-	-	-
BRM	-	-	2.08	3
CHR11	1.79	17	1.68	23
CHR17	2.71	16	1.57	23
INO80	1.06	2	-	-
REF6	-	-	2.52	2
LUH	-	-	3.9	3
SEU	-	-	-	-

Figure S4. Confirmation of interaction for selected protein complex partners. (A) Co-immunoprecipitation using inflorescences of CHR17- and REF6-GFP lines, probed with the SEP3 antibody. Stability of the interaction of CHR17 and SEP3 in the absence of longer DNA fragments was tested by benzonase treatment, and IP efficiency was found to be reduced. E, eluate (proteins eluted from the magnetic beads); FT, flow through; I, input (crude nuclei protein extract). **(B)** Results of LC-MS-based complex isolation of CHR17-GFP and REF6-GFP. MADS proteins were confirmed as complex partners of CHR17 and REF6. Identification of other nucleosome remodelers, especially in the CHR17-GFP IP, suggests the formation of larger structures. **(C)** Comparison of PI-GFP complex isolation in the presence and absence of DNA [ethidium bromide (EtBr)/benzonase treatment]. EtBr was used to release protein complexes from the DNA and benzonase to degrade all forms of DNA and RNA.

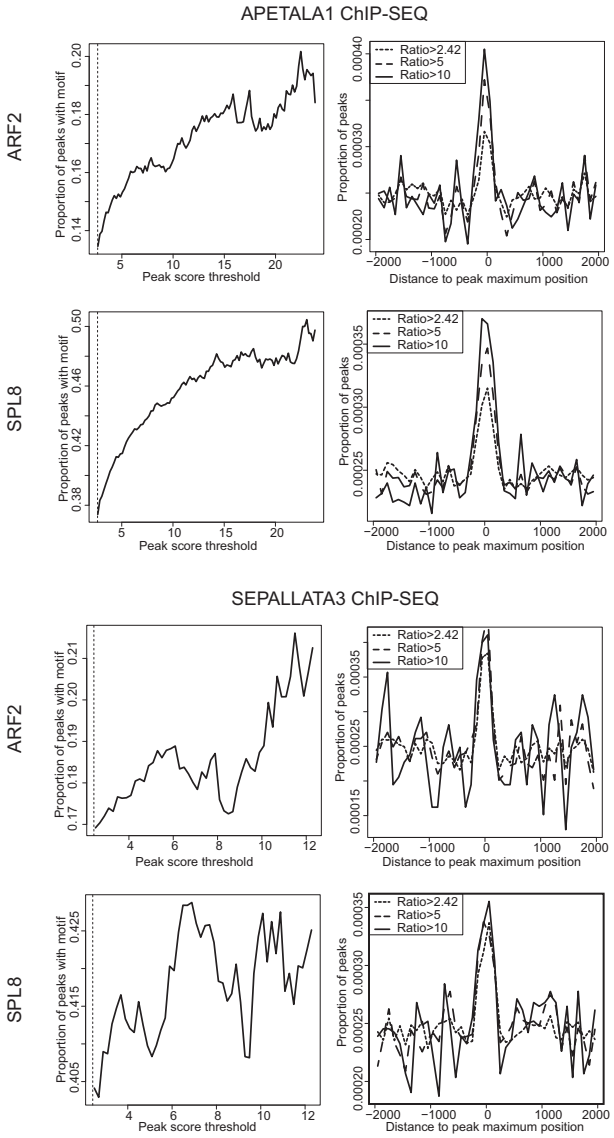
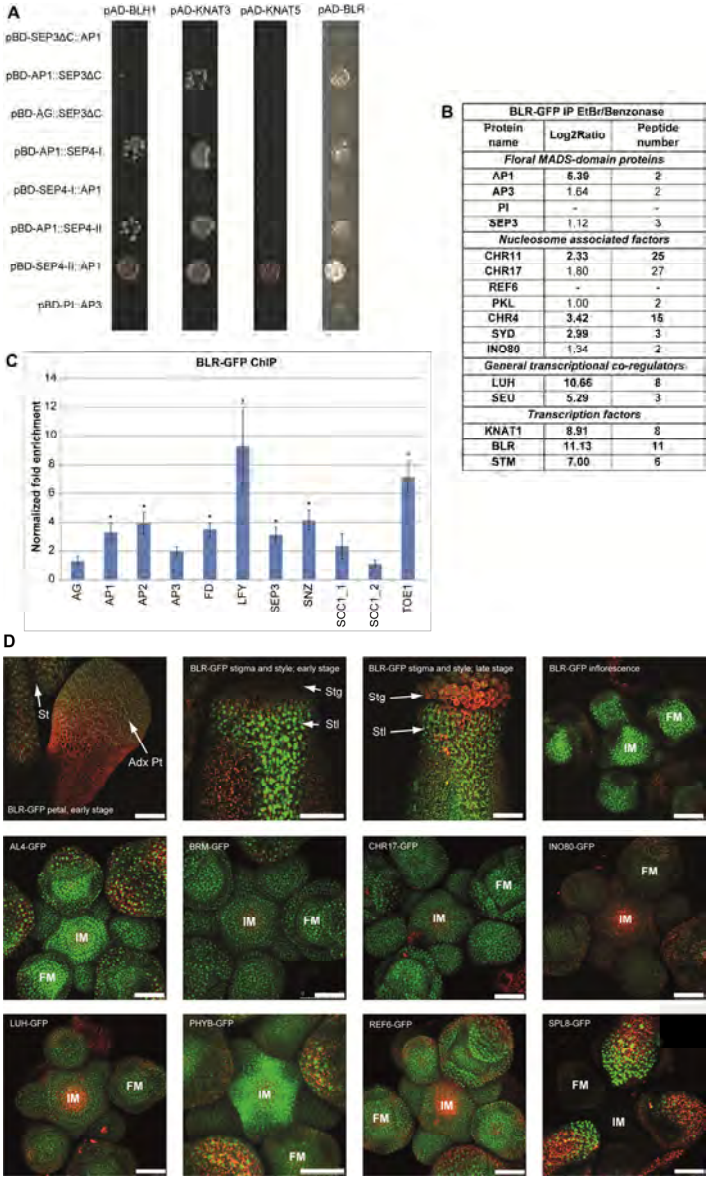


Figure S5. Enrichment of ARF and SPL8 DNA-binding motifs in SEP3 and AP1 ChIP-SEQ peaks. Left: For each significant peak, the nucleotide sequence 100 bp around the position of the maximum peak score location were extracted and associated with the peak score value. To obtain the proportion of peaks with a given DNA-binding site consensus (ARF2: 'TGTCTC'; SPL8: 'GTAC') at a given peak score threshold level, the proportion of sequences with at least one DNA-binding site consensus was calculated. Right: The proportion of the distance of the DNA-binding site consensus to the peak score location was calculated as the distance from the center position of the DNA consensus to the peak score location for each peak with a score bigger than the corresponding threshold at a given FDR level. All graphs were generated with R software; enrichment was calculated with its package 'Biostrings'.



BLR-GFP iP EtBr/Benzonase		
Protein name	Log2Ratio	Peptide number
Floral MADS-domain proteins		
AP1	6.39	2
AP3	1.64	2
PI	-	-
SEP3	1.12	3
Nucleosome associated factors		
CHR11	2.33	25
CHR17	1.80	27
REF6	-	-
PKL	1.00	2
CHR4	3.42	15
SYD	2.99	3
INO80	1.94	2
General transcriptional co-regulators		
LUH	10.66	8
SEU	5.29	3
Transcription factors		
KNAT1	8.91	8
BLR	11.13	11
STM	7.00	6

Figure S6. Interaction and expression of selected MADS-domain protein complex partners. (A) Y3H experiment reveals direct interaction of AP1/SEP4 and AP1/ SEP3ΔC dimers with KNAT3 and, to a lesser extent, with BLH1. Interactions between MADS-domain protein dimers and KNAT5 were included as negative interaction controls. **(B)** Results of LC-MS/MS-based complex isolation of BLR-GFP. AP1 was confirmed as a complex partner of BLR. Also, other known and unknown interaction partners of BLR were significantly enriched. Ethidium bromide was used to release protein complexes from the DNA and benzonase to degrade all forms of DNA and RNA. **(C)** BLR-GFP ChIP analyzed by qPCR in four technical replicates; material was obtained from inflorescence tissue of pBLR: BLR-GFP and subjected to ChIP with antibodies specific to GFP.

Legend continued on following page

Figure S6. Legend, continued.
Results are presented as fold enrichment of input chromatin (more than three-fold enrichment is marked with an asterisk). Error bars represent SE of the mean. **(D)** The expression patterns of pBLR:BLR-GFP (6.1-kb promoter) on various stages of flower development. BLR is predominantly expressed in the center of IM and FM (whorl 2, 3, and 4), in the style of gynoecium and in the petal tip. The expression patterns of other potential interaction partners of MADS-domain proteins: pAL4:AL4-GFP (2.8-kb promoter), pBRM:BRM-GFP (1.8-kb promoter), pCHR17: CHR17-GFP (0.45-kb promoter), pINO80:INO80-GFP (2.0-kb promoter), pLUH:LUH-GFP (4.5-kb promoter), pPHYB:PHYB-GFP (2.0-kb promoter), pREF6:REF6-GFP (0.8-kb promoter), and pSPL8:SPL8-GFP (2.5-kb promoter). Our expression data are in line with previous reports on the mRNA expression of *BLR*, *SEU*, *LUG*, *KNAT3*, and *BRM* in the inflorescence (Serikawa et al., 1997; Conner and Liu, 2000; Franks et al., 2002; Roeder et al., 2003; Farrona et al., 2004). GFP signal is indicated in green; chloroplast and other 'background' signal are indicated in red. Adx Pt, adaxial site of petal; FM, flower meristem; IM, inflorescence meristem; Sp, sepal; St, stamen; Stg, stigma; Stl, style. (Scale bars, 50 μm.)

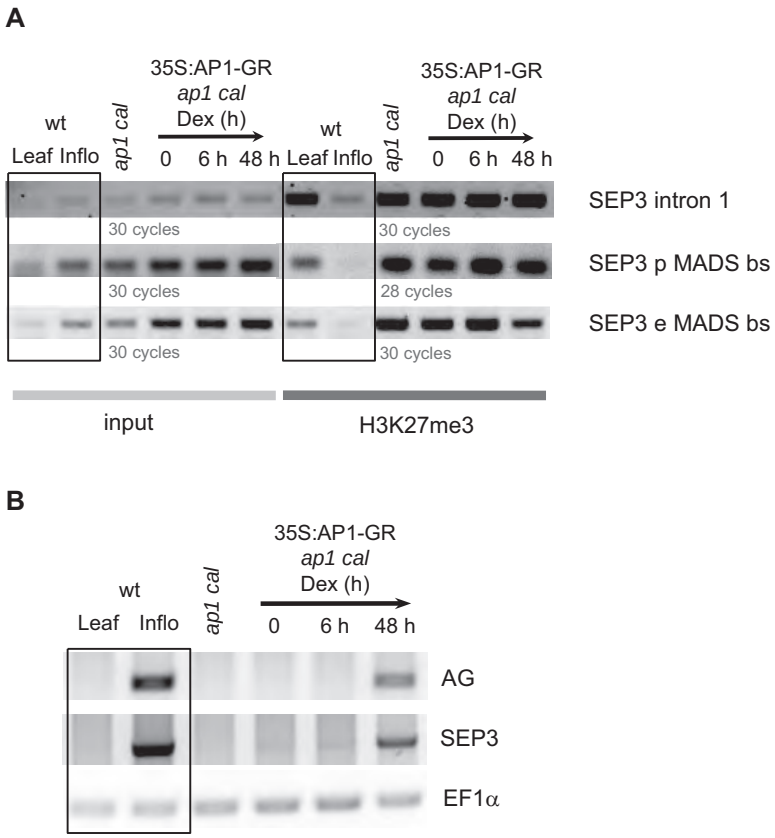


Figure S7. Dynamics of histone modifications in the SEP3 promoter and changes in gene activity. (A) ChIP-PCR analysis of H3K27me3 in the distal *SEP3* enhancer (e), *SEP3* promoter (p), and first *SEP3* intron. In contrast to enhancer and promoter region, no change in H3K27me3 is detectable in the *SEP3* intron 2 d after *AP1* induction. **(B)** RT-PCR analysis of *SEP3* and *AG* expression before and after *AP1* induction at the same time points as for the analysis of histone modifications.

Chapter 3

Proteomics insights into plant signaling and development

Kerstin Kaufmann[1], Cezary Smaczniak[1], Sacco de Vries[2], Gerco C. Angenent[3,4], Rumyana Karlova[1]

[1] Laboratory of Molecular Biology, Wageningen University, 6708PB Wageningen, The Netherlands
[2] Laboratory of Biochemistry, Wageningen University, 6703 HA Wageningen, The Netherlands
[3] Business Unit Bioscience, Plant Research International, 6708PB Wageningen, The Netherlands
[4] Centre for BioSystems Genomics, 6708PB Wageningen, The Netherlands

Proteomics **11**, 744-55 (2011)

Abstract

Mass spectrometry-based proteomics is used to gain insight into the abundance and subcellular localization of cellular signaling components, the composition of molecular complexes and the regulation of signaling pathways. Multicellular organisms have evolved signaling networks and fast responses to stimuli that can be discovered and monitored by the use of advanced proteomics techniques in combination with traditional functional analysis. Plants are multicellular organisms and products of tightly regulated developmental programs that respond to environmental conditions and internal cues. Plant development is orchestrated by inter- and intracellular signaling molecules, receptors and transcriptional regulators, which act in a temporal and spatially coordinated manner. Here we review recent advances in proteomics applications used to understand complex cellular signaling processes in plants.

Introduction

Plants are sessile organisms and therefore need to adjust their metabolism, growth and development to a highly dynamic environment. They also lack a somatic immune system to defend against pathogen attack. Therefore, plants have evolved different pathways to respond rapidly and efficiently towards different external and internal factors such as light, temperature, nutrient deficiency, invasion of pathogens and hormones. These signals are perceived by plasma membrane-localized or cytosolic receptors as is shown in **Figure 1**. Transmembrane receptor kinases (RKs) are classified into several groups based on the structure of the extracellular domains. Initial steps in a standard signal transduction cascade include perception of the signal (ligand) or sets of signals leading to conformational change of the receptor and changes in its interaction partners, which modulate receptor activity. This results in transmission of secondary signals and induction of specific phosphorylation cascades or other posttranslational modifications (e.g. ubiquitylation). Eventually, the signal is transmitted to the nucleus, where transcription factor complexes induce changes in gene expression. The attenuation of the signaling cascade primarily controls receptor degradation, which is in case of membrane-bound receptors usually coupled to internalization (Citri and Yarden, 2006).

Many signaling processes in plants converge at the level of gene regulation. The *Arabidopsis* genome encodes almost 1900 transcriptional regulators according to the current gene ontology (GO) classification (www.arabidopsis.org), which represent about 7% of all protein-coding genes in this plant species. Developmental transitions usually require changes in the relative abundance of key-regulatory transcription factors that act as repressors or activators of specific developmental programs. In

addition, interactions between transcription factors and recruitment of general cofactors can modulate transcriptional regulation.

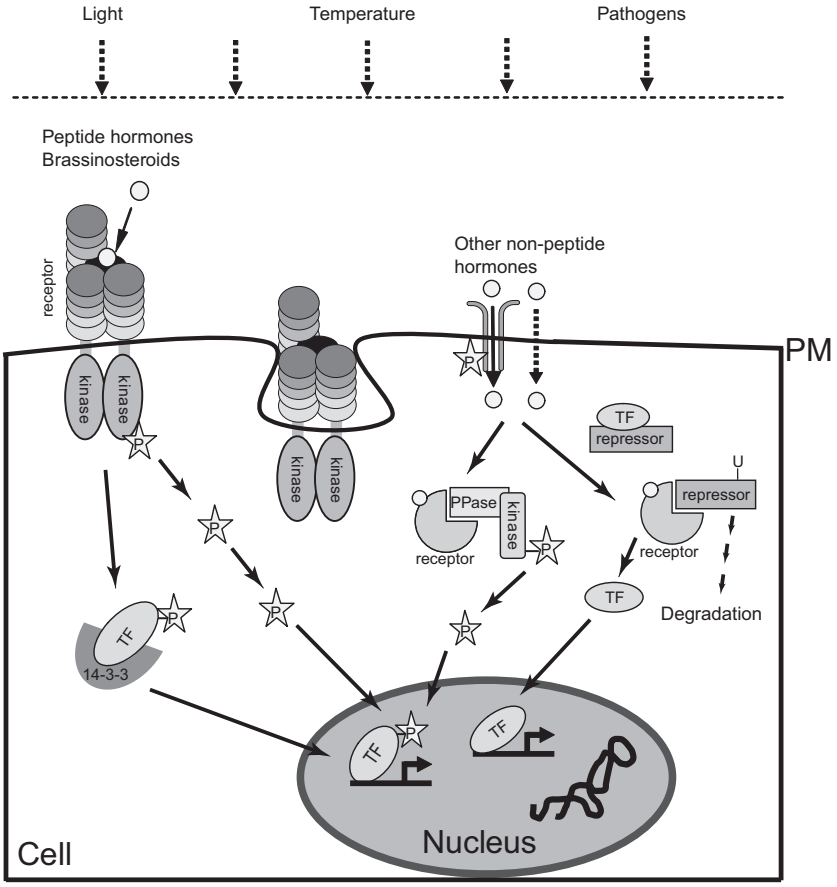


Figure 1. Schematic representation of examples of cellular signaling cascades in plants. The plant cell perceives external or internal signals like light, temperature and invasion of pathogens. Intercellular signaling occurs via peptide and non-peptide hormones. The transmembrane receptor-like kinases can transduce signals trough the plasma membrane (PM) of the cell. Alternatively, signals can be perceived by intracellular receptors. Upon excitation of the receptors, a signaling cascade is activated involving kinases, phosphatases or ubiquitin ligases. This activation is often associated with post-translational modification of the proteins, such as phosphorylation (P) or ubiquitylation. Finally, the signal is transduced to transcription factors (TF) triggering transcriptional responses.

The term 'proteome' was coined by Marc Wilkins (Wilkins et al., 1996) and describes the entire set of proteins expressed by an organism, tissue or cell. 'Proteomics' can be defined as the comprehensive analysis of presence, localization, modifications or interactions of proteins at the subcellular, cellular, organ or organism levels (James, 1997). The rise of proteomics as an area of research is tightly linked to

the revolution of mass-spectrometry (MS)-based technologies in the past 20 years. Proteomics is being used in plant sciences for systematic identification and functional characterization of proteins and protein complexes.

In this review, we focus on recent advances in proteomics applications used to understand cellular signaling and developmental processes in plants. After describing general approaches to unravel proteomes of organs and subcellular compartments, we will focus on proteomics approaches used to study different steps of a typical signal transduction cascade: starting with the proteomic characterization of mobile peptide ligands, followed by the identification of their receptors, receptor protein complex partners and downstream signaling cascades. We introduce quantitative proteomics methods used to identify new components in signaling pathways, in particular, phosphorylation cascades. Since targeted degradation of proteins, e.g. by ubiquitylation, plays a role in several known signaling pathways, approaches to identify new targets of ubiquitylation will be reviewed. Most signaling cascades result in changes in gene regulation, and examples of the LC-MS based characterization of specific transcription factor and chromatin remodeling complexes will be presented.

Proteomics: From whole organism to subcellular protein 'catalogs'

The sequencing of complete genomes of model organisms as well as major crop species is currently leading to a rapid advance in our understanding of biological systems. While genome sequences tell us about the theoretical potential of encoded gene products, they do not give information on their qualitative and quantitative occurrence in the plant. Since most genes (*Arabidopsis thaliana*: 82%; TAIR9) encode for proteins, the compilation of organ- and developmental stage-specific proteomes can help us to unravel the functional properties and activities of the genome.

'Genome-scale' proteomics or 'proteogenomics' efforts aim at comprehensive identification of proteomes to improve genome annotation (Baerenfaller et al., 2008; Castellana et al., 2008; see also ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR9_genome_release/readme_TAIR9.txt). In a proteogenomics effort in the model plant *Arabidopsis thaliana*, Baerenfaller et al. applied linear quadrupole (LTQ) MS using protein extracts from six different organs and developmental stages (Baerenfaller et al., 2008). In total, they identified approximately 13,000 proteins, among which, 57 corresponded to new gene models. This represents nearly 50% of all predicted *Arabidopsis* gene models. The results showed that specific GO categories were overrepresented in the proteome datasets of different plant organs, indicating the presence of organ-specific subproteomes. The effort to comprehensively identify the proteome of *Arabidopsis* was extended by Castellana et al. (Castellana et al., 2008), who identified 1473 new or revised gene models.

The changes in protein abundance during development can be analyzed by quantitative proteomics techniques. Classical approaches for quantitative proteomics make use of two-dimensional gel electrophoresis (2-DE), combining isoelectric focusing in the first dimension and size separation in the second dimension. This approach has been extensively used in the plant field (for review, see Hochholdinger et al., 2006), however the sensitivity and reproducibility of the method are limited. Differential gel electrophoresis (DIGE) addresses these issues by preincubation of the protein samples with sensitive fluorescent dyes, which allows the simultaneous analysis of several samples in a single gel, thus eliminating issues of reproducibility across multiple gels and simplifying the downstream analysis (Viswanathan et al., 2006; for review, see Thelen and Peck, 2007). Also DIGE has found many applications in studies of plant signaling, for instance in studying proteomics changes in response to environmental conditions (e.g. Amey et al., 2005; Ndimba et al., 2005; Bindschedler et al., 2008; Xu and Huang, 2008; Alam et al., 2010) and to identify differences in proteomes of plant organs or developmental stages (e.g. Holmes-Davis et al., 2005; Mooney et al., 2006; Hebeler et al., 2008; Lyngved et al., 2008). For example, a number of proteomic analyses of embryo and seed development in different crop plant species have been performed (Koller et al., 2002; Gallardo et al., 2003; Catusse et al., 2008; Dam et al., 2009; Gomez et al., 2009; Mak et al., 2009; Pawlowski, 2009; Sghaier-Hammami et al., 2009; R. Thompson et al., 2009). Seeds are important for food production and plant breeding, and proteomics tools can for example assist in the identification of protein allergens (Thelen, 2009). 2DE-based proteomic analysis of mature seeds from the crop plant *Beta vulgaris*, which is a major source of sucrose, led to the identification of more than 750 proteins and revealed their expression in root, cotyledons and perisperm within the seed (Catusse et al., 2008). The results provide a proteome-wide snapshot of metabolic activity in the seed, and show that mature seeds are equipped with enzymes to mobilize major reserve compounds during germination. Furthermore, the data show the presence of components of the 26S proteasome/ubiquitin system in seeds. Proteasome activity is known to be important for the control of hormonal activity, particularly via the degradation of DELLA repressor proteins in gibberellic acid (GA) signaling and of AUX/IAA repressors in auxin signaling (for review, see Santner and Estelle, 2010). Degradation of DELLA proteins is important, since these proteins are negative regulators of GA and repress germination.

Further dissection of protein localization, e.g. presence in specific subcellular compartments, can give indications towards protein function as well as towards mechanisms of protein targeting and trafficking. The characterization of proteomes of subcellular organelles is an ongoing process, and studies provided new insights into proteomes of chloroplasts, mitochondria (S. L. Yang et al., 2003; Kleffmann et al.,

2004; Huang et al., 2009), vacuoles (Carter et al., 2004; Jaquinod et al., 2007), nuclei (for review, see Erhardt et al., 2010), peroxisomes (Reumann et al., 2007; Reumann et al., 2009), plasma membranes (Alexandersson et al., 2004; Komatsu, 2008; Sadowski et al., 2008), the cell wall (Bayer et al., 2006; Zhu et al., 2006; Jamet et al., 2008) and the extracellular space (for review, see Agrawal et al., 2010).

Approaches to study organelle proteomes often make use of (2-DE) gel- or MS-based quantification, since these allow the comparison with control samples and assist in the 'high-resolution' dissection of proteomes within organelles. In contrast to gel-based quantification, MS-based methods usually use peak intensities of tryptic peptides in an MS run for quantification of their corresponding proteins. In order to correct for technical variation between different MS runs, which may interfere with correct peak alignment, metabolic and chemical labeling have been introduced into plant proteomics (Ippel et al., 2004; Wienkoop and Weckwerth, 2006; Bindschedler et al., 2008; for review see Thelen and Peck, 2007). MS-based quantification can alternatively be done by MS/MS spectral counting. Approaches such as localization of organelle proteins by isotope tagging (LOPIT) overcome the need for isolation of highly pure (sub)organelle fractions (Dunkley et al., 2004; Dunkley et al., 2006; Lilley and Dupree, 2007). LOPIT uses labeling by isobaric tags for relative and absolute quantification (iTRAQ) and partial separation of organelles by density gradients, and it relies on the quantitative co-separation of proteins with 'marker proteins' for which the localization is known.

Cell-to-cell communication and long-distance transport are essential in signaling cascades controlling cell identity and plant response to environmental and internal cues. Therefore, we will discuss recent proteomics approaches to identify transported and secreted proteins in more detail here.

Long-distance signaling in plants occurs usually via trafficking of signaling molecules, such as hormones, proteins and RNA molecules. Proteins can be transported over longer distances via xylem and phloem sap (for review, see Turgeon and Wolf, 2009). Proteomics has been used for the analysis of phloem sap from several crop plant species (e.g. Haebel and Kehr, 2001; Barnes et al., 2004; Walz et al., 2004; Giavalisco et al., 2006; Aki et al., 2008; Lin et al., 2009). Most studies identified only relatively low numbers of proteins. The most recent study (Lin et al., 2009) identified however more than 1100 potentially transported proteins in phloem sap of pumpkin (*Cucurbita maxima*). The study used a combination of cation- or anion-exchange chromatography and SDS-PAGE or in-solution digest followed by LC-MS/MS. Proteins involved in RNA binding, translation, ubiquitin-mediated proteolysis and trafficking were found to be overrepresented in the phloem sap proteome, and also proteins controlling developmental processes. For instance, TOPLESS and related proteins are found. These proteins are transcriptional corepressors that interact with

transcription factors, and their phloem transport could modulate the activities of these transcription factors. Another example for a transported protein is the flowering inducer FLOWERING LOCUS T (FT) (Corbesier et al., 2007). Transport of FT homologs from pumpkin and rice was confirmed by proteomics (Aki et al., 2008; Lin et al., 2009).

Not only long-distance, but also cell-to-cell communication within tissues is mediated by hormones and extracellular 'peptide hormones' (Agrawal et al., 2010). Specialized membrane-localized cellular import and efflux carrier proteins can control hormonal transport, and plasmodesmata mediate movement of proteins from cell to cell (Lucas and Lee, 2004). The global identification of secreted proteins or peptides is therefore important for our understanding of signaling pathways in plants (Agrawal et al., 2010). The purification and identification of extracellular proteins from native plant tissues remains a challenge, because of contamination with high-abundant proteins that are released from (degenerating) cells during extraction, while extracellular signaling proteins are often small, post-translationally modified (especially glycosylated) and present in low concentrations. Recent studies used DIGE to distinguish cell wall proteins from proteins in fluid (Soares et al., 2007) or alternatively used extraction of extracellular proteins from cell culture, followed by identification with multidimensional protein identification technology (MudPIT), (Cho et al., 2009). MudPIT makes use of multidimensional (usually 2-D) chromatography coupled to MS/MS in order to increase the sensitivity of peptide identification in complex protein samples.

Unraveling signal transduction cascades using proteomics approaches

Signaling processes usually involve direct physical contacts between different components in a pathway, in order to transfer a 'signal' from receptors to transcription factors or other intracellular effector proteins. Combinatorial interactions between signaling proteins can be crucial for determining their cell-type specific functions, subcellular localization and stability. Therefore, the identification of protein complexes and post-translational modifications of signaling proteins is essential to understand signal transduction cascades. The signal is often transmitted from receptors via phosphorylation of intermediate and effector proteins. Protein phosphorylation ensures fast and reversible response to different stimuli. These responses are dependent on protein kinases and phosphatases, which, respectively, catalyze phosphorylation and dephosphorylation of specific substrates. Plants contain approximately twice the number of protein kinases as found in mammals (de la Fuente van Bentem and Hirt, 2007). Proteomics approaches are being used to study changes in phosphorylation in response to variation in light or temperature (Bonardi et al., 2005; El-Khatib et al., 2007), invasion of pathogens (for review, see Quirino et al., 2010), hormones (El-

Khatib et al., 2007; H. Li et al., 2009; Chen et al., 2010), and salt stress (Chitteti and Peng, 2007). An alternative commonly used mechanism for signal transduction is by targeting of repressor proteins for degradation via ubiquitylation (for review, see Vierstra, 2009).

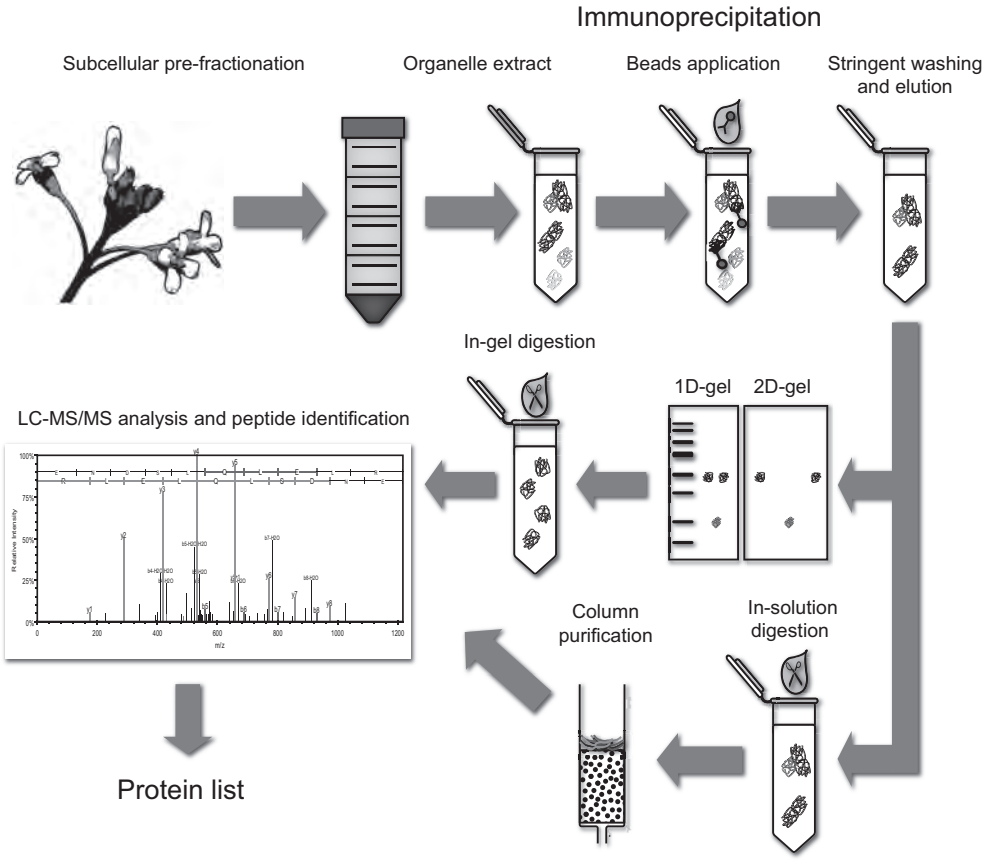


Figure 2. Isolation of protein complexes from plants by affinity purification. The procedure involves an optional prefractionation step for enrichment of specific subcellular compartments, followed by the isolation of the protein complexes using immunoprecipitation of (usually tagged) proteins. The identification of proteins in the complex requires proteolysis using trypsin and/or other enzymes. This can be done in an 'on-bead' or 'in-solution' approach. Alternatively, the eluates can be loaded on a 1-D or 2-D gel and individual bands visible in a Coomassie- or silver-stained gel can be excised. Peptide identification by LC-MS/MS or MALDI-TOF followed by searches of the deduced peptides against a protein database, leading to the generation of a list of proteins present in the IP sample.

Chromatographic separation techniques and density gradient centrifugation are well established in the characterization of protein complexes, often in combination with PAGE. They have been successfully used to characterize major protein complexes

in chloroplasts, mitochondria and cell membranes, as well as more specific components of the general transcriptional machinery in the nucleus (Backstrom et al., 2007) and chloroplasts (Olinares et al., 2010; Schroter et al., 2010). Since signaling proteins usually act in a transient and/or cell-type specific manner, the isolation of their complexes often involves affinity purification (**Figure 2**), which enables a sensitive identification of interacting proteins. Tandem affinity purification (TAP) approaches, Strep tags and biotin tags have been successfully used in plants (Zhong et al., 2003; Rohila et al., 2004; Witte et al., 2004; Rubio et al., 2005; Chang, 2006; Van Leene et al., 2007; Van Leene et al., 2008). Alternatively, protein fusions to green fluorescent protein (GFP) are being used, which allow the direct visualization of the protein expression and subcellular localization *in planta* (Karlova et al., 2006). Combination of affinity purification and separation by size exclusion and/or blue native-PAGE potentially enables the detection of distinct complexes formed by one protein (Remmerie et al., 2009). Recently, the first systematic proteomics efforts to unravel 'interactomes' of specific signaling processes have been accomplished. Proteins of the 14-3-3 family are components of many signaling pathways and bind to a wide variety of client proteins in a phosphorylation-dependent manner. Chang et al. (Chang et al., 2009) performed TAP-tag purification of a generic subunit of 14-3-3 protein complexes that was expressed from a constitutive promoter. Complex partners were identified by a quantitative, MudPIT-based strategy. This approach revealed 101 new potential 14-3-3 clients, indicating that 14-3-3s are some of the most connected nodes in the emerging protein–protein interaction network of plants. Another recent proteomics study characterized the core cell cycle interactome in *Arabidopsis* cell cultures; complex partners of 102 cell-cycle associated proteins, constitutively expressed as fusion to an improved version of the TAP tag (GS-tag), were isolated (Van Leene et al., 2010).

In the following, we will introduce proteomics approaches that were used to identify components of specific plant signaling cascades, and to study their interactions and transmission of signals.

Signaling peptides and their modifications

Extracellular peptides are important components of signaling cascades. An example for a peptide involved in cell-to-cell communication that is localized in the extracellular space is the CLAVATA3 (CLV3) peptide, which restricts the size of the stem cell zone in meristems (Rojo et al., 2002). The *Arabidopsis* genome contains 27 genes encoding CLV3-like peptides (Cock and McCormick, 2001) and for some of them a role in plant development has been demonstrated (for review, see Wang and Fiers, 2010). Other secreted peptides with roles in signaling and development have been identified from different plant species (Amano et al., 2007; Hara et al., 2007; for

review, see Matsubayashi and Sakagami, 2006). Many of these small signaling peptides are processed from longer precursor proteins and secreted as mature peptides into the apoplast (Matsubayashi and Sakagami, 2006). Since secreted peptides are not easily found by 'untargeted' proteomics approaches, Ohyama et al. (Ohyama et al., 2008) established a targeted method to isolate and identify apoplastic peptides and their biochemical modifications from liquid culture of seedlings. It was found that peptides accumulate in the medium in liquid culture and can be enriched by *o*-chlorophenol extraction followed by acetone precipitation prior to LC-MS/MS. By combination of results from this approach with data of transgenic overexpression and endogenous gene expression patterns, the authors identified a novel family of secreted peptides with a role in root growth (Ohyama et al., 2008). They also identified post-translational modifications, in particular, glycosylation and arabinosylation of the CLV3 peptide, which are required for high-efficiency binding to one of its membrane-bound receptors (Ohyama et al., 2009).

Receptor complexes

Membrane-located RKs play important role in plant signaling pathways (reviewed by Torii, 2004). These receptors account for ~2.5–4% of all proteins encoded by plant genomes (Shiu et al., 2004). Nevertheless, the ligands and interacting partners only for a few of them are identified. For instance, CLV3 was found to bind directly to the membrane-bound RK CLAVATA1 (CLV1) and to another membrane-bound complex containing the RKs CORYNE and CLAVATA2 (reviewed in Wang and Fiers, 2010). Other examples are ligand peptide–RK interactions triggering vascular differentiation (Hirakawa et al., 2008), self-incompatibility in Brassicaceae (Takayama et al., 2001) and immune response (Montoya et al., 2002; Chinchilla et al., 2007; Yamaguchi et al., 2010).

Brassinosteroids (BRs) are the only non-peptide hormones identified so far for which a membrane-bound receptor has been identified, named brassinosteroid-insensitive 1 (BRI1) (Nam and Li, 2002; for review, see Karlova and de Vries, 2006). BRs regulate cellular expansion, differentiation and proliferation in plants. BRI1 belongs to a large family of leucine-rich repeat (LRR) RKs that includes over 220 members in *Arabidopsis* and about 400 in rice (Shiu et al., 2004). Transmembrane RKs consist of an extracellular domain, a single transmembrane-spanning domain and a cytosolic kinase domain. Upon ligand binding (BRs), the BRI1 kinase domain is activated and interacts with BRI1 associated kinase 1 (BAK1/SERK3) (Li et al., 2002; Nam and Li, 2002). By co-immunoprecipitation combined with MALDI-TOF, BRI1 and BAK1 were identified as interaction partners of somatic embryogenesis receptor-like kinase 1 (SERK1) (Karlova et al., 2006). BAK1/SERK3 and SERK1 are members of the SERK subfamily of LRR RKs which consist of five members in *Arabidopsis*.

SERK receptors were shown to be involved in different signaling pathways. Aside from its role in BR signaling, BAK1/SERK3 also acts as co-receptor in flagellin/FLS2 signaling (Chinchilla et al., 2007). Other proteins found to interact with the SERK1 receptor were CDC48A and a 14-3-3 protein (Karlova et al., 2006). Another proteomics study confirmed the interaction of 14-3-3 proteins with BRI1 and BAK1 receptors *in vivo* (Chang et al., 2009).

Most non-peptide hormone receptors are not membrane-bound but encode F-box proteins that target repressor proteins for proteolytic degradation (auxin, gibberellin, jasmonic acid) (for a recent review, see Santner and Estelle, 2010). In contrast, abscisic acid (ABA) acts via a cytosolic phosphorylation pathway (Nishimura et al., 2009; Hubbard et al., 2010). ABA mediates resistance to abiotic stress and controls developmental processes in plants and is recognized by several cytosolic receptors, and one of them is pyrabactin resistance 1 (PYR1). A proteomics approach was used for the identification of the protein complex(es) of ABA-insensitive 1 (ABI1), which is a protein phosphatase 2C (PP2C) that plays a critical role as a negative regulator early in ABA signal transduction. The results confirmed the interaction of PYR/RCARs proteins and PP2C *in vivo* (Nishimura et al., 2010). Already 5 min after stimulation with ABA in *Arabidopsis* 9 out of 14 PYR/RCARs proteins were found to interact with ABI1. The results of this study also confirm previously found interactions of ABI1 with ABA-responsive signaling kinases.

Phosphoproteomics of receptors and signaling cascades

Phosphoproteomics approaches allow to study phosphorylation in specific regulatory proteins and to identify new components of signaling cascades. Technical progress has greatly improved the MS-based identification of phosphorylated residues in proteins, in particular, the new generation of highly sensitive and accurate mass spectrometers in combination with phosphopeptide enrichment methods, mainly immobilized metal affinity purification (IMAC) and titanium dioxide (TiO_2). Furthermore, quantitative phosphoproteomics approaches enable the identification of phosphorylation events in response to particular stimuli (Benschop et al., 2007; Tang et al., 2008b; H. Li et al., 2009); and thereby the isolation of new members of signaling pathways (Schulze, 2010).

Several studies have been carried out to identify new components of BR (Deng et al., 2007; Tang et al., 2008a), ethylene (H. Li et al., 2009) and elicitor signaling (Benschop et al., 2007). The experimental set-up in these approaches involves the comparison of phosphoproteomes of 'treated' plants and controls. Common treatments are elicitor/hormone treatment in wild-type plants and specific mutants, changing light conditions (see, e.g. Reiland et al., 2009) and various stress treatments (for review, see Kersten et al., 2009; Schulze, 2010). Different quantitative proteomics

techniques have been applied like label-free quantification, DIGE, iTRAQ or metabolic labeling. Quantitative phosphoproteomics can identify new signaling components, which is especially useful in pathways where genetic approaches have been nearly exhausted.

One of the first large-scale quantitative phosphoproteomics in plants was done on metabolic $^{14}N/^{15}N$ labeled *Arabidopsis* tissue cultured cells (Benschop et al., 2007). Changes in phosphorylation of membrane-associated proteins were analyzed for cells treated with the flagellin peptide flg22 and the fungal elicitor xylanase. The perception of general elicitors by plant cells initiates signaling events, which are associated with phosphorylation and occur within 15 s across the plasma membrane, as shown for the FLS2-BAK1 complex (Schulze et al., 2010). Almost 60% of all phosphopeptides (enriched by TiO_2 affinity chromatography) identified in the xylanase-treated cells were also found in the flg22-treated cells. However, phosphorylated peptides of the FLS2 receptor, which binds directly to the flg22 peptide, were not detected. A similar approach was used by Nuhse et al. (Nuhse et al., 2007), but instead of metabolic labeling the authors used iTRAQ. Several differentially phosphorylated proteins were identified, as well as novel phosphorylation sites in proteins known to be involved in elicitor signaling.

Phosphoproteomic studies resulted in identification of three new components in BR signaling, named BR-signaling kinases (BSK1, BSK2 and BSK3) (Tang et al., 2008b). The authors used seedlings of the BR-deficient *det2-1* mutant, which were treated with BRs. Using 2-D DIGE-based quantification and subsequent phosphopeptide analysis (enriched using IMAC) of excised spots, 19 plasma membrane proteins were detected that responded to the BR treatment (Deng et al., 2007; Tang et al., 2008a).

These studies show that quantitative phosphoproteomics can be used as a powerful tool in addition to genomics and reverse genetic approaches to identify new signaling components involved in plant signal transduction pathways.

Several MS-based studies have been performed to identify phosphorylation sites of individual signaling proteins in plants (for review, see also Kersten et al., 2009). These studies make use of immunoprecipitation to purify a protein of interest, followed by the identification of phosphopeptides by LC-MS/MS. These studies yielded new insights into hormonal signaling, in particular, the BR, ABA and auxin pathways, which will be introduced below.

Identification of phosphorylation sites of the BR receptor BRI1 *in vivo* and *in vitro* in combination with site-directed mutagenesis has provided details on the BR signaling cascade starting with autophosphorylation of BRI1, via interaction and transphosphorylation of its co-receptors to the specific activation of transcription by brassinazole-resistant 1 (BZR1) (for review, see Tang et al.). Although BRI1 and its

co-receptors, SERK1 and BAK1, were considered to be typical serine/threonine (Ser/Thr) protein kinases, recent proteomics studies showed tyrosine phosphorylation sites in BRI1, SERK1 and BAK1 kinase domains, and thus they were proven to be dual-specificity kinases (Karlova et al., 2009; Oh et al., 2009). In a comprehensive analysis of the SERK protein family, the autophosphorylation activity of the kinase domains of the five SERK proteins was compared and the phosphorylated residues were identified by LC-MS/MS. Differences in kinase activity ranged from high activity for SERK1, intermediate for SERK2 and SERK3 and a low activity for SERK4 and SERK5 (Karlova et al., 2009), as indicated by the number of phosphorylated sites.

Phosphoproteomics shed light also on the action of components in the ABA signaling pathway, where several phosphopeptides were identified in the SNF1-related protein kinase 2 (SnRK2) kinase after stimulation with ABA (Umezawa et al., 2009). Binding of ABA to its receptor PYR1 triggers its interaction with the group A PP2C ABI1 and thereby inhibits phosphatase activity, by enabling autophosphorylation and downstream signaling of SnRK2. ABI1 and related PP2Cs thus act as gatekeepers of ABA-mediated signaling by inactivating SnRK2 in the absence of ABA.

Another example of a phosphoproteomics application comes from auxin signaling. Michniewicz et al. identified *in vivo* phosphorylation sites in the *Arabidopsis* PIN1 protein (Michniewicz et al., 2007). PIN proteins facilitate export of auxin from cells (Petrasek et al., 2006). The authors showed that protein phosphatase 2A and the Ser/Thr kinase PINOID act antagonistically in phosphorylation of PIN1. The decision about targeting of PIN1 to apical or basal cell membranes requires reversible phosphorylation of this protein, by facilitating this polarized auxin transport across tissues.

Signaling via the ubiquitine-26S proteasome system

Protein ubiquitylation not only plays a role in hormonal responses (for review, see Santner and Estelle), but also in perception of light, pathogen response, chromatin structure and developmental processes (for review, see Vierstra, 2009). Ubiquitin is a 76-amino-acid polypeptide that can be covalently attached to proteins that are then targeted for degradation. Almost 1700 *Arabidopsis* proteins have been connected to the ubiquitin-26S proteasome system based on genomic studies (Vierstra, 2003). While several core ubiquitin enzyme ligase complexes have been identified, only a few targets of the system had been characterized until recently. To tackle this problem, several proteomic approaches to identify targets of the ubiquitylation pathway have been applied in plants, making use of ubiquitin signature peptides produced by trypsin proteolysis (Peng et al., 2003). Two of these studies used immunopurification with antibodies against endogenous or tap-tagged ubiquitin followed by LC-MS/MS (Igawa

et al., 2009; Saracco et al., 2009). In an alternative approach, the use of a ubiquitin-binding domain that is present in a number of proteins was used as a 'bait' for the purification of ubiquitylated proteins (Manzano et al., 2008). Maor et al. combined this type of affinity purification with MudPIT (Maor et al., 2007). Using this approach, the authors identified almost 300 potentially ubiquitylated proteins in *Arabidopsis* suspension cultures. Together with the ubiquitylated proteins identified in the other methods, several hundreds of new potential targets of the ubiquitylation pathway have thus been identified, many of which with known regulatory functions. It is expected that this is still only a subset of all ubiquitylated proteins, since a number of known ubiquitylated proteins were not identified by these first proteomic studies.

Other post-translational modifications with roles in signaling pathways have been identified, for instance (ubiquitin-related) sumoylation and modification by thioredoxins (see Montrichard et al., 2009 and Miura and Hasegawa, 2010 for review). The comprehensive identification of post-translational modifications by proteomics techniques is expected to greatly contribute to our understanding of complex interplay of signaling mechanisms in environmental response and development.

Protein complexes in transcriptional regulation

Transcriptional control in eukaryotes results from the interplay between specific transcription factors, chromatin remodeling factors and general transcriptional machinery. Reports on the isolation of transcription-associated complexes in plants are still mostly limited to components of the general transcriptional machinery that is present in the nucleus or chloroplasts. More recently, attempts have been made to isolate specific components in transcriptional regulation, which will be introduced here.

Jasmonates are plant hormones with roles in plant defense and development (Browse, 2009). The response to this hormone is mediated by jasmonate ZIM-domain (JAZ) proteins, which interact physically with the transcription factor MYC2 and thereby repress the expression of specific target genes. Isolation of interaction partners of constitutively expressed JAZ protein from cell culture, following a TAP-based approach (Van Leene et al., 2007), identified general corepressors of the TOPLESS (TPL) family as interaction partners of JAZ (Pauwels et al., 2010). This interaction was found to be mediated by an adaptor protein designated as novel interactor of JAZ (NINJA). These interactions thus shed light on how genes are repressed in response to hormones.

Polycomb-group (PcG) proteins are widely conserved proteins in plants and animals that provide a 'cellular memory' of gene repression by trimethylation of histone 3 lysine 27 (H3K27me3) (reviewed by Calonje and Sung, 2006). Notably,

genetic and molecular data suggest that different paralogous PcG proteins have functionally diverged, suggesting the presence of developmental stage-specific protein complexes with partially different functions (Chanvivattana et al., 2004; Hennig and Derkacheva, 2009). Recently, one PcG protein complex has been isolated by immunoaffinity purification followed by LC-MS/MS (De Lucia et al., 2008). The results suggest a dynamic association of a core VRN2 PcG protein complex with additional subunits after prolonged cold treatment of the plants. This dynamic complex formation can be linked to changes in the activity of the complex at 'target' genomic loci (De Lucia et al., 2008). This forms the molecular basis of action of this complex during vernalization response, which is the induction of flowering by low temperatures.

Concluding remarks and future challenges

The revolution of next-generation DNA sequencing technologies facilitated an unprecedented explosion in our knowledge on the genome sequences of model and crop plants. The next challenge is to understand the functional 'output' of the genome, e.g. the abundance and the localization of its gene products, as well as their molecular interactions and the regulation of their activities. Since more than 80% of all genes present in the genome of the model plant *A. thaliana* encode for proteins, proteomics tools are imperative for the completion of this task. A clear technical limitation here is the availability and the applicability of sufficiently sensitive biochemical procedures and workflows. Aside from general plant-specific difficulties like the presence of cell walls or recalcitrant tissue, regulatory proteins are often low abundant, or expressed in a highly tissue-specific fashion. Plants are multicellular organisms where a protein not only can be expressed in different tissues but also can have different functions depending on the presence of associated proteins. This complicates even more the use of proteomics protocols, which have been developed for uniform cells in most cases. While research in animal systems can make use of cell cultures in order to discern cell-type specific proteomes and interactomes, this is not an option in most cases in plants, since plant cells in culture lose their identity rapidly and de-differentiate. An additional obstacle is that regulatory proteins can also be unstable or post-translationally modified, reducing the chance that they will be readily detected using standard high-throughput methods. However, strategies to tackle these problems are emerging.

Because of their highly relevant applications for *in vivo* analysis of protein function, advanced proteomics tools are expected to become more widely used in the analysis of signaling and developmental processes in plants. The emerging generation of genomics, proteomics and metabolomics approaches needs bioinformatics methods for data integration and network reconstruction. For plant research, the integration of

-omics data with quantitative genetics data is expected to contribute to our understanding of complex regulatory networks underlying important phenotypic traits such as yield, pathogen resistance, and nutrient perception and utilization (Baginsky et al., 2010).

Acknowledgements

We apologize to all authors of publications that could not be cited in this review due to space constraints. We thank the NGI Centres, Netherlands Proteomics Centre (NPC) and the Centre for Biosystems Genomics (CBSG) for funding. R.K. and K.K. are funded by the Dutch Organization for Scientific Research (NWO) in the framework of the ERA-PG Program.

Chapter 4

Proteomics-based identification of low-abundance signaling and regulatory protein complexes in native plant tissues

Cezary Smaczniak[1,2,*], Na Li[3,4,*], Sjef Boeren[3], Twan America[1], Walter van Dongen[3], Soenita S. Goerdayal[5,6], Sacco de Vries[3], Gerco C. Angenent[1,7], Kerstin Kaufmann[1,2]

[1] Business Unit Bioscience, Plant Research International, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands

[2] Laboratory of Molecular Biology, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands

[3] Laboratory of Biochemistry, Wageningen University, Dreijenlaan 3, 6703HA Wageningen, The Netherlands

[4] Present address: Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, West Lincui Road, Chaoyang District, Beijing 100101, P.R. China

[5] Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[6] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[7] Centre for BioSystems Genomics, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands

[*] These authors contributed equally to this work.

Nat. Protoc. **7**, 2144-2158 (2012)

Abstract

Owing to the low abundance of signaling proteins and transcription factors, their protein complexes are not easily identified by classical proteomics. The isolation of these protein complexes from endogenous plant tissues (rather than plant cell cultures) is therefore an important technical challenge. Here, we describe a sensitive, quantitative proteomics-based procedure to determine the composition of plant protein complexes. The method makes use of fluorophore-tagged protein immunoprecipitation (IP) and label-free mass spectrometry (MS)-based quantification to correct for nonspecifically precipitated proteins. We provide procedures for the isolation of membrane-bound receptor complexes and transcriptional regulators from nuclei. The protocol consists of an IP step (~6 h) and sample preparation for liquid chromatography-tandem MS (LC-MS/MS; 2 d). We also provide a guide for data analysis. Our single-step affinity purification protocol is a good alternative to two-step tandem affinity purification (TAP), as it is shorter and relatively easy to perform. The data analysis by label-free quantification (LFQ) requires a cheaper and less challenging experimental setup compared with known labeling techniques in plants.

Introduction

Protein interactions are essential for the perception and transmission of signals in cellular signaling cascades and transcriptional regulation. Often, proteins form large complexes that can maintain and execute their specific biological function only in a multimeric form. A common method for the detection of protein interactions in eukaryotic tissues is IP, followed by the detection of interaction partners in a western blot. For this method, two antibodies are required: one against the 'bait' protein that is used for IP, and the other against the potential interaction partner to be used in the western blot. This targeted method has obvious drawbacks in that two antibodies are needed, either against the native proteins or against the protein tags, and in that additional information is required about possible existence of certain interactions. Therefore, this method is not suitable for a large-scale detection of unknown protein interactions. In contrast, IP followed by protein identification by LC-MS/MS-based approaches does not have those drawbacks.

First, large-scale approaches to isolate and characterize multimeric protein complexes from plants using MS were based on TAP, with a gene expressing a specific dual tag and two affinity-based purification steps (Rohila et al., 2004; Rubio et al., 2005; Rohila et al., 2006). Double-affinity purification used in TAP-tagged approaches often requires relatively large amounts of tissue and moderate-to-large amounts of proteins that form relatively stable interactions (**Table 1**). In *Arabidopsis*, cell culture combined with TAP has been successfully used to isolate protein

complexes regulating the cell cycle (Van Leene et al., 2010). However, plant cells in culture attain mixed identities and therefore have limited applicability in studying cell type-specific signaling complexes. In addition, isolation of signaling complexes with two-step affinity purification approaches could lead to loss of interaction partners, as they are often low in abundance and the interactions are usually relatively transient under native conditions (Table 1).

Table 1. Comparison of different protein complex characterization methods. Only the most common methods are mentioned, and some important advantages and disadvantages are summarized here (Berggard et al., 2007; Zhou and Veenstra, 2007; Kaufmann et al., 2011; Pflieger et al., 2011; Dunham et al., 2012).

Method	Advantages	Disadvantages
Single-step immunoprecipitation with antibodies against the bait protein followed by LC-MS/MS	• no prior knowledge of protein complex composition required • no protein fusion tag necessary • physiological levels of the bait protein maintained	• specific antibody required • requires sufficiently high endogenous expression level of bait protein and interaction partners for detection by LC-MS/MS • requires efficient coupling of antibody to magnetic beads, bead size and coupling efficiency may differ from pre-coupled microbeads • antigens may be not accessible to antibody in native complexes
Single-step immunoprecipitation with antibodies against the tagged protein followed by LC-MS/MS	• no prior knowledge of protein complex composition required • use of generic antibodies (usually optimized for this purpose)	• protein tag required • altered expression of the bait protein (if other than endogenous promoter is used) • protein tags can change of or even inhibit the activity, interactions or stability of the protein (functionality of the tagged protein can be tested by mutant complementation) • interaction partners with abundance that is too low may not be detected
Tandem affinity purification followed by LC-MS/MS	• no prior knowledge of protein complex composition required • reduced non-specific binding/background proteins	• transient or weak interactions can be lost due to two rounds of protein complex purification associated with incubation and washing steps • moderate to high levels of bait protein (and interaction partner) expression are required • protein tags may influence protein activity/stability (see above)
Co-immunoprecipitation	• standard approach for *in vivo* confirmation of protein-protein interaction • relatively easy to perform no LC-MS/MS required	• prior knowledge of protein interaction partners is required • specific antibodies against potential complex partners are required (for every single interaction different antibodies are needed), or, if tagged proteins are used, the tags may influence protein function (see above)

Prefractionation techniques and modern LC-MS/MS instruments have enabled a highly efficient detection of low-abundance proteins in complex mixtures. This shifts the bottleneck in the identification of proteins from their mere detection toward discrimination between specific interaction partners and nonspecific background (Blagoev et al., 2003; Rinner et al., 2007; America and Cordewener, 2008; Hubner et

al., 2010). A well-established method for a highly sensitive comparative analysis of protein complexes, which has been widely applied in animal cell lines, is SILAC (stable isotope labeling by amino acids in cell culture) (Blagoev et al., 2003). SILAC makes use of protein labeling with 'light' and 'heavy' forms of selected amino acids (e.g. arginine, lysine). Isotopically labeled proteins present in the target tissue allow for pairwise peptide abundance comparison between differentially labeled ('light' versus 'heavy') samples in the same LC-MS run. SILAC has also been introduced to quantify proteins of plant cell cultures (Clough and Bent, 1998; Gruhler et al., 2005); however, metabolic labeling of intact plants is elaborate and incomplete (Gruhler et al., 2005). More recently, LFQ has been adapted for complex characterization in animal cell lines (Hubner et al., 2010), in which quantitative bacterial artificial chromosome-GFP interactomics (QUBIC) – which combines the expression of GFP transgenes in bacterial artificial chromosomes (TransgeneOmics) (Poser et al., 2008) and LC-MS/MS-based protein quantification approaches – has been used to unravel mammalian protein interactome. Similarly to SILAC, this method is based on LC-MS peptide peak abundance quantification, but it allows for the comparison of samples that were analyzed in separate LC-MS runs, and therefore does not require any kind of labeling. As in other quantification methods, the comparison assumes that the abundance of most proteins in 'treated' and 'nontreated' samples is not changed by experimental conditions, and hence their ratio distributions should be equal, allowing for proper comparison of differentially expressed/enriched proteins. It is quantitatively less accurate compared with SILAC but, with proper use of control samples, is sufficient to determine protein compositions of complexes in pull-down experiments (Hubner et al., 2010).

Here we present an efficient method that combines affinity-based complex isolation of endogenously expressed, fluorophore-tagged proteins from intact plant tissues and the identification of interaction partners by label-free MS-based analysis (**Figure 1**). We provide experimental strategies for transcriptional regulators and membrane-bound receptor proteins, as well as quantitative data analysis approaches for IP experiments in plants.

Our protocol has been successfully applied to characterize MADS-domain transcription factor complexes in *Arabidopsis thaliana* (Smaczniak et al., 2012b) confirming protein interactions in endogenous floral tissues that were proposed in the groundbreaking 'floral quartet' model about 11 years ago (Theissen and Saedler, 2001). In addition, the protocol was effectively used for the *in vivo* identification of dynamin-related proteins associating with the PIN-FORMED (PIN) auxin efflux carriers in *Arabidopsis* (Mravec et al., 2011). The protocol, followed by TiO_2 bead enrichment of phosphopeptides, was also used to study *in planta* phosphorylation of

the *Arabidopsis* basic helix-loop-helix transcription factor SPEECHLESS (Gudesblat et al., 2012).

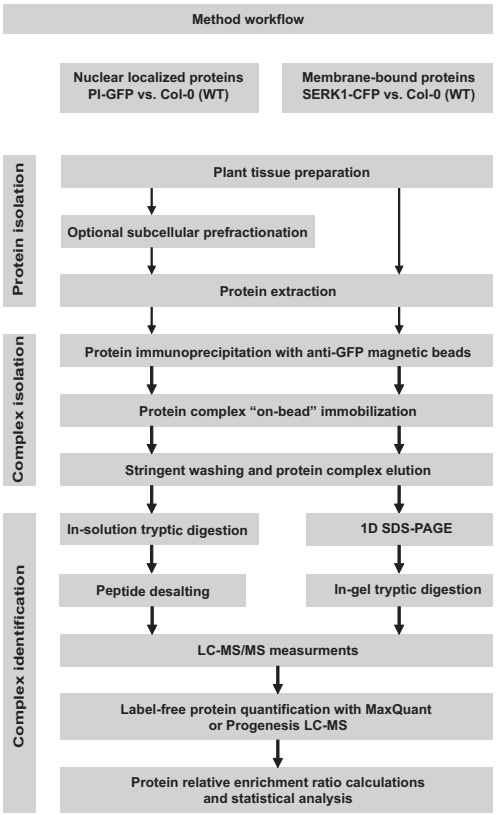


Figure 1. Experimental and simplified data analysis workflow. For detailed description of the data analysis scheme look in the main text for MaxQuant (Cox et al., 2009; Cox et al., 2011) and Supplementary Method 1 for Progenesis LC-MS (http://www.nonlinear.com/products/progenesis/lc-ms/overview/).

Development of the protocol

To develop and optimize this procedure, we chose two proteins for which interaction partners have been identified previously: the membrane-bound leucine-rich repeat receptor-like kinase (LRR-RLK) family protein somatic embryogenesis receptor-like kinase 1 (SERK1) and the floral homeotic MADS-domain transcription factor PISTILLATA (PI). On the basis of IP followed by LC/MALDI-TOF MS analysis, it was previously shown that SERK1 associates with the brassinosteroid receptors brassinosteroid insensitive 1 (BRI1) and BRI1-associated kinase 1, which was verified by blue native gel electrophoresis, genetics and FRET-fluorescence lifetime imaging microscopy (FRET-FLIM) (Karlova et al., 2006). Later studies identified two other LRR kinases (At1G27190 and At3G28450) as SERK1 interaction partners (Karlova, 2008). The PI protein heterodimerizes with the MADS-domain protein APETALA3 and forms higher-order protein complexes with other MADS-domain transcription factors as determined by yeast *n*-hybrid, gel retardation and FRET-FLIM experiments (Immink et al., 2010; Smaczniak et al., 2012b). Thus, SERK1 and PI proteins provide well-characterized testing systems for method development.

The sample preparation procedures were optimized separately for nuclear proteins and for membrane-bound receptors. A protein tagged with GFP or a GFP derivative such as cyan fluorescent protein (CFP) or yellow fluorescent protein (YFP) and expressed from its endogenous promoter (**Figure 2**) was precipitated with GFP

antibodies coupled to paramagnetic microbeads (diameter ~200 nm, Miltenyi Biotec), which provided not only relatively low background (nonspecifically precipitated proteins) but also fast binding kinetics (Hubner et al., 2010). According to the GFP antibody specifications provided by the manufacturer, this antibody also binds GFP variants including enhanced blue fluorescent protein (EBFP), EGFP, ECFP and EYFP. We found that nuclei isolation before IP produced optimal IP results in the case of the PI protein, whereas for SERK1 only a short cell lysis step preceded the IP. The optimal choice of beads used for the IP was crucial. We found that GFP antibodies coupled to magnetic beads markedly improved protein complex partner detection compared with CNBr Sepharose beads. Monoclonal GFP-specific antibodies coupled to magnetic beads had better affinity for GFP (K_d ~5 nM) than polyclonal GFP-specific antibodies coupled to Sepharose beads (K_d ~20 nM). Determination of the maximum binding capacity of the beads, the concentration of the GFP-tagged protein in the plant extracts and the K_d for anti-GFP/YFP interaction allowed us to determine the optimal IP conditions, in which ~50% of the tagged protein is routinely precipitated (**Figure 3**).

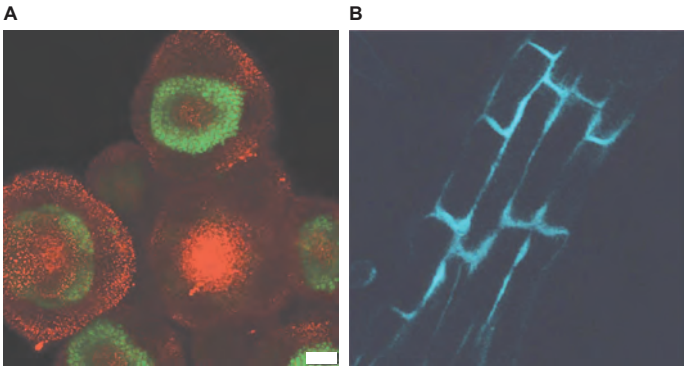


Figure 2. Confocal images of the fluorophore-tagged proteins in native plant tissues. (A) PI-GFP nuclei localized expression profile in early floral meristems of *Arabidopsis thaliana* flowers. GFP florescence is localized in a circular pattern in the floral bud, representing the second and third floral whorls where the *PI* gene is expressed. **(B)** SERK1-CFP membrane localized expression profile in *Arabidopsis thaliana* root tissue. Scale bar, 25 µm.

We used 8 M urea to elute soluble proteins and their interaction partners from the beads for hydrophilic proteins and SDS-PAGE sample loading buffer for membrane-bound hydrophobic proteins. The digested (in-solution or in-gel trypsin digestion) and purified peptide samples were applied onto a reversed-phase HPLC column with an integrated electrospray emitter connected to a hybrid mass spectrometer (LTQ-Orbitrap XL, Thermo Scientific) The Orbitrap type of mass spectrometers generate high-resolution data with a large dynamic range, and thus they are able to detect very-low-abundance proteins in complex samples. This is important for the identification of interaction partners of natively expressed transcription factors or signaling proteins.

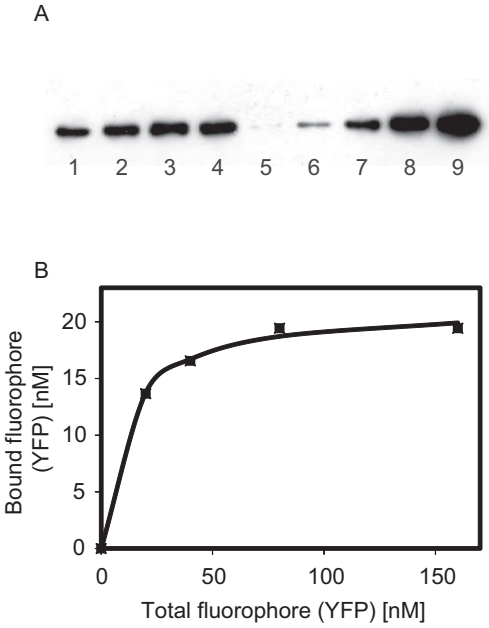


Figure 3. Determination of the binding capacity and affinity of the anti-GFP beads. The affinity and binding capacity of the anti-GFP beads were determined by immunoprecipitation of different concentrations of pure fluorophore (YFP). **(A)** Quantitative immunoblotting. Lanes 1 - 5: YFP precipitated by anti-GFP beads from solutions containing 10, 20, 40, 80 and 0 pmol YFP; Lanes 6 – 9: Known amounts of pure YFP used for calibration (0.2, 0.4, 0.8 and 1.6 ng, respectively). **(B)** Saturation curve plotted from the data visualized in (A) to estimate K_d of anti-GFP – fluorophore (YFP) interaction.

Identification of peptides in the eluates, after digesting the proteins with trypsin, revealed about 300 proteins in the PI-GFP IP samples and 230 proteins in the SERK1-CFP IP samples on the basis of the MaxQuant/Andromeda (Cox et al., 2011) peptide database search and initial filtering, thereby suggesting a substantial amount of background proteins in the eluates. To distinguish specifically immunoprecipitated proteins from the background, we applied a LFQ strategy and compared protein abundances between IP samples and IP controls. There are two LFQ methods that are often used in quantitative proteomics (Zhu et al., 2010). First is the spectral counting method, which compares the number of identified MS/MS (MS2) spectra for peptides of a particular protein and can be used with the data obtained with any type of mass spectrometer. The second method is quantification using the MS (MS1) peak intensity/abundance (extracted ion chromatogram) measurement that allows the separation of the identification process, which uses both MS2 and MS1 data, from the quantification process that takes place only at the MS1 level. Both methods are suitable for analyzing protein abundance changes in large-scale proteomics experiments (Old et al., 2005; Ning et al., 2012). However, with our experimental setup, spectral counting did not perform well, as the amount of MS2 counts in the control samples was very low or even zero for known interaction partners, as well as some other proteins, which hampered protein abundance normalization and proper background-level estimation. In our experiments, we used MS1 peptide peak intensities/abundances for quantification, as the MS1 data also contain complete peak elution profiles required for relative protein quantification. Proper alignment of high-resolution MS1 peaks from several LC-MS runs is essential for accurate quantification. In addition, when MS1 peptide peak alignment is correct, it is not necessary to

identify all MS1 peaks from every LC-MS run (in contrast to the spectral counting method), as a single identification can characterize well-aligned peptide peaks in other runs, ultimately allowing for proper abundance comparison between those peptide peaks. We tested software packages for protein LFQ: MaxQuant (v1.2.2.5, Max Planck Institute of Biochemistry, Martinsried, Germany) (Cox and Mann, 2008) and Progenesis LC-MS (v2.6, Nonlinear Dynamic, Newcastle, UK). MaxQuant is freeware that was developed especially to process high-resolution, Orbitrap-type data. At the current stage of software development, MaxQuant is unable to process data obtained from other types of mass spectrometers. Progenesis LC-MS, in contrast, is a commercial software package that processes data obtained from many different types of mass spectrometers directly or in standard formats (e.g. .mzXML or .mzML).

To correct for the variability in total protein amount in the IP samples and controls, we used a normalization approach assuming that most background proteins were unaffected by our experimental conditions. The normalization procedures are incorporated into both software applications. Low or zero MS1 intensity values in the control data sets can strongly impair the ratio calculation for low-abundance proteins, such as the interaction partners in our data sets. Therefore, there is a need for imputation of a minimal quantity value for peptides that were not quantified (could not be normalized) to calculate approximate protein ratios. We tested several imputation strategies of missing values before (peptide intensity noise imputation) or after (lowest protein abundance imputation) data normalization. We calculated protein ratios by dividing the combined and normalized peptide intensities/abundances of a particular protein in the IP samples with the corresponding values in the controls. Proteins identified with at least two peptides (including one unique peptide) that are markedly enriched in the sample at a permutation-based false discovery rate (FDR) of 0.01 were considered potential interaction partners of the bait protein (**Figure 4**).

Experimental design

Plant material and protein isolation

In our IP approach, we use native plant tissues expressing a fluorophore-tagged version of a protein under the control of its endogenous promoter. Besides fluorophore tags, also other tags such as the shorter human influenza hemagglutinin (HA) tag may be used. We recommend testing the functionality of fusion proteins using mutant complementation before the IP experiments. Another alternative strategy is to covalently couple custom-made antibodies against endogenous proteins to magnetic beads, which can then be used for IP.

The optimal choice of the plant tissue for IP experiments is crucial for the characterization of the complex partners. Until now, we successfully performed

complex isolation from *Arabidopsis* flower buds, overproliferating meristematic tissues and seedlings, but our protocol can also be applied to any kind of plant tissue in which the expression of the bait protein is well characterized. Before conducting an IP experiment, the expression and the intracellular localization of the bait protein in the target tissue should be confirmed. The choice of tissue subcellular fractionation techniques or total protein extraction strongly depends on the characteristics of the bait protein expression and its stability during the isolation procedures. Transcription factor protein complexes can be isolated using total protein extract or, alternatively, after nuclei isolation. We suggest that both strategies be tested when setting up a new complex isolation experiment.

The characteristics of each tissue (e.g. cell size and 'density', nucleus/cytoplasm ratio) force an optimization of the minimal amount of plant material that should be used for a successful IP. The amounts can vary from 0.75 g of fresh plant material for meristematic tissues to 5 g for seedlings or roots per single IP experiment. When the expression of the bait protein is low, larger amounts of plant tissue are needed, which might be hard to obtain in a relatively short time, or enrichment techniques could be applied to select for the cells that express the bait protein (Deal and Henikoff, 2011). In contrast, the use of total protein extracts can lead to a higher amount of background proteins that can affect the detection of low-abundance proteins present in a complex.

'Background' proteins and potential false positives

In our IP protocols, we do not use any protein-protein or protein-DNA cross-linking agents (e.g. formaldehyde), as it could result in the identification of a higher number of false positives. Some protein complexes, owing to their biological nature (e.g. transcription factors), are stabilized in the presence of DNA (Melzer et al., 2009; Smaczniak et al., 2012b). However, the incomplete fragmentation by sonication used in our protocol may lead to the isolation of protein complexes together with other proteins that are attached to DNA in the nuclear chromatin. These 'interactions' could be a source of false positives. To test whether proteins that are enriched in the IP are false positives or whether they depend on the presence of DNA, we recommend using a control in which an additional step of DNA digestion by a nuclease (for example, Benzonase) is introduced. Benzonase has also been used in protocols on animal cell lines (Hubner et al., 2010).

Control experiments

The idea of using quantitative proteomics for characterization of protein complexes requires selection of suitable IP control samples for proper comparison. We used wild-type Arabidopsis plants as controls to our transgenic fluorophore-tagged plant lines. By using this type of control, we were able to correct for the nonspecific

binding of the GFP antibodies to other proteins. The choice of wild-type *Arabidopsis* plants as controls to fluorophore-tagged plant lines does not correct for potential interactions between a tag (fluorophore) and other proteins within the immunoprecipitated sample. Hence, if a plant line expressing GFP under the control of the same promoter as the bait protein is available, it should be preferably used as a control. In addition, a small amount of immunoprecipitated proteins in the control samples can theoretically cause problems with the imputation of missing protein abundance values in the data analysis. Alternatively, native antibodies against the protein of interest (the bait) can be used (e.g. coupled to magnetic beads) for IP in the wild-type plants, and a corresponding mutant plant line can be used as controls.

Replicates

To reliably estimate protein abundance ratios, we advise performing IP experiments with a minimum of two or three biological replicates each analyzed in two separate technical replicates. Generally, for the LFQ of protein levels, increasing the number of biological replicates will lead to better estimation of protein ratios. During the quantification procedures, replicates are used to calculate the statistical differences between samples and controls, as well as the normalization factors that allow multiple sample comparison.

Sample processing and LC-MS/MS

After the IP, protein eluates are digested (either in-solution or in-gel) with proteomics-grade trypsin. Alternatively, other proteases can be used to digest proteins that rarely contain arginine and lysine residues (e.g. endoproteinase Glu-C). Thereafter, the digest could be desalted on a solid-phase extraction column with the C18-bonded silica stationary phase or the hydrophilic-lipophilic balanced (HLB) copolymer filling. The disadvantage of the desalting step is the loss of hydrophilic and very small peptides that could potentially improve the total protein coverage. After MS measurements, the protein LFQ is important for correct characterization of the complex partners. Alternatively to MS1 peak intensity comparison, both MaxQuant and Progenesis LC-MS give the number of MS2 counts, which can be used for the spectral counting method of relative protein abundance calculations.

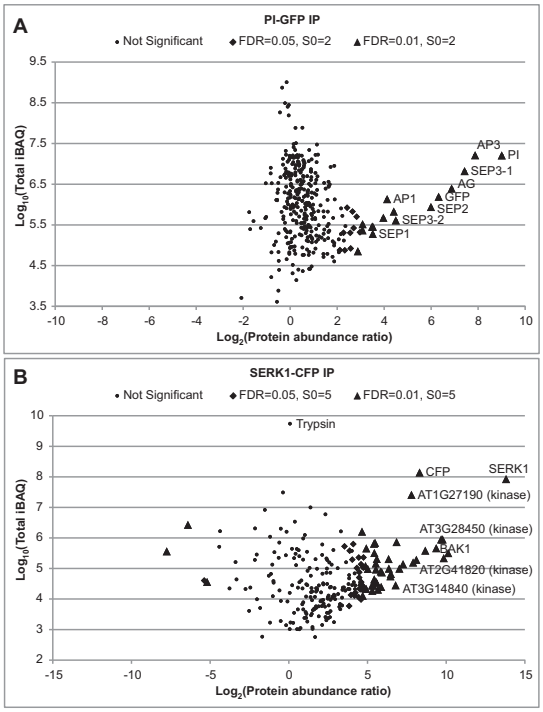
Data analysis

Introducing a statistical approach, where missing LFQ (label-free quantification) or iBAQ (intensity-based absolute quantification) (Schwanhausser et al., 2011) values are imputed with an artificial abundance background value (Hubner et al., 2010), might cause a bias towards low abundance levels for immunoprecipitated proteins (the calculated relative ratio will be lower than the true ratio). In contrast,

imputation of missing values in the controls allows for comparison of protein abundances present only in the IP samples.

Materials, procedure, troubleshooting and timing

Materials, procedure, troubleshooting and timing sections are available upon request or at http://www.nature.com/nprot/index.html website.



C

Protein Symbol	Peptide Number		Sequence Coverage [%]		p-value
	All	Unique	All	Unique	
PI-GFP IP					
PI	13	13	61.1	61.1	7.47E-05
AP3	10	10	50	50	5.34E-05
SEP3-1	9	1	41.6	5.6	2.28E-05
AG	7	7	41.7	41.7	1.77E-05
GFP	6	6	33.8	33.8	5.36E-07
SEP2	4	3	18.8	13.6	2.25E-07
SEP3-2	9	1	41.8	6	1.82E-06
AP1	5	5	24.6	24.6	1.33E-02
SEP1	3	2	12.6	7.6	2.73E-03

D

Protein Symbol	Peptide Number		Sequence Coverage [%]		p-value
	All	Unique	All	Unique	
SERK1-CFP IP					
SERK1	31	21	55.7	38.2	5.04E-05
AT2G41820 (LRR-RLK)	17	17	24.2	24.2	2.93E-04
AT3G28450 (LRR-RLK)	17	17	30.6	30.6	6.15E-05
BAK1	11	1	18.3	1.8	1.72E-03
CFP	18	18	64.7	64.7	6.30E-05
AT1G27190 (LRR-RLK)	36	30	52.6	47.4	2.26E-05
AT3G14840 (LRR-RLK)	9	9	9.9	9.9	2.28E-04

Figure 4. Protein interaction profiling using MaxQuant. (A) and **(B)** graphical representation of the (normalized) protein abundance ratios between the samples: PI-GFP (A) or SERK1-CFP (B) and the wild type control, plotted against the iBAQ intensities for a particular protein. **(C)** and **(D)** lists of identified known interaction partners with number of identified peptides (all and unique with their sequence coverage. The statistical significance was calculated essentially as suggested previously (Hubner et al., 2010) by permutation based FDR estimation with the FDR set to 0.01 or 0.05 and the s0 parameter set to 2 (for PI-GFP IP) or 5 (for SERK1-CFP IP). The triangles indicate significant protein abundance differences between samples and controls within FDR=0.01, the diamonds mark significant differences within FDR=0.05 while the dots correspond to not significantly enriched proteins.

Anticipated results

By following this procedure, we were able to reduce IP complexity with the PI-GFP protein as a bait from ~300 identified proteins to ~15 proteins using relative quantification by MaxQuant (**Figure 4**). We identified the bait protein PI and its

heterodimerization partner APETALA3 with the highest ratios among all enriched proteins, and with a sequence coverage of 50–60%. Protein interaction profiling of PI-GFP IPs using Progenesis LC-MS gave comparable results to MaxQuant (**Supplementary Methods** and **Supplementary Figure 1**). We also identified other MADS domain proteins that, together with the PI, are responsible for floral organ specification as enriched. We found additional proteins to be enriched in the IP samples, which represent novel candidates for protein complex partners of the PI protein (**Supplementary Data 1**).

In the SERK1 membrane receptor immunoprecipitate, we identified the bait protein SERK1 with the highest intensity and sequence coverage among the enriched proteins. We identified the previously identified interactors LRR-RLK At1G27190, BAK1 and At3G28450 among the five most enriched proteins, with sequence coverage ranging from 20 to 52% (**Figure 4**). We observed a novel candidate interactor, LRR-RLK At2G41820 interacting with SERK1. We confirmed these interactions with SERK1 by FRET-FLIM/bimolecular fluorescence complementation in *Arabidopsis* protoplasts (**Supplementary Methods** and **Supplementary Figure 2**). We found that the previously identified SERK1 interactor BRI1 was not highly enriched in the SERK1 complex isolation under our experimental conditions, possibly as a result of the more transient nature of the interaction with SERK1. Proteins enriched after SERK1 IP also include other membrane proteins (**Supplementary Data 2**), with slightly lower enrichment ratios and peak intensities.

Author Contributions

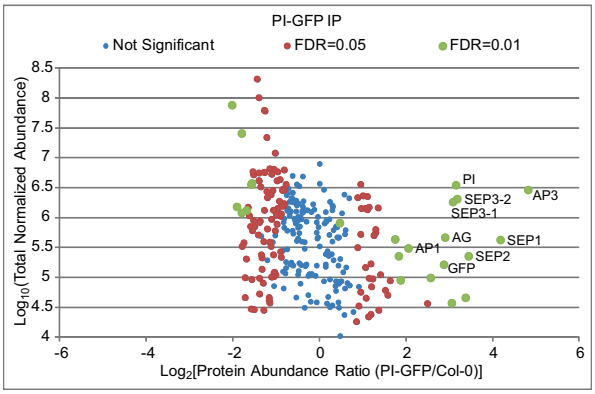
C.S., N.L. and K.K. designed the research; C.S., N.L. and W.v.D. performed experiments; S.B. recorded LC-MS/MS measurements; C.S. and S.B. analyzed the data; S.B., T.A., S.d.V. and G.C.A. commented on the manuscript; S.S.G. provided trypsin digestion and peptide purification protocols for soluble proteins; K.K. supervised the project; C.S., N.L. and K.K. wrote the manuscript.

Acknowledgements

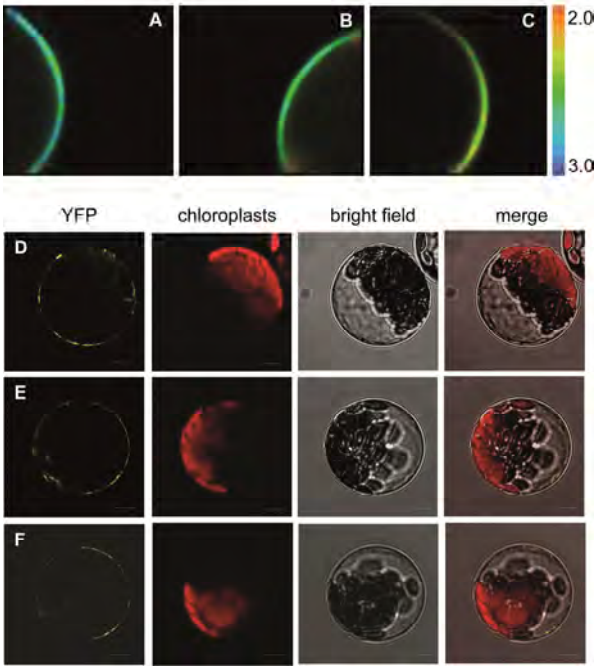
C.S. and G.C.A. were supported by a Proteomics Hotel grant from the Netherlands Proteomics Centre (NPCII-T2.3) and Centre for BioSystems Genomics (CBSG2012). All proteomic measurements were recorded at Biqualys Wageningen (http://www.biqualys.nl). N.L. was supported by the Netherlands Organization for Science, ALW grant no. 61.62.5000.86. N.L., W.v.D. and S.d.V. thank B. Möller for sharing details of IP protocols and helpful discussions. K.K. thanks M.C. O'Flaherty and R. Karlova for advice during initial stages of the project, as well as J.H. Cordewener and J.J.M. Vervoort for useful discussions. We thank J. Muiño for support in bioinformatics and statistical analysis.

Supplementary information

Supplementary figures



Supplementary Figure 1. Protein interaction profiling using Progenesis LC-MS on the example of PI-GFP IPs. Graphical representation of the normalized protein abundance ratios between the PI-GFP IP samples and controls plotted against total protein normalized abundance (summed peptide MS1 peak abundances – areas under peak). Imputation of the missing values with the lowest normalized protein abundance value.



Supplementary Figure 2. Interactions of SERK1 with At1g27190, At2g41820 and At3g28450 confirmed by FRET-FLIM and BiFC in *Arabidopsis* protoplasts. (A-C) Fluorescence lifetime images of *A. thaliana* leaf protoplasts transiently expressing SERK1-sCFP **(A)**, SERK1-sCFP + At2g41820-sYFP **(B)** and SERK1-sCFP + At1g27190-sYFP **(C)** for 16 hrs. sCFP lifetime distributions are presented as pseudocolor images; the color bar shows the lifetime distribution, ranging from τ = 2.0 ns (red) – 3.0 ns (blue). Average lifetimes, determined on at least 45 protoplasts in three independent experiments, are 2.62 ± 0.06 ns for SERK1-sCFP, 2.37 ± 0.06 ns for SERK1-sCFP/At1g27190-sYFP and 2.46 ± 0.09ns for SERK1-sCFP/At2g41820-sYFP. **(D-F)** Bimolecular fluorescence complementation of SERK1-YFP^C with At1g27190-YFP^N **(D)**, At2g41820-YFP^N **(E)** and At3g28450-YFP^N **(F)** expressed in *Arabidopsis* leaf protoplasts for 16 hrs.

Supplementary Methods and Datasets

Supplementary methods and datasets are available upon request or at http://www.nature.com/nprot/index.html website.

Chapter 5

DNA-binding specificity of floral homeotic protein complexes: quantitative binding differences predict organ-specific target genes

Cezary Smaczniak[1,2,*], Jose M. Muiño[1,3,4,*], Gerco C. Angenent[1,5], Kerstin Kaufmann[2]

[1] Business Unit Bioscience, Plant Research International, 6708PB Wageningen, The Netherlands
[2] Laboratory of Molecular Biology, Wageningen University, 6708PB Wageningen, The Netherlands
[3] Laboratory of Bioinformatics, Wageningen University, 6708PB Wageningen, The Netherlands
[4] Present Address: Computational Molecular Biology, Max Planck Institute for Molecular Genetics, D-14195 Berlin, Germany
[5] Centre for BioSystems Genomics, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands
[*] These authors contributed equally to this work.

Manuscript in preparation.

Abstract

MADS-domain transcription factors (TFs) have important functions throughout plant development. How these TFs, which possess a conserved DNA-binding domain and bind highly similar consensus DNA sequences, acquire different functional specificities remains enigmatic. The ability of MADS-domain TFs to multimerize suggests that MADS-domain proteins may form complexes with different DNA-binding specificities. To test this idea, we used an *in vitro* high-throughput SELEX-seq approach to discriminate DNA-binding specificities of several MADS-domain protein homo- and heterodimers. By mapping the *in vitro* bound DNA sequences to the *Arabidopsis* genome and compare it with the available *in vivo* DNA-binding data we were able to characterize complex-specific binding events at genome wide scale. Using this strategy, we could show that not only the specificity but also the affinity to a particular DNA binding site can successfully discriminate different MADS-domain complexes. Taken together, these data advance our knowledge on the molecular mechanisms of MADS-domain TF binding and target gene selection.

Introduction

The exact molecular mechanisms of DNA recognition are still unknown for many TFs. Particularly intriguing is the question, how closely related TF protein family members ultimately control distinct biological processes. DNA recognition by proteins resides from the primary DNA sequence and its structural properties (Rohs et al., 2009). In addition, the chromatin organization, such as nucleosome occupancy, strongly affects the recognition and binding of TFs to DNA in eukaryotes (Kaplan et al., 2009; Zhang et al., 2012). Another aspect that contributes to biological specificity comes from the ability of TFs to form higher-order protein complexes and their combinatorial interactions. The interplay of these regulatory mechanisms makes gene regulation a tightly controlled process.

MADS-domain TFs are present in all eukaryotic organisms. Especially in plants, they form a large family of more than 100 members (Parenicová et al., 2003) and have important roles in the regulation of many developmental processes (reviewed by Smaczniak et al., 2012a). Remarkably some MADS-box genes acquired more than one function in different organs or developmental stages. To explain the variety of regulatory functions of MADS-box genes, we need to understand the molecular mechanisms of their target gene recognition and regulation. MADS-domain proteins bind regulatory regions of their target genes through their highly conserved, 56 amino acid N-terminal DNA binding domain called the MADS domain (Schwarz-Sommer et al., 1990). *In vitro* studies revealed that MADS-domain proteins bind a 10 bp DNA consensus sequence, CC[A/T]$_6$GG, called the CArG-box (Pollock and Treisman,

1990; Schwarz-Sommer et al., 1992; Huang et al., 1993; Riechmann et al., 1996a; Riechmann et al., 1996b). Thousands of CArG-box-like sequences are present in the genome of *Arabidopsis*, many of which seem not to be bound by MADS-domain factors (de Folter and Angenent, 2006). Besides that, the majority of DNA sites (82%) that are bound by MADS-domain proteins *in vivo* do not contain perfect CArG-boxes (Kaufmann et al., 2009; Zheng et al., 2009; Kaufmann et al., 2010c; Deng et al., 2011).

In the functionally most well characterized class of plant MADS-domain proteins, the MIKC-type proteins, the MADS domain is followed by the I (intervening) and K (keratin-like) domains and a highly-variable C-terminus (reviewed by Kaufmann et al., 2005b). DNA-binding is mediated by the N-terminal and central regions of the MADS-domain. MADS-domain proteins bind to individual CArG-boxes as dimers, and according to structural analysis of MADS domains from animal and yeast representatives (Pellegrini et al., 1995; Tan and Richmond, 1998; Huang et al., 2000), dimerization is mostly mediated by β-sheets in the C-terminal part of the MADS domain. Additionally, residues in the I-domain play a role in dimerization, which presumably forms an α-helical structure analogous to animal MADS-domain proteins (Riechmann et al., 1996a). The K-domain may also contribute to the stabilization of dimeric interactions, but it is also required for the formation of quaternary protein complexes (Egea-Cortines et al., 1999; Yang and Jack, 2004). *In vitro* studies suggest that quaternary MADS-protein complexes can bind to two CArG-boxes simultaneously, thereby creating DNA loops (Egea-Cortines et al., 1999; Melzer and Theissen, 2009; Melzer et al., 2009; Smaczniak et al., 2012b).

In *Arabidopsis*, MADS-box genes were initially characterized in flower development, where four different classes of MADS-box genes (A-C and E class) act together to specify different floral organs (Haughn and Somerville, 1988; Coen and Meyerowitz, 1991; Pelaz et al., 2000). According to the 'floral quartet' model, each type of floral organ is specified by a distinct combination of four functional classes of MADS-domain proteins that form quaternary protein complexes and bind two CArG-boxes in the regulatory regions of target genes (Theissen and Saedler, 2001). However, how the target gene specificity of these protein complexes is achieved is still not understood. Some part of the functional specificity may come from the ability of MADS-domain proteins to form homo- and heteromeric protein complexes, which has been suggested by domain swaps experiments (Krizek and Meyerowitz, 1996; Riechmann et al., 1996a). Recently, the interactions of MADS-domain proteins suggested in the 'floral quartet' model were characterized *in vivo* in *Arabidopsis* (Smaczniak et al., 2012b) confirming the combinatorial mode of action for MADS TFs.

Here we show that several MADS-domain protein dimers bind diverse CArG-box-like sequences with different affinities, which unravels their DNA-binding specificities. In our approach we make use of systematic evolution of ligands by exponential enrichment (SELEX) followed by high-throughput sequencing (seq), which in addition to DNA-binding specificities, enable us to measure the DNA-binding affinities for each combination of MADS TF dimer and DNA sequence. Mapping of the *in vitro* bound sequences to the genome of *Arabidopsis* and comparing them with the available *in vivo* ChIP-seq binding profiles, allow us to discriminate target genes for each particular MADS-domain dimer.

Results

SELEX-seq for MADS-domain TF complexes

In our SELEX approach we made use of *in vitro* translated MADS-domain proteins that were incubated with the dsDNA library, which contained a region of randomized nucleotides. Each dsDNA library was labeled with a specific bar code for multiplexing, and PCR primer flanking sites for amplification and DNA sequencing procedures (Jolma et al., 2010). We used two sets of libraries that contained different lengths of the randomized nucleotide region (20N and 40N, respectively). Each round of SELEX was performed with immobilized, target protein-specific antibodies and is summarized in **Figure 1A**. During the procedure, DNA sequences bound by MADS-domain TF complexes were isolated by immunoprecipitation and, after short PCR amplification, used in the subsequent SELEX rounds (**Figure 1A**). For each MADS-domain TF dimer combination at least three rounds of SELEX were performed and the evolved pools of sequences were characterized by high-throughput sequencing after each round (R1-R3). To demonstrate the enrichment of TF-DNA complexes, two to three additional rounds of SELEX were performed and the pool of sequences from each round was labeled with biotin and studied by EMSA (**Figure 1B** and **C**). MADS-domain TF complexes bind the evolved pool of DNA as dimers and as higher-order complexes (composed of two dimers). No binding was visible in the control experiment that contained DNA and the *in vitro* translation mix only. As an initial check for the enrichment, the 6th round of SELEX for SEPALLATA3 (SEP3) homodimer was sequenced by standard Sanger sequencing. More than 80% of the sequenced fragments (21 out of 25 sequences) contained a CArG-box-like motif (**Figure S1A**).

High-throughput sequencing of the first three rounds of SELEX (R1-R3) for the various dimer combinations showed enrichment of the putative MADS-domain TF consensus binding sites (the 'perfect' CArG-boxes) in the evolved pools of sequences (**Figure 1D**). The enrichment of the CArG-boxes in SELEX for APETALA1 (AP1) homodimer was low ($3{\cdot}10^{-3}$%), although still higher comparing to

enrichment of any random 10 bp sequence ($5 \cdot 10^{-5}$%) (**Figure S1B**). There was no enrichment of the CArG-box motif in the control experiments (**Figure 1D**).

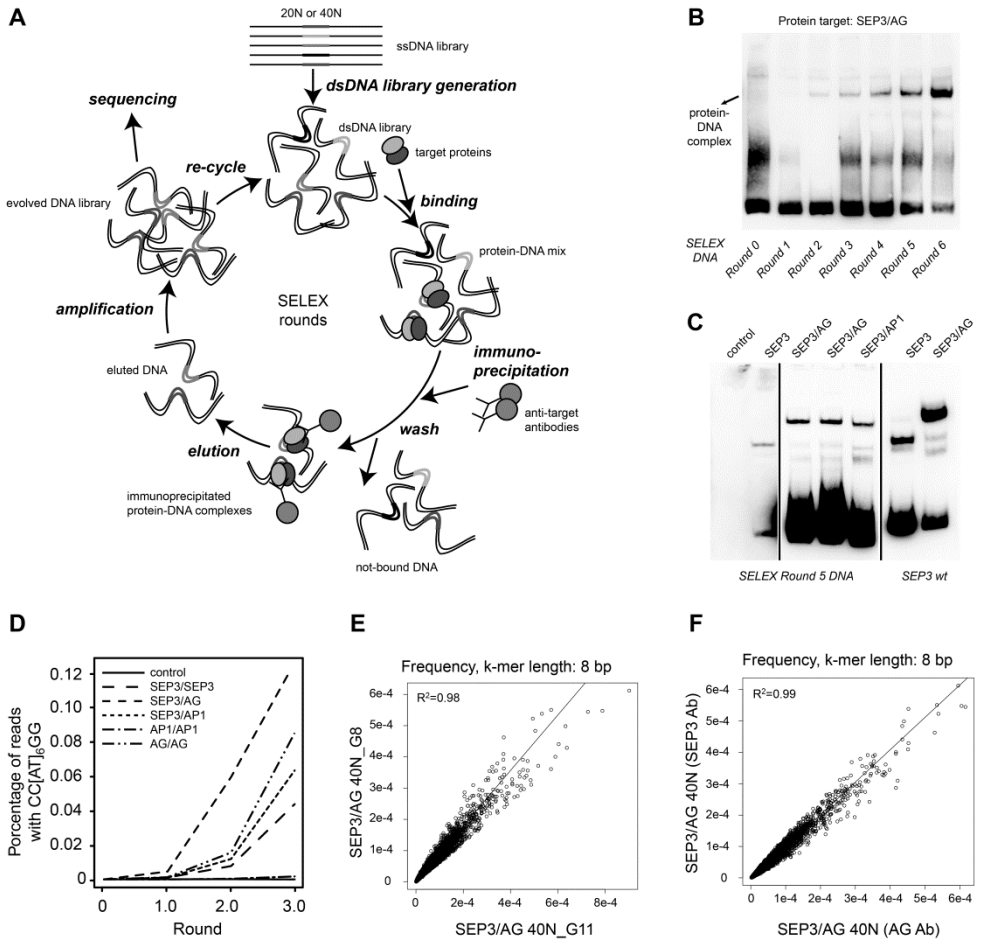


Figure 1. SELEX-seq experiment analysis. (A) Overview of experimental setup for the SELEX experiments performed in this study. **(B)** EMSA analysis of the DNA library obtained at different rounds of SELEX for the SEP3/AG complex. **(C)** EMSA analysis of the 5th round of SELEX libraries for different MADS-domain TF complexes. **(D)** Enrichment of the known, perfect CArG-box consensus sequences ($CC[A/T]_6GG$) in the sequenced SELEX rounds. Control: SELEX-seq performed with protein synthesis mix alone (without MADS-domain proteins). **(E and F)** SELEX-seq quality control and reproducibility of the libraries. **(E)** Dot plot graph represents comparison between 8-mer frequencies in each of the 40N libraries (G8 and G11 respectively). **(F)** Dot plot graph represents comparison between 8-mer frequencies in each of the 40N libraries obtained after immunoprecipitation with different antibodies (SEP3 and AG antibodies respectively).

To estimate the affinities of MADS-domain TF dimers to the DNA fragments we sequenced the initial libraries used in our experiments (R0) and calculate the relative enrichments between R1-R3 and R0 round of SELEX using the method

proposed by Slattery et al. (Slattery et al., 2011; see also Materials and Methods). To check the reproducibility of the SELEX we performed two additional experiments in which we compared the frequencies of 8 bp DNA fragments (8-mers) in SELEX-seq derived 40N DNA sequences. In the first SELEX-seq experiment, we used the same dimer [SEP3/AGAMOUS (AG)] that was incubated with two different 40N libraries. In the second experiment, we used the same dimer (SEP3/AG) that was immunoprecipitated with different antibodies (SEP3 and AG antibodies respectively). Both experiments showed very strong correlation (R^2 = 0.98 in the first and R^2 = 0.99 in the second set-up) showing great reproducibility of our experimental procedure (**Figure 1E** and **F**).

DNA binding specificities for MADS-domain TF complexes

To determine DNA binding specificities for MADS-domain TF complexes we performed SELEX on several combinations of *in vitro* synthesized MADS-domain proteins and different (20N and 40N) bar-coded DNA libraries. We studied the DNA specificity of the following MADS-domain protein complexes: SEP3 homodimer (SEP3/SEP3), AP1 homodimer (AP1/AP1), AG homodimer (AG/AG), SEP3/AP1 heterodimer (sepal and petal specific dimer) and SEP3/AG heterodimer (stamen and carpel specific dimer). We determined the optimal length of the randomized fragment that sufficiently predicts the specificity of the bound MADS-domain complexes by calculating the information gain with Kullback-Leibler divergence (Slattery et al., 2011). The optimal length of the k-mer in the evolved libraries varied between 10 and 12 bp (**Figure S2**).

We used k-mer length of 12 bp to identify the most enriched sequences in each library. The SELEX-seq derived estimated relative affinities to 12 bp DNA sequence fragments were plotted in a heatmap (**Figure 2**), which could differentiate MADS-domain dimers into clusters. We observed that different MADS-domain protein complexes show different affinities for the enriched 12-mers. We organized them into four groups based on a similar affinity pattern: group A (controls), group B (SEP3/SEP3, SEP3/AP1 and AP1/AP1 dimers), group C (AG/AG dimer) and group D (SEP3/AG dimer). Additionally, we selected six groups of 12-mers (Group 1-6) that were specific for the selected groups of dimers (**Figure 2**). To identify consensus binding sites of MADS-domain TF complexes in each 12-mer group we extracted the full length sequences (40N) that contained a particular 12-mer and performed DNA motif discovery using GADEM algorithm (Li, 2009).

Comparison between different 12-mer group motifs revealed differences in nucleotide composition between different MADS-domain binding sites. Most of the resolved motifs showed relative similarity to the perfect CArG-box sequence. In these motifs two cytosines [C at position 5 (C5) and C6] and two guanines (G13 and G14)

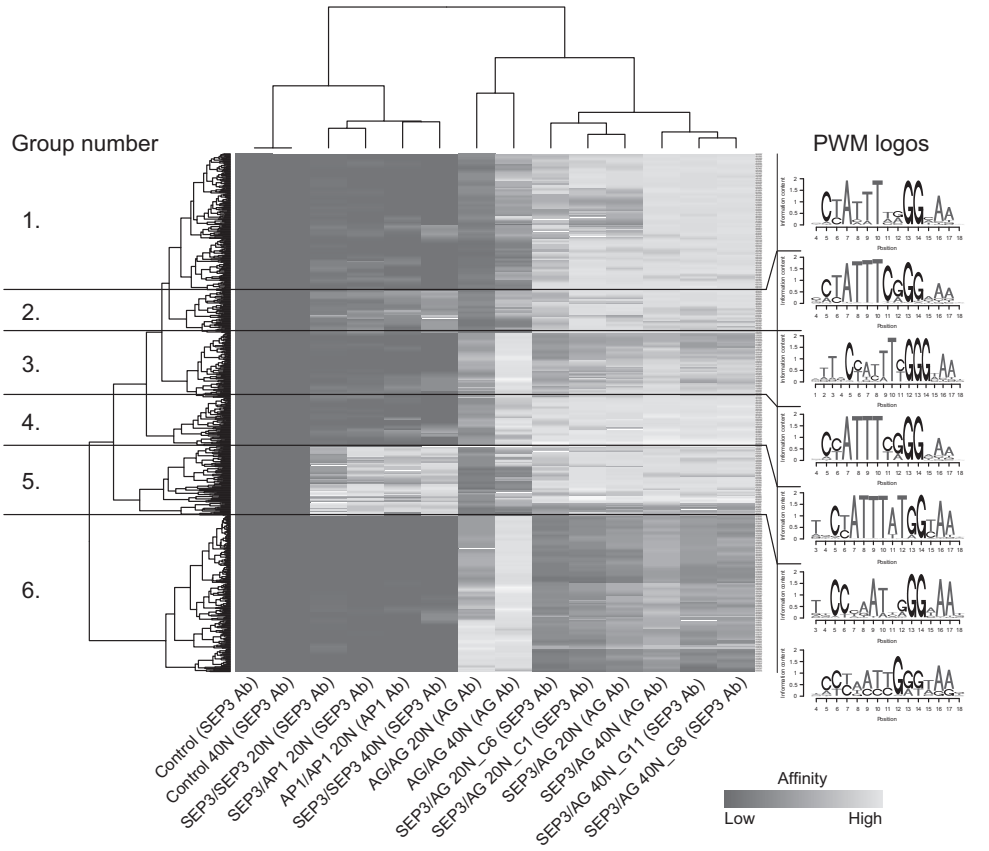


Figure 2. MADS-domain TF complexes DNA binding specificities. Heat map of 12-mer (DNA-binding sequence fragments) affinities enriched in the 3rd round of SELEX for all studied MADS TF complexes and corresponding sequence logos built from the position weight matrices (PWMs) for all 40N library sequences containing group specific 12-mers.

are separated by the middle, six-nucleotide fragment (position 6-12) (**Figure 2**). The most prominent differences between these motifs are in the composition of this middle fragment. In a perfect CArG-box, the middle fragment is composed of adenines (A) and thymines (T) only [therefore it is called adenine rich ('Ar')]. The perfect CArG-box binding pattern was shown by SEP3/SEP3, SEP3/AP1 and AP1/AP1 complexes (Group 5), however, with a strong preference to a specific adenine rich sequence: 'ATTTAT'. The motif discovery showed also that C6 and G13 are not completely conserved and can be replaced by either A or T, creating even longer A/T stretch in bindings sites specific for this group of dimers. On the other hand, the motif specific for AG homodimer had shorter A/T stretch in almost all cases (Group 3 and 6) and sometimes allowed C at position 8 in the A/T rich fragment.

The shorter A/T stretch characteristic for the AG/AG dimer was also reported before in SELEX experiments followed by Sanger sequencing (Huang et al., 1993). Not surprisingly, the SEP3/AG dimer showed consensus motif intermediate between SEP3/SEP3 and AG/AG dimers. This is especially visible in the Group 1 consensus site where the first part of the motif (position 4-9) resembles the left 'half-site' of the SEP3/SEP3 specific binding site (Group 5) and the other 'half-site' (position 10-17) reflects the AG/AG specific motif (Group 6). The lower affinity of AG/SEP3 to Groups 3 and 6 consensus sites, which are more specific for AG/AG homodimers, indicates that that these AG homodimers are underrepresented in the SELEX mixtures containing both SEP3 and AG proteins.

Analysis of the consensus binding sites revealed also that non-CArG-box-like sequences can be bound by MADS-domain protein dimers. This is shown by the second, overrepresented motif specific for the AG/AG dimer (Group 6). Although the C5,6 and G13,14 positions are relatively conserved, the middle fragment containing C9,10,11 is far from the perfect A/T stretch present in a CArG-box. Furthermore, until now, the MADS-box consensus site was represented by a 10 bp nucleotide fragment. Our SELEX experiments showed, however, that the possible consensus binding site for MADS-domain proteins could be longer. This is supported by the presence of highly conserved A16 and A17 and sometimes T2 and T3 nucleotides in the discovered motifs.

Mapping of in vitro binding sites to in vivo ChIP binding profiles – validation of the in vitro SELEX-seq approach

To validate our *in vitro* SELEX-seq approach we mapped all dimer-specific 13-mers to the genome of *Arabidopsis*. We choose a length of 13 bp, because it is the minimum length for the alignment application 'soapv2' (R. Li et al., 2009) to perform mapping of the sequence reads to a genome. Additionally, using estimated 13-mer affinity values we were able to profile the DNA binding sites identified by SELEX to the genome of *Arabidopsis* in a quantitative manner and compare them with the ChIP-seq data obtained for SEP3 and AP1 (Kaufmann et al., 2009; Kaufmann et al., 2010c) (**Figure 3A**).

Global comparison of binding site peak scores for SEP3/AG SELEX-seq and SEP3 *wt* ChIP-seq showed relatively good correlation (R^2=0.42) (**Figure 3A**), where highly scored ChIP-seq peaks revealed also high score of the predicted binding by SELEX-seq in the same genomic region. Additionally, the proportion of ChIP-seq peaks with SELEX-seq binding sites increases with higher ChIP-seq scores in AP1 and SEP3 ChIP-seq data, confirming good correlation between SELEX-seq and ChIP-seq experiments (**Figure 3C**). There were on average three SELEX-seq peaks within a single ChIP-seq peak, showing that multiple binding sites in close proximity can

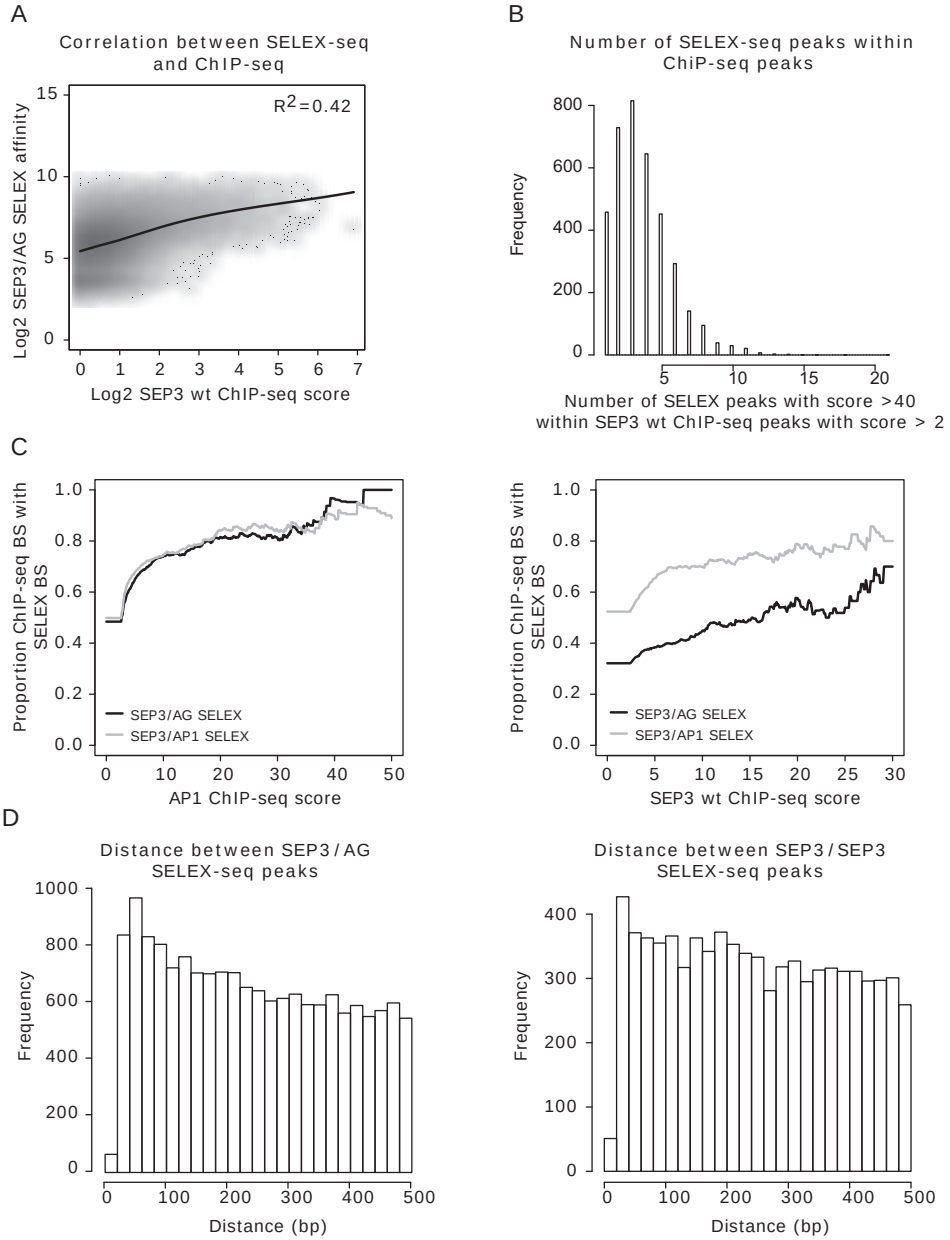


Figure 3. Comparison between *in vitro* SELEX-seq and *in vivo* ChIP-seq. (A) Correlation between estimated affinities of SELEX-seq 13-mers mapped to the genome of *Arabidopsis* and corresponding ChIP-seq peak scores. **(B)** Number of SELEX-seq peaks within corresponding ChIP-seq peaks. **(C)** Proportion of ChIP-seq binding sites with SELEX-seq binding sits as a function of ChIP-seq score. Left: AP1 ChIP-seq; right: SEP3 wt ChIP-seq. **(D)** Preferred distances between subsequent SELEX-seq peaks in the genome of *Arabidopsis*. Left: SEP3/AG SELEX-seq; right: SEP3/SEP3 SELEX-seq.

together constitute a functional binding site for MADS-domain TF complexes (**Figure 3B**), which is in agreement with the 'floral quartet' model (Theissen and Saedler, 2001), where tetrameric MADS-domain protein complexes bind to two binding sites at short distance from each other. The presence of more than two binding sites in close proximity could also suggest potential formation of higher-order MADS-domain protein complexes (hexameric or higher) at these sites. Moreover, the preferred distance between SELEX-seq binding sites was around 50 bp (**Figure 3D**), which is in an agreement with the previously calculated mean distance between CArG-boxes within SEP3 ChIP-seq binding sites (Kaufmann et al., 2009).

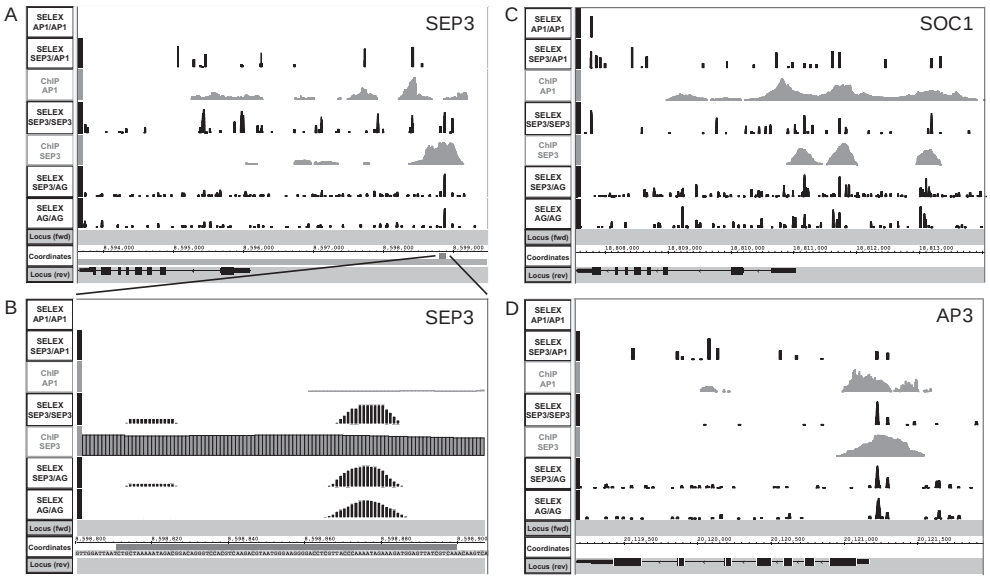


Figure 4. Examples of SELEX-seq binding sites mapped to the genome of *Arabidopsis*. (A) Promoter region of *SEP3* with SELEX-seq and related ChIP-seq binding profiles. **(B)** *SEP3* promoter region (-3.1 kb from ATG) with the regulatory sequence (purple) containing two CArG-box-like sequences that were correctly predicted as binding sites by SELEX-seq experiments. **(C-D)** Examples of MADS-domain TF target genes with ChIP-seq and SELEX-seq binding profiles.

To further analyze binding sites obtained with SELEX-seq we looked at particular genes regulated by MADS-domain proteins. The upstream promoter regions of the selected MADS-box genes: *SEP3*, *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS* 1 (*SOC1*) and *APETALA3* (*AP3*) were predicted by our SELEX-seq approach as binding regions of several MADS-domain TF dimers (**Figure 4**). The position of the SELEX-seq peaks is in very good agreement with the position of ChIP-seq peaks for AP1 and SEP3 *in vivo* binding. However, the resolution of SELEX-seq binding profiles is much higher comparing to profiles

generated by ChIP-seq. By combining the results of SELEX-seq and ChIP-seq we were able to predict exactly which sites in the regulatory regions of the target genes are bound. One of these binding regions, positioned 3.1 kb upstream in the promoter of SEP3, had two binding sites in close proximity that showed differential affinity (represented by the SELEX peak height) and specificity (not all dimer combinations

Table.1 Examples of genes predicted as specific targets of SEP3/AP1 (sepal-petal identity complex, AP1-domain) or SEP3/AG (stamen-carpel identity complex, AG-domain).

Gene family	Gene	Function
AP1-domian		
Homeobox genes	*PRESSED FLOWER* (*PRS*) [or *WUSCHEL RELATED HOMEOBOX 3* (*WOX3*)]	Regulation of lateral axis-dependent development in *Arabidopsis* flowers and cell proliferation (Matsumoto and Okada, 2001).
	KNOTTED1-LIKE HOMEOBOX GENE 4 (*KNAT4*)	Development of the *Arabidopsis* shoot apical meristem (Truernit et al., 2006; Truernit and Haseloff, 2007).
	TRIPTYCHON (*TRY*)	Trichome development on sepals (Schnittger et al., 1998).
SQUAMOSA-PROMOTER BINDING PROTEIN (SBP)-like (SPL) genes	*SPL2* and *SPL8*	Proper development of lateral organs (Shikata et al., 2009), micro- and megasporogenesis, trichome formation on sepals, and stamen filament elongation (Unte et al., 2003).
AP2 TFs	*RELATED TO AP2 2* (*RAP2.2*)	Pathogen resistance and ethylene responses (Zhao et al., 2012).
AG-domain		
MADS-box TFs	*AG* and *SHATTERPROOF1* (*SHP1*)	Carpel and ovule development (Liljegren et al., 2000).
AP2 TFs	*SCHLAFMUTZE* (*SMZ*)	Flowering repression (Mathieu et al., 2009).
FRI-related genes	*FRIGIDA* (*FRI*)	Major regulator of transition from vegetative to reproductive phase (Clarke and Dean, 1994)
Basic helix-loop-helix (bHLH) TFs	*SPEECHLESS* (*SPCH*)	Cell asymmetric divisions (MacAlister et al., 2007)
SUPERMAN-family genes [C(2)H(2)-type zinc finger genes]	*SUPERMAN* (*SUP*)	Flower-specific gene that controls the boundary of the stamen and carpel whorls (Sakai et al., 1995).

bound to both binding sites) for MADS-domain complexes. These results are in agreement with our previous *in vitro* EMSA studies of the same regulatory fragment (Smaczniak et al., 2012b) where these two binding sites showed variable binding efficiencies for different MADS-domain protein complexes with one site being superior in importance to the other. Another example are the binding sites bound by AP1-containing complexes *in vivo*, visualized by ChIP-seq binding profiles, which are located 2 and 2.5 kb upstream in the promoter of SEP3. These sites are mostly predicted by SELEX-seq results obtained for SEP3-AP1 dimers. Altogether, these results suggest that the SELEX-seq derived binding motifs are able to discriminate between favored binding sites for MADS-domain TF complexes.

SELEX-seq/ChIP-seq discrimination for specific target genes of MADS-domain TF dimers

Because our SELEX-seq experiments are able to discriminate between different MADS-domain dimers based on binding specificities, we tested whether SELEX-seq results of several MADS-domain dimers can be used to assign a DNA binding event identified by ChIP-seq to a particular dimer, and therefore be able to predict organ-specific target genes. As an example, we focused on the SEP3/AP1 (sepal and petal specific) and SEP3/AG (stamen and carpel specific) dimers. We pooled the DNA regions identified as bound in the SEP3 and AP1 ChIP-seq experiments and we only used the affinity ratios of the SEP3/AP1 and SEP3/AG dimers obtained by our SELEX-seq experiment to classify these pooled regions. The SELEX-seq affinity ratios were able to correctly distinguish differentially bound MADS-domain TF complexes (**Figure 5**). Next, we calculated affinity ratios between SEP3/AP1 and SEP3/AG SELEX-seq binding sites present within SEP3 ChIP-seq peaks, which allowed us to group SEP3 target genes into either SEP3/AP1 or SEP3/AG targets (based on two-fold ratio difference). To narrow the number of target genes obtained by such analysis we compared it with the organ-specific expression data (Jiao and Meyerowitz, 2010). We found 157 genes specific for sepals and petals (AP1-domain) and 176 genes specific for stamens and carpels (AG-domain) that were common between ChIP-seq, SELEX-seq and RNA-seq data (**Figure 5B**). Among genes characteristic for AP1-domain as well as for the AG-domain we found representatives of several TF families (**Table 1**).

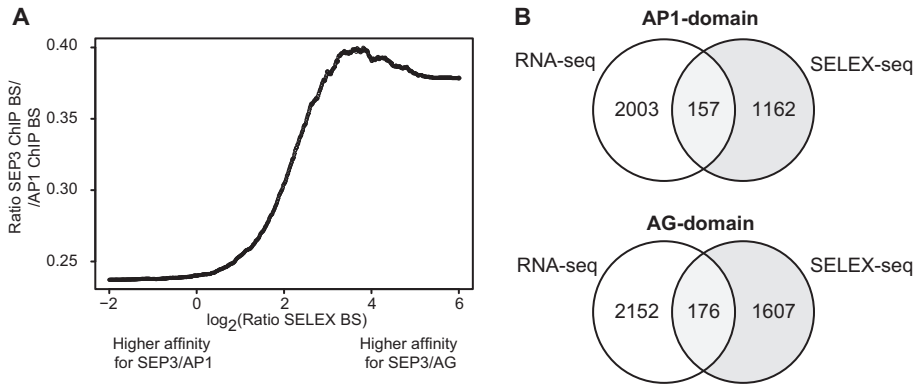


Figure 5. SELEX-guided target gene discrimination for MADS-domain protein dimers. (A) Prediction of dimer-specific target genes based on the SELEX affinity ratios and ChIP-seq score ratios. **(B)** Venn diagram: comparison between domain-enriched genes that exhibited significant (two-fold with $p < 0.05$) up-regulation as compared with the other domain (AP1 vs. AG domain) (Jiao and Meyerowitz, 2010) and target genes, specific for either SEP3/AP1 or SEP3/AG dimer based on the SELEX-seq analysis.

Discussion

The variety of functions of MADS-domain TFs in the life cycle of *Arabidopsis thaliana* suggests specific transcriptional regulation of their multiple target genes. How exactly MADS-domain TFs achieve their functional specificity is not yet fully understood. Here, we showed that part of the specificity comes from the distinct interactions of MADS-domain TF dimers with the DNA. By making use of the SELEX-seq approach we were able to distinguish specific and different binding sites for diverse MADS-domain TF dimers. Taking together our high-throughput *in vitro* data, the *in vivo* binding data and the organ specific expression data (Kaufmann et al., 2009; Jiao and Meyerowitz, 2010), we assigned a role to specific MADS-domain dimers in the regulation of target genes in a floral organ specific manner.

DNA binding specificities of MADS TFs

Although, many *in vitro* studies aimed to unravel the DNA binding specificities of MADS-domain TFs, only until recently, with the usage of high-throughput sequencing technologies, we are fully able to explore binding proprieties of these TFs. The selection of bound sequences from a pool of random oligonucleotides followed by Sanger sequencing was used previously to elucidate DNA-binding properties of several homodimeric MADS-domain TFs: e.g. AG, AGAMOUS-LIKE 15 (AGL15), SEPALLATA1 (SEP1), SEPALLATA4 (SEP4), and SHP1 (Huang et al., 1993; Huang et al., 1995; Huang et al., 1996; Tang and Perry, 2003). These experiments usually produced less than a hundred DNA sequences from which the consensus was built. Some features of these consensus logos are common to our SELEX-seq built logos using MADS-domain homodimers. For example, for most MADS-domain TFs the consensus motif resembles the perfect CArG-box of $CC[A/T]_6GG$ similar to our SELEX-seq logos (**Figure 6**). Moreover, positions 2, 3, 16, and 17 occupied by either T or A support our finding that the consensus binding site for MADS-domain proteins is longer than 10 bp. The A-stretch on the right end of the CArG-box was reported before for the in vivo consensus binding sequences of SEP3, AGL15 and

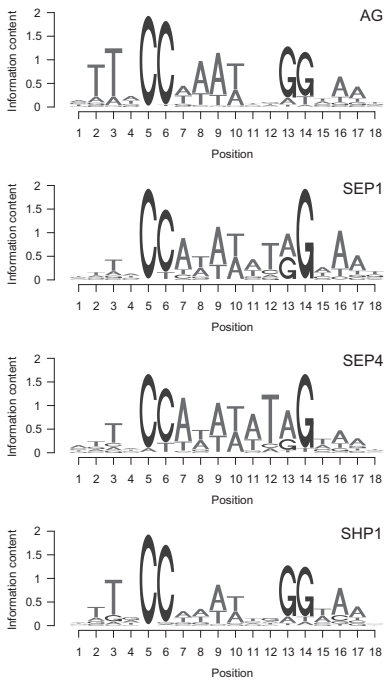


Figure 6. Binding sequence logos of MADS-domain proteins built from PWMs (according to Huang et al., 1993, 1995, 1996).

FLOWERING LOCUS C (FLC) (Kaufmann et al., 2009; Zheng et al., 2009; Deng et al., 2011), which suggest the extended length of the functional MADS-domain TF binding site. Furthermore, the lack of a T-stretch on the left end of the CArG-box in the consensus sequence characteristic for heterodimers and not for homodimers suggests that binding of heterodimers could be asymmetrical.

Consensus logos built for AG and SHP1 protein homodimers (Huang et al., 1993; Huang et al., 1996), which belong to the same MADS-domain protein subfamily, are very similar to one of the logos built for the AG/AG homodimer in our SELEX-seq approach. This suggests that binding might be specific for the MADS-domain TF subfamilies, a characteristic that must have originated from the DNA-binding domain sequence, which is highly conserved within MADS-domain protein subclades (Parenicová et al., 2003). This may be related to the partially redundant biological functions of members of the same subfamily [e.g. SHP1, SHATTERPROOF2 (SHP2) and SEEDSTICK (STK)] in ovule determination, Pinyopich et al., 2003).

Although position weight matrices (PWMs) or consensus sequences give substantial information on the DNA sequence characteristics that determine TF binding, they fail, for example, to visualize the dependencies between nucleotide positions or, most importantly, the affinities to particular DNA structures. Previously, it was reported that the MADS-domain protein complexes AP1/AP1, AP3/PI and AG/AG bind to the same DNA fragments containing CArG-boxes, however with different affinities (Riechmann et al., 1996a; Riechmann et al., 1996b). Based on these data and the functional analysis of chimeric proteins, where plant MADS domains were swapped with human MADS domains and the resulting chimeric proteins were still able to rescue plant MADS-box gene mutants, it was suggested, somewhat controversially, that the functional specificity is independent of the DNA-binding specificity of MADS-domain proteins (Riechmann and Meyerowitz, 1997; Krizek et al., 1999). Our high-throughput analysis, on the other hand, shows that there are differences in binding between selected MADS-domain dimers. These differences are represented in a sequence and affinity based manner – where both factors could play a major role in target gene selection and regulation. Our results support also the notion that different MADS-domain protein complexes could bind overlapping sets of binding sites, although with different affinities, which would allow for active competition between different MADS-domain protein complexes for the same target genes *in vivo*.

MADS domain protein - DNA complex structure and its functional specificity

Our SELEX-seq results with the SEP3/AG dimer suggest that MADS-box TF binding sites can be considered as two half-sites reflecting binding specificities from

SEP3 and AG, respectively. The differences in the core CArG-box that are preferentially bound by different dimers also suggest that DNA binding specificity is determined by the structure or the sequence of the DNA-binding MADS domain. Previous domain swaps experiments showed that functional specificity could be determined by the MADS and I domain for AP1 and AG proteins (Krizek and Meyerowitz, 1996). These studies, in relation to the MADS domain swaps between plant MADS-domain TFs and human MADS-domain proteins serum response factor (SRF) and Myocyte-specific enhancer factor 2A (MEF2A) (Riechmann and Meyerowitz, 1997; Krizek et al., 1999), show that the I region is very important for the *in vivo* functionality of these proteins. It is well possible that the dimer combination (mainly determined by the I region) determines the structure of the dimer, which contributes to the binding specificity of MADS-domain TF dimers.

Several studies aimed to unravel the MADS domain DNA binding determinants of MADS-domain TFs. The crystal structures of the MADS domain bound to the DNA was characterized for human SRF and MEF2A proteins and yeast Minichromosome maintenance 1 (Mcm1) protein (Pellegrini et al., 1995; Tan and Richmond, 1998; Huang et al., 2000) but not for any of plant MADS-domain TFs. These structural data provided substantial information about direct contacts of particular amino acid residues in the MADS domain to the DNA sequence. Residues R143, R157, K163 and R164 (corresponding to position 3, 17, 23 and 24 in **Figure 7**) of the SRF protein play crucial roles in DNA binding. These four amino acid residues seem to be highly conserved also in plant MADS-domain TFs (**Figure 7**). There are a number of critical differences in the protein-DNA interactions observed for the human and yeast MADS-domain TFs. In particular, the N-terminal end of the recognition MADS α-helix and the N-terminal extension of the MADS domain appear to be important for both DNA-binding affinity and specificity of these two TFs (Sharrocks et al., 1993; Nurrish and Treisman, 1995; Huang et al., 2000). Comparing to plant MADS-domain proteins, the N-terminal extension exist in the AG subclade only and might play a role in the specificity of DNA binding to AG-regulated target genes.

Positions 14, 15, and 16 of the MADS domain are highly variable within the MADS-domain TF family. Positions 14 and 16 of the MADS domain occupied by K and R respectively in SRF make direct contacts with the DNA (Pellegrini et al., 1995). It was shown for SRF and MEF2A that the 14th residue of the MADS-domain plays an important role in the dimer mediated DNA bending (West et al., 1997), which could have a role in regulating the biological specificity for those TFs. Non-conserved positions 14-16 within plant MADS-domain TFs, especially position 14, suggest that individual members of the MADS-domain TF family might induce DNA-bending to a different extent.

MADS-domain TFs bind DNA through interactions of the N-terminal part of the MADS domain with the CArG-box A/T-rich region of the minor groove (Pellegrini et al., 1995) causing substantial bending of the DNA. Although crystal structures for plant MADS-domain TFs are not available, it was shown that also plant MADS-domain proteins bend the DNA significantly towards the minor groove (Riechmann et al., 1996b; Melzer et al., 2009). Additionally, it was reported that DNA-bending of MADS-domain complexes could be DNA-sequence specific (West et al., 1997; West et al., 1998), which supports the importance of the DNA sequence in the regulation of a protein-DNA complex structure. According to the floral quartet model (Theissen and Saedler, 2001), MADS-domain TF complexes bend the DNA in order to bind two different binding sites simultaneously as a quaternary protein complex. What are the determinants of this characteristic binding is not well understood. Above we discussed that the protein complex can differentially bend the DNA upon sequence-specific DNA binding. However, it was also shown that the intrinsic properties of the DNA like its shape (bend) can also influence DNA binding (Rohs et al., 2009).

Figure 7. MADS domain sequences of the MADS-domain TF family. MADS domain sequence logos of some subfamilies of the MADS MIKC Type TFs.

DNA sequence shapes the DNA-binding site

The A/T-rich region of the CArG-box sometimes can be considered as an 'A-tract'. An A-tract is a stretch of 4–6 adenine base pairs in the DNA sequence (Haran

and Mohanty, 2009) that causes DNA bending towards the minor groove (Koo et al., 1986) and when distributed with the helical periodicity in the genome, leads to intrinsic global curvature of the DNA. Moreover, A/T-rich regions, whether they are composed of the ApT steps (a stretch of adenines followed by a thymine) or TpA steps (a stretch of thymines followed by an adenine), can narrow or widen the minor groove of the DNA respectively. Based on our SELEX-seq sequence logos we can infer that there are differences in the length of the A/T-rich regions in the consensus sequence specific for either homo- or heterodimers. For example, the AG homodimer usually prefers to bind sequences with shorter A-tracts while the SEP3 homodimer binds sequences with long A-tracts. The SEP3/AG heterodimer binds sequences with moderate A-tract length. This suggests that not only DNA sequence but also DNA structure, especially the width of the CArG-box minor groove, plays a role in the recognition of specific binding sites by particular MADS-domain protein complexes, a phenomenon that was also observed for other transcriptional regulators (Rohs et al., 2009; Slattery et al., 2011). Although A-tracts could have a role in the recognition and binding specificity of MADS-domain TFs, it is static and cannot account for a dynamic binding of TFs which is needed for the regulation of gene expression in a temporal and spatial manner. Therefore, differential DNA bending induced by the TF dimer plays most likely a more prominent role in TF-DNA binding and gene regulation (Haran and Mohanty, 2009).

The role of MADS TF multimerization in MADS DNA binding specificity

In general, MADS-domain consensus binding sequences are relatively similar, as we discussed already above. Therefore, it seems unlikely that the functional specificity of MADS-domain TFs can be attributed to the DNA-binding specificity of a MADS-domain protein dimer only. Relatively recently, high throughput *in vivo* DNA binding experiments showed that MADS-domain proteins bind the DNA in places that lack the canonical CArG-box (or CArG-box-like) sequences (Kaufmann et al., 2009; Zheng et al., 2009; Kaufmann et al., 2010c; Deng et al., 2011). Moreover, targeted *in vivo* immunoprecipitation experiments of several MADS-domain TFs revealed that MADS-domain proteins can form larger complexes with other transcriptional regulators (Smaczniak et al., 2012b), and as such could bind the DNA. Our SELEX-seq experiments showed the specificity for binding sites for several MADS-domain homo- and heterodimers, which revealed differences depending on the composition of the MADS-domain dimer. Comparing SELEX-seq with ChIP-seq binding profiles we were able to crack part of the MADS cis-regulatory code, e.g. which dimer combination binds to a particular DNA sequence *in vivo* and subsequently may regulate the expression of the corresponding target gene. Whether the presence of other cofactors that could interact with the MADS-domain TFs could

modulate MADS-domain binding specificity remains an important question to be resolved in future studies.

Materials and methods

SELEX-seq

The dsDNA libraries were obtained from the ssDNA sequences by single-cycle PCR amplification with a complementary primer essentially as described before (Jolma et al., 2010). The dsDNA libraries contained either 20 or 40 random nucleotide fragments flanked by specific barcodes that allowed for later characterization when multiplexed in high-throughput sequencing. The dsDNA libraries contained all necessary features required for direct sequencing with an Illumina Genome Analyzer (Jolma et al., 2010).

Proteins dimers were synthesized using TNT SP6 Quick Coupled Transcription/Translation System (Promega) following the manufacturer's instructions in a total volume of 20 µl and equimolar expression plasmid concentrations. The binding reaction mix was prepared essentially as described previously for EMSA experiments (Egea-Cortines et al., 1999; Smaczniak et al., 2012b) and contained 20 µl of *in vitro*-synthesized proteins and 50-100 ng of dsDNA library in a total volume of 120 µl. The binding reaction was incubated on ice for 1 h followed by 1 h immunoprecipitation with protein specific antibodies coupled to magnetic beads (MyOne, Invitrogen) in thermomixer at 4 °C with constant mixing at 700 rpm. Magnetic beads with attached antibodies where prepared in advance according to manufacturer's instructions (MyOne, Invitrogen) with purified antibodies resuspended in 1X PBS; 0.5 mg of beads was used for a single binding reaction. After immunopreciptiation, beads were washed 5 times with 150 µl of binding buffer without salmon-sperm DNA and bound DNA was eluted with 50 µl 1X TE in thermomixer at 90 °C with full mixing speed. Afterwards, magnetic beads were immobilized and the supernatant was transferred to a 1.5-ml tube. DNA fragments were amplified with 10 to 15 cycles of PCR with SELEX round-specific primers (Jolma et al., 2010) and the total amplicon was used in the subsequent SELEX round. The amplification efficiency was checked on the agarose gel by comparing to a known concentration of a standard probe. Samples for sequencing, after amplification, were cut out from agarose gel and purified using MinElute Gel Extraction Kit (Qiagen). Different libraries were multiplexed by mixing in an equmolar amounts in the Elution Buffer (Qiagen) and sequencing was performed on the HiSeq 2000 sequencer (Illumina).

SELEX-seq data analysis

Data obtained from HiSeq 2000 system were extracted and grouped according to library specific barcodes. Sequence reads that didn't pass the filter quality of CASAVA 1.8 or that were present in the library in an unexpected high number (>1,000) were eliminated. Data analysis was essentially performed as described before by (Slattery et al., 2011). Relative affinity for each possible k-mer of length *m* was calculated as the ratio between the frequencies of k-mers in round 0 to round 3, and normalize to 1 by dividing for the highest affinity-predicted k-mer. Frequency of k-mers in round 0 was predicted by a MonteCarlo model of order 6.

To *in silico* predict genomic regions bound by a given MADS-domain dimer based on our SELEX-seq experiments, we obtained the affinity value for each k-mer of length 13 bp, and mapped them to the TAIR10 genome with the soapv2 (R. Li et al., 2009) allowing no mismatches and without sequence reads that map into multiple locations. The 13 bp regions, where a 13-mer reads correctly mapped, were given a score value equal to the estimated affinity of that particular 13-mer and regions where several 13-mers overlapped were given the score equal to the sum of the affinities of each 13-mer.

DNA PWMs and logos

PWMs were calculated based on the extracted 12-mer-containing sequences with the GADEM algorithm (Li, 2009) and DNA sequence logos were built with the 'seqLogo' R script.

EMSA

Biotin-labeled SELEX-derived sequences were produced by PCR with biotin-labeled primers and purified from 2% agarose gel. Electrophoretic mobility shift assays were performed essentially as described before (Smaczniak et al., 2012b).

Acknowledgments

We would like to thank Arttu Jolma and Jussi Taipale for providing barcoded ssDNA libraries and sharing details on the dsDNA library preparation and amplification.

Supporting Information (SI)

SI Figures

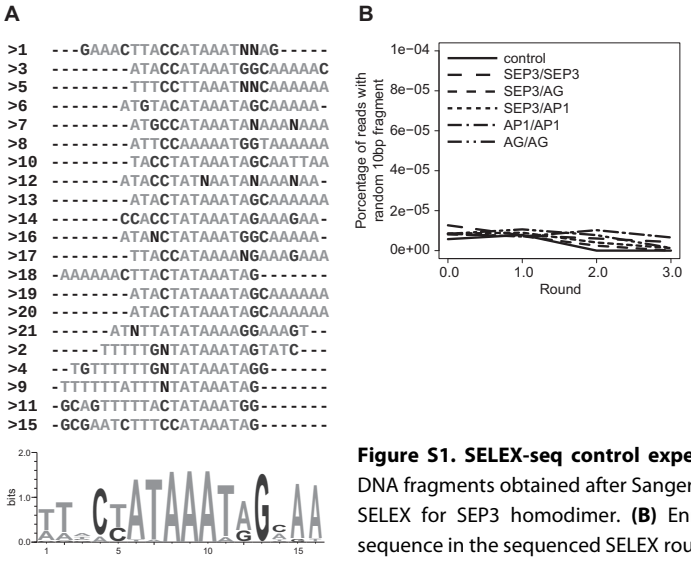


Figure S1. SELEX-seq control experiments. **(A)** Alignment of 21 DNA fragments obtained after Sanger sequencing of the 6th round of SELEX for SEP3 homodimer. **(B)** Enrichment of a random 10 bp sequence in the sequenced SELEX rounds.

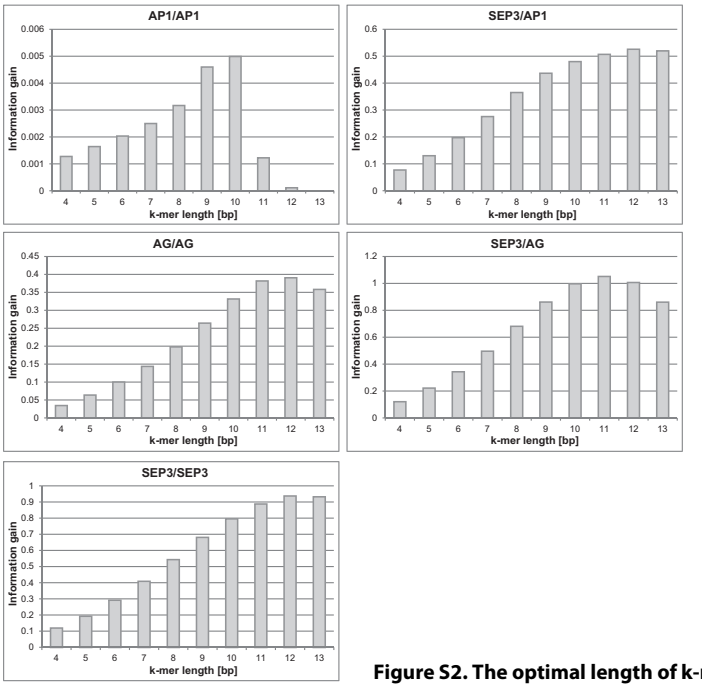


Figure S2. The optimal length of k-mers in SELEX-seq libraries.

Chapter 6

Going beyond the CArG-box: understanding the mechanism of DNA recognition by plant MADS-domain transcription factors

Jose M. Muiño[1,2,3,*], Cezary Smaczniak[1,4], Gerco C. Angenent[1], Kerstin Kaufmann[4], Aalt D.J. van Dijk[1]

[1] Business Unit Bioscience, Plant Research International, 6708PB Wageningen, The Netherlands
[2] Laboratory of Bioinformatics, Wageningen University, 6708PB Wageningen, The Netherlands
[3] Present Address: Computational Molecular Biology, Max Planck Institute for Molecular Genetics, D-14195 Berlin, Germany
[4] Laboratory of Molecular Biology, Wageningen University, 6708PB Wageningen, The Netherlands
[*] To whom correspondence should be addressed: Tel: +49-30-841-31169; Fax: +49-30-8413-1152; E-mail: muino@molgen.mpg.de

Manuscript in preparation.

Abstract

MADS-domain transcription factors in plants act as key regulators of many developmental processes. Despite the wealth of information that exists about these factors, the mechanisms by which they recognize their cognate DNA-binding site, called the CArG-box (consensus CC[A/T]$_6$GG), and how different MADS-domain proteins achieve DNA-binding specificity is still largely unknown. We used information from *in vivo* ChIP-seq experiments, *in vitro* DNA-binding data and evolutionary conservation to address these important questions. We found that structural characteristics of the DNA play an important role in the DNA-binding of plant MADS-domain proteins. The central region of the CArG-box largely resembles a structural motif called 'A-tract', which is characterized by a narrow minor groove and may assist bending of the DNA by MADS-domain proteins. Periodically spaced A-tracts outside the CArG-box suggest additional roles for this structure in the process of DNA-binding of MADS-domain proteins, such as for the binding of higher-order protein complexes. We examined the role of temperature on DNA-binding of MADS-domain proteins using *in vitro* temperature experiments and found that temperature affects binding affinities possibly due to changes in DNA structure. Furthermore, the data show that structural characteristics of the CArG-box do not only play an important role in the MADS-domain protein binding, but can also partly explain differences in the DNA-binding specificity of different MADS-domain proteins and their heteromeric complexes.

Introduction

The MADS domain is a conserved DNA-binding domain present in a eukaryote-wide family of transcription factors (TFs). MADS-domain proteins typically contact their cognate binding site (BS), the CArG-box (consensus: CC[A/T]$_6$GG) as dimers. Structural analysis of animal and yeast MADS-domain protein dimers revealed that central parts of their MADS domains form an antiparallel coiled-coil, made of two amphipathic α helices – one from each subunit, which is positioned in the minor groove of the central [A/T] part of the CArG-box. The N-terminal regions penetrate into the minor groove and stabilize bending of the DNA. The C-terminal part of the MADS-domain forms mainly β-sheets that allow protein dimerization (Pellegrini et al., 1995; Tan and Richmond, 1998; Huang et al., 2000).

The family of MADS-box genes has mostly expanded in plants, and in particular in flowering plants. Two major classes of MADS-domain proteins can be distinguished: type I proteins, which are a heterogeneous group of proteins having only the MADS-domain in common, and the type II proteins, which have a highly conserved domain structure (Smaczniak et al., 2012a). In type II proteins, which are

also called MIKC-type proteins, the MADS-domain is followed by an intervening (I) domain, which likely forms an alpha helix and contributes to the selection of dimer partners. After the I-domain, a keratin-like (K) domain is located, which presumably assembles into coiled-coil structures enabling dimeric and higher-order complexes formation, followed by a highly variable C-terminus which has roles in transcriptional regulation (Kaufmann et al., 2005b). MIKC-type genes function as master regulators of developmental phase transitions as well as meristem and floral organ specification in flowering plants. They function together in a combinatorial manner, since the proteins interact with each other forming heterodimers and higher-order complexes (Egea-Cortines et al., 1999; Honma and Goto, 2001; de Folter et al., 2005; Smaczniak et al., 2012b; for review, see Immink et al., 2010).

Several variants of the CArG-box exist (Nurrish and Treisman, 1995). Their main distinguishing feature is the length of the [A/T] rich region in the central portion of the motif. Different MADS-domain proteins also differ in their ability to bend DNA around their binding sites (Riechmann et al., 1996b). For example, the mammalian MADS-domain factor Myocyte-specific enhancer factor 2A (MEF2A), which hardly induces DNA-bending, has the consensus binding motif CTA[A/T]$_4$TAG, while the consensus sequence of serum response factor (SRF) and yeast Minichromosome maintenance 1 (MCM1) reflects the standard CC[A/T]$_6$GG consensus (Pollock and Treisman, 1991; Nurrish and Treisman, 1995). Also in plants, some differences in the bending upon MADS-domain TF binding event have been reported (Huang et al., 1996; West et al., 1998), although the main determinants of binding site recognition and specificity have remained enigmatic. Given their various important and specialized roles in plant development, understanding the mechanisms of DNA-binding site recognition of plant MADS-domain TFs is an intriguing question.

The identification of *in vivo* DNA binding events of MADS-domain TFs at genome-wide scale provides new opportunities to study parameters and factors influencing DNA-binding site recognition. Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) or hybridization to tiling arrays (ChIP-CHIP) has allowed to generate genome-wide binding maps of several MADS-domain TFs involved in floral transition (Immink et al., 2012; Tao et al., 2012) and flower development (Kaufmann et al., 2009; Kaufmann et al., 2010c; Wuest et al., 2012). Especially a study on the floral MADS-domain TF SEPALLATA3 (SEP3), which acts as a mediator of higher-order interactions among floral MADS-domain proteins, has revealed that the consensus CArG-box has only poor predictability for DNA-binding *in planta* (Kaufmann et al., 2009): only 7.7% of all perfect CArG-boxes are bound by SEP3. The data suggested that dependencies between nucleotides as well as nucleotides outside the core CArG-box motif may contribute to binding site

recognition, and therefore redefinition of the CArG-box as motif representing the DNA binding event of MADS-domain TFs is needed.

In this paper, we analyze the DNA structural properties of CArG-box regions bound by specific MADS domain TFs. We found that properties associated with DNA curvature and flexibility are overrepresented among functional CArG-boxes. Our results also suggest that curvature of the DNA may play a role in determining the DNA-binding specificity of different MADS-domain dimers.

Materials and methods

Bioinformatics analysis of ChIP experiments

ChIP-seq experiments were analyzed as previously described (Kaufmann et al., 2010a). Sequence reads were mapped to the *Arabidopsis thaliana* (TAIR9) genome and significant read-enriched regions were detecting using CSAR (Muino et al., 2011). For ChIP-chip experiments, probe sequences were remapped to the TAIR9 *Arabidopsis* genome with the Starr package (Zacher et al., 2010). Only probes that mapped to unique locations were retained. Subsequently, CisGenome (Ji et al., 2008) was used to detect potential binding regions, using the hidden Markov model to combine intensities of neighboring probes. In this case the score value range between 0 and 1, where '1' was the most significant.

The maximum ChIP score value for the 10 bp region of each CArG-box motif present in the *Arabidopsis* genome was obtained from the ChIP-seq or ChIP-CHIP analysis described above. The datasets for SEP3 (Kaufmann et al., 2009), SOC1 and SVP (Tao et al., 2012) were re-analyzed in this study.

Predicting DNA structures

Dinucleotide properties were obtained from the DiProDB database (Friedel et al., 2009). They were used to estimate several properties of the DNA at each dinucleotide step. From these properties, we calculated average differences between the set of regions identified as bound by SEP3 in our ChIP-seq analysis comparing to unbound regions. Using these properties we calculated the DNA structure characteristics using X3DNA as previously described (Lu and Olson, 2008).

DNA conservation studies

The aligned DNA sequences of 80 *Arabidopsis thaliana* accessions were obtained from the 1001 genome project (http://www.1001genomes.org; release 2010_05_12). CArG-box motifs were located and associated with the SEP3 ChIP-seq score in the Col-0 accession and their corresponding sequences were extracted from other *Arabidopsis* accessions. Depending on the A-tract length, each CArG-box motif was classify in two groups, functional (length 4-6) and non-functional (length <4).

Among the sequences that show at last one nucleotide variation, the proportion of conserved A-tract lengths was calculated.

EMSA

Electrophoretic mobility shift assays were performed with *in vitro* synthesized proteins and the biotin-labeled DNA fragment, essentially as described before (Smaczniak et al., 2012b). The oligonucleotide sequence was derived from the first intron of the AG locus and contained a single CArG-box: 5'-TGAATATTATATATATT-CCAAATAAGG-AAAGTATGGAACGTT. ssDNA oligonucleotides were synthesized with the biotin attached to the 5'-end of the forward strand. Two complementary oligonucleotide strands were annealed before the EMSA.

Kd values estimation for protein-DNA complexes

K_d values were estimated essentially as described before (Riechmann et al., 1996b) based on the *in vitro* EMSA experiments by incubating a fixed amount of *in vitro* translated proteins (2 µl of the reaction mix) with increasing amounts of the DNA probe. The DNA probe used in the K_d calculations was the same as the '*SEP3 wt*' fragment described before (Smaczniak et al., 2012b).

Results

CArG-boxes bound by SEP3 complexes have particular DNA structural properties

Available methods for ChIP-seq data analysis aim to identify genomic regions that are bound by the protein of interest independently of the underlying DNA sequence. The posterior association of a particular DNA sequence/motif with these binding regions is complicated by the lack of resolution of ChIP-seq experiments. To overcome this problem, we modified the R package CSAR (Muino et al., 2011) to generate read-enrichment score values at each single-nucleotide position, and to extract the maximum score value within the region defined by the presence of our sequence/motif of interest (see Materials and Methods). This allowed us to focus on the DNA sequences/motifs bound by the TF avoiding the challenging task of defining the limits of the DNA binding region.

To understand the specificity of the SEP3 DNA binding, we studied the influence of DNA structural properties of 'functional' CArG-boxes as identified by the SEP3 ChIP-seq data. We estimated more than 110 DNA structural properties as defined in the dinucleotide property database (DiProDB, Friedel et al., 2009) for each dinucleotide step of the 50 bp region around all CArG-boxes in the *Arabidopsis* genome (7,742). We tested for difference in average dinucleotide properties for every nucleotide of the 10 bp CArG region using a *t*-test and among the top 10 most

significant different properties were: 'Flexibility slide', 'Tilt stiffness', and 'Minor groove width'; properties related to the DNA flexibility and curvature.

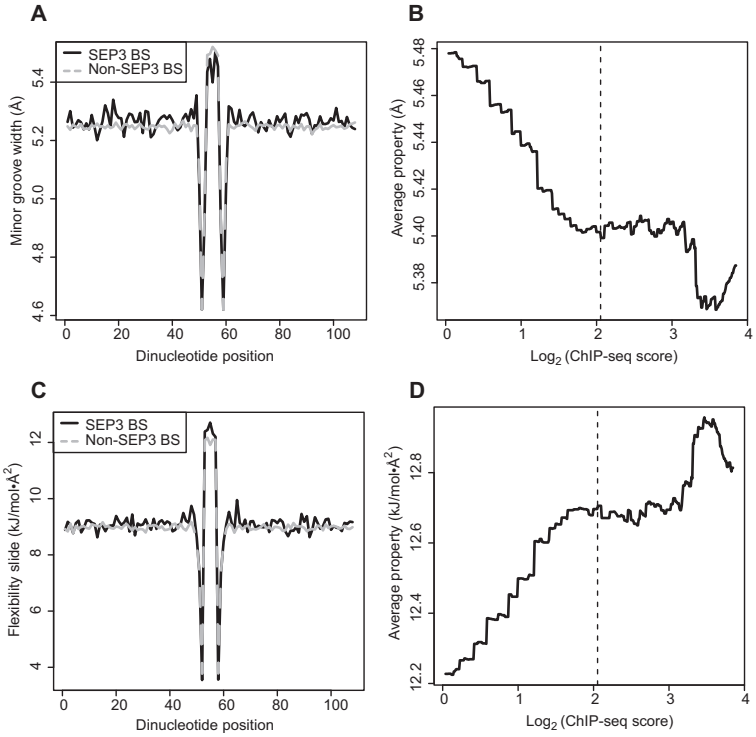


Figure 1. DNA structure properties of functional CArG-box regions. Average dinucleotide values for the property 'Minor groove width' **(A)** and 'Flexibility slide' **(C)** in CArG-box regions bound by SEP3 (black) and unbound (dashed grey). The 10 bp CArG-box motif is located in position 50-60. *T*-test statistic confirms a significant difference ($p<0.05$) between these two set of regions. To confirm that this differences are associated with the SEP3 ChIP-seq score, the relationship between the properties 'Minor groove width' **(B)** and 'Flexibility slide' **(D)** of the dinucleotide position, which show the highest significant difference in graph **A** and **C** and the SEP3 ChIP-seq score (log2 score) at the X-axis, is plotted. Dashed vertical line in (**D**) indicates the SEP3 ChIP-seq threshold for FDR<0.05. Both properties show a strong correlation with the SEP3 ChIP-seq score.

Figure 1 shows an average dinucleotide property among the CArG-box regions bound and unbound by SEP3 for 'Flexibility slide' and 'Minor groove width'. **Figure 1 B** and **D** shows the relationship between the average value of the DNA property and the SEP3 ChIP-seq score, illustrating that for the 'Flexibility slide' the average value increases with the ChIP-seq score, and for the 'Minor groove width' the regions bound by SEP3 are narrower.

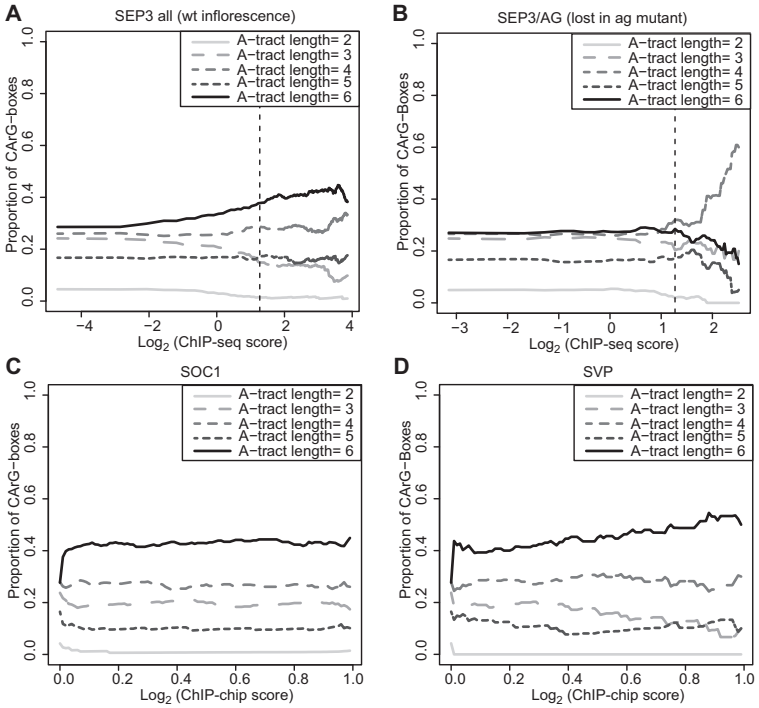


Figure 2. Enrichment of A-tract motifs in CArG-box regions. The proportion of CArG-box motifs with an A-tract element inside (length 4-6) increase depending on the ChIP score. In contrast, the proportion of CArG-box motifs without an A-tract (length 2-3) decreases with the ChIP score. **(A)** SEP3 ChIP-seq. **(B)** SEP3 ChIP-seq regions that loose the binding event in the *ag* mutant. **(C)** SOC1 ChIP-chip. **(D)** SVP ChIP-chip.

A-tracts are overrepresented in SEP3-bound CArG-box sequences

The structural properties of functional CArG-boxes that are detected in our analysis show striking similarities with the properties of DNA elements known as A-tracts. A-tracts have been defined as four to eight consecutive A/T base pairs without a TpA step (Stefl et al., 2004). The consensus of one A-tract can be described with the motif: $N_iA_mT_nN_j$, where $m+n>=4$ and the total length of the motif being 10 bp. DNA regions containing in-phase A-tract repeats show a narrower minor groove width and higher bendability towards the minor groove than other AT-rich regions.

Because of their structural and sequence similarities, we studied how the presence of one or more A-tracts in the CArG-box region (510 bp) influences the binding of SEP3. **Figure 2 A** and **B** shows that the proportion of DNA regions containing an A-tract (4>m+n>6) inside of the 10 bp CArG-box increases with the ChIP-seq score threshold used, supporting the idea of its positive relationship. In contrast, the proportion of regions without an A-tract inside the CArG-box tends to

decrease with the threshold used. A-tracts of length 4 and 6 show the highest enrichment in the SEP3 ChIP-seq experiment studied; this observation also holds for other MADS-domain TF ChIP(-chip) experiments (**Figure 2 C** and **D**).

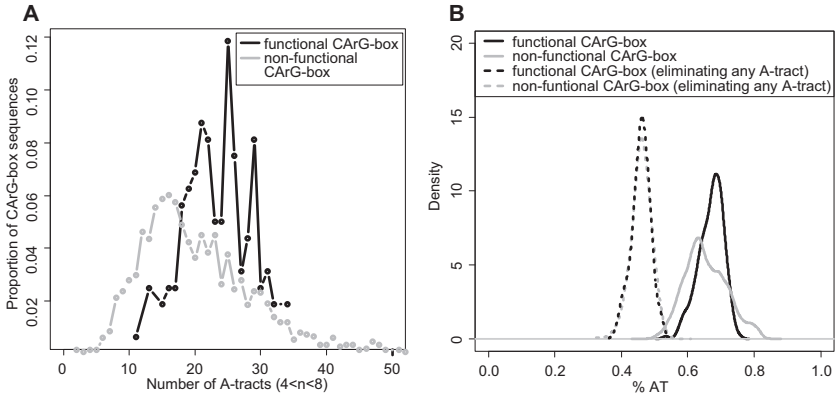


Figure 3. Multiple A-tracts in functional CArG-box regions. (A) Distribution of multiple A-tracts elements for 510 bp CArG-box regions bound by SEP3 (black) or not bound by SEP3 (grey). **(B)** This difference is not due to a different AT content of the regions, since when the A-tracts elements are eliminated both set of regions have the same AT-content (dashed line), only when the A-tracts elements are considered the set of regions have a different AT-content distribution (continuous line).

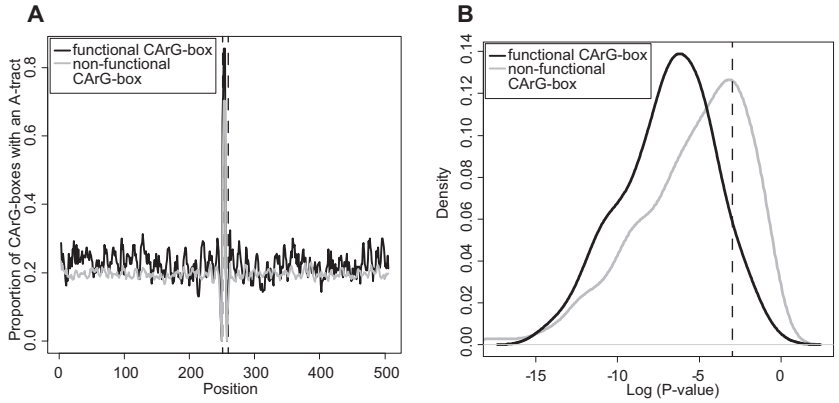


Figure 4. Periodically distribution of A-tract elements. For regions 500 bp around the CArG-box motif (position 250-260). **(A)** The proportion of regions with an A-tract (length 4-6) in the position determine by the x-axis. Regions with a SEP3 ChIP-seq binding event are indicated in black, and regions without a binding event are indicated in grey. For each CArG-box region with an arbitrary length of 150 bp the periodicity of A-tract position can be tested using the Fisher's test. **(B)** The distribution of log p-values for the Fisher's test for each sequence in the group of regions bound by SEP3 (black), and not bound (grey). P-values lower than 0.05 (indicated with a dashed vertical line) indicate a statistically significant periodicity for the A-tact location. Regions bound by SEP3 show a distribution with more significant p-values.

The surroundings of CArG-boxes (510 bp) bound by SEP3 are also characterized by a higher presence of A-tracts than non-bound CArG-box regions (**Figure 3A**), this is not due to a different CG/AT content, since when we eliminate the A-tracts from the studied regions the difference in AT-content is almost identical (**Figure 3B**). These A-tracts are not located randomly, but they show a significant periodicity of 11 bp (Fisher's periodicity test; $p < 0.05$; **Figure 4**).

A-tract DNA curvature may play a role in the DNA-binding specificity of MADS domain proteins

Because the A-tract length is related with the degree of curvature of the DNA region where it is located and because several MADS-domain protein homo- and heterodimers bend the DNA *in vitro* at different degrees (Riechmann et al., 1996b; West et al., 1997), we studied the preference of A-tract length of individual MADS-domain protein complexes. DNA regions detected by the SEP3 ChIP-seq experiment in wild-type (wt) but not in the *agamous* (*ag*) mutant (Kaufmann et al., 2009) are expected to be mainly bound by protein complexes containing SEP3 and AG. These DNA regions are enriched in CArG-boxes with an A-tract of length 4 (**Figure 2B**), in contrast to the preferences of length 4 and 6 in the wt data (**Figure 2A**). These results indicate that some MADS-domain protein complexes, e.g. the SEP3-AG heterodimer, may have a preference for CArG-boxes with particular A-tract properties.

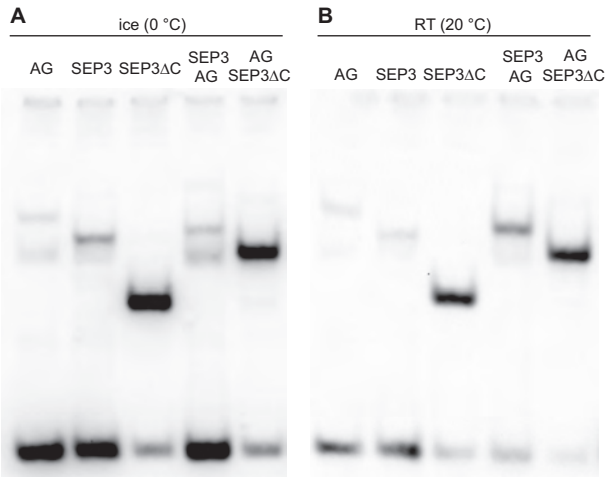


Figure 5. Identification of MADS-box dimers in EMSA. EMSA of several MADS dimers incubated on ice (0 °C) **(A)** and at room temperature (20 °C) **(B)** with a probe representing a part of the AG intron. From this figure we can identify the position of each dimer bound to DNA. At 0 °C, AG homodimer and SEP3-AG heterodimer show two bands but the lower one disappears at room temperature.

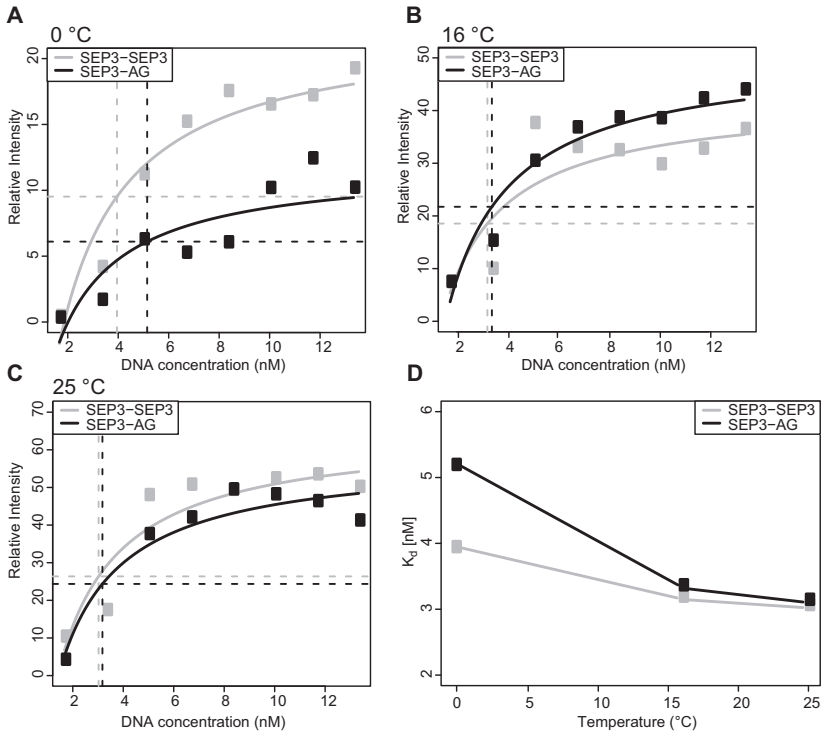


Figure 6. K_d calculations of MADS-domain-DNA complexes based on EMSA. (A-C) Calculation of the K_d from a series of EMSAs done at different temperatures. **(D)** Summary of the K_d obtained from these experiments. The change in affinity is temperature-dependent and its temperature-dependency is specific for each dimer.

DNA curvature of regions containing A-tracts strongly depends on the temperature. It has been shown that these DNA fragments increase progressively in curvature as the temperature decreases (Prosseda et al., 2010). This property enables us to experimentally study the importance of DNA curvature in the DNA binding affinity and specificity of MADS-domain proteins. We used electrophoretic mobility shift assay (EMSA) experiments to study the *in vitro* DNA affinity of two MADS-domain protein dimers (SEP3 homodimer and SEP3-AG heterodimer) to a 40 bp probe representing the *AG* intron at different temperatures (**Figure 5**). Strikingly, the relative binding of these dimers changes dramatically with the temperature; the SEP3 homodimer showing stronger binding at lower temperatures, while the SEP3-AG heterodimer at higher temperatures. This supports the hypothesis that SEP3-AG may have higher affinity for less curved DNA than the SEP3 homodimer, and that the DNA curvature may play a role in the DNA specificity of these two dimers. This is in agreement with the enrichment of short length A-tracts associated with the *in vivo* SEP3-AG binding, since decreasing length of A-tracts is associated with lower degrees

of DNA curvature. This temperature-dependent change in affinity is not related with a change in the proportion of SEP3 homo- and heterodimer in the SEP3/AG mixtures at both temperatures, because at low temperature we can observe a band at the position of the SEP3 homodimer when we incubated SEP3 together with AG and the DNA probe (**Figure 5**). It is unlikely that the temperature of the EMSA experiment have an important influence on the protein-protein interaction in this system. However, we have observed two bands when AG is incubated with the studied probe at 0 °C which may indicate two isoforms of the AG-DNA complex. At room temperature only one band remains. When we incubated SEP3, AG and the probe at 0 °C, also two bands appears, one corresponding to SEP3-AG and another to the second AG isoform-DNA complex. This second band also tends to disappear with the increasing temperature. This could reflect the different affinity to the DNA or differential formation of both protein complexes of AG isoforms depending on the temperature, since the *in vitro* translation is always done at the same conditions; we expect to have always the same relative concentration of each isoform. The fact that one band disappeared at high temperature (**Figure 5B**) can only be explained by a temperature-dependent presence of the AG-DNA complex, which could be explained by several mechanisms, among them: change in DNA affinity, change in protein-protein affinity (**Figure 6**) and change in degradation rate.

A-tract length is conserved among Arabidopsis ecotypes

Another evidence for the importance of A-tract length can be given by DNA sequence conservation studies. The proportion of 10 bp Col-0 CArG-box sequences with conserved length of their A-tract among the 80 sequenced *Arabidopsis* ecotypes (1001 genome project) is higher in regions bound by SEP3 TF complexes than in regions without SEP3 binding (**Figure 7**). In contrast, the proportion of CArG-box sequences with conserved length of consecutive A and T base pairs for non-functional A-tracts (length < 4 bp) decreases with the SEP3 ChIP-seq score. This supports not only the functionality of the A-tract inside the CArG-box sequence but also the importance of its length.

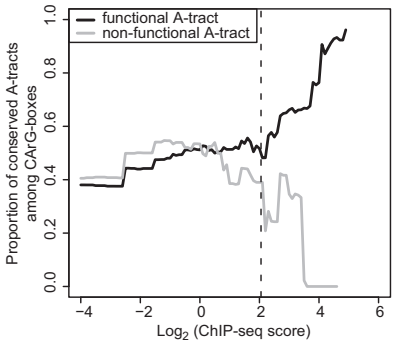


Figure 7. Conservation of the A-tract length in functional CArG-box regions. The average proportion of conserved A-tract length located inside the CArG-box motif among the 81 *Arabidopsis* ecotypes are shown in relation to the SEP3 ChIP-seq score threshold for functional A-tracts with length 4-6 (black) and for AT-regions with length 2-3 (grey). Only CArG-box with a functional A-tract has a positive relationship with the SEP3 ChIP-seq score threshold.

Predicting functional CArG-boxes

We can use the information obtained in this study to improve our definition of the DNA binding event of MADS-domain proteins. Although it is possible to predict the structural parameters of small DNA fragments and therefore their general characteristics, structure prediction alone was not very informative. We predicted the DNA structure of the 250 bp region around the *Arabidopsis* CArG-box sequences using X3DNA (Lu and Olson, 2008), using as starting parameter the dinucleotide properties from the DiProDb database, next we calculate the end-to-end distance of the predicted DNA structure as a measure of DNA curvature for the set of functional and non-functional CArG-boxes separately. **Figure 8** shows the distribution of the end-to-end distances of CArG-box regions. Functional CArG-boxes regions show an average shorter distance which suggests higher level of curvature, although this difference is just marginally significant (p-value=0.067). This poor result could be explained by the difficulty to predict the DNA structure. However, when we used the periodically distribution of A-tract elements on functional CArG-box regions against randomly choose regions, the difference is significant and it can be used as a predictor of 'functional' CArG-box sequences (**Figure 4B**).

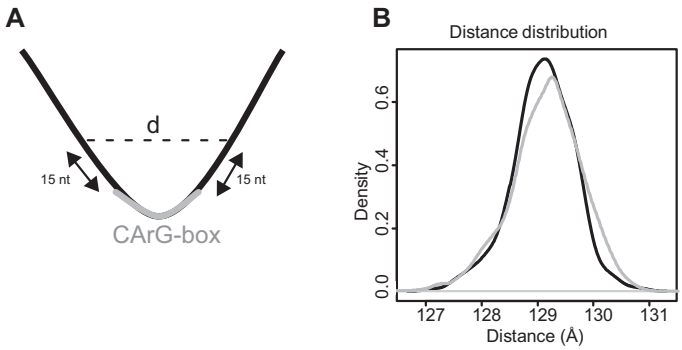


Figure 8. End-to end distance distribution of CArG-box regions. (A) The DNA structure of 50 bp around the CArG-box location was predicted using X3DNA. In order to obtain a value for the DNA curvature, the distance between base pairs positioned 15 nt from both ends of the CArG-box motif was calculated. Shorter distances indicate a higher level of curvature. **(B)** The distribution of end-to-end distances for bound regions by SEP3 in the ChIP-seq experiment (black) and not bound regions (grey). The mean difference of these two sets of regions is not significant (p-value=0.067).

Discussion

The 10 bp DNA sequence motif known as the CArG-box represents the DNA-binding consensus of MADS-domain TFs. Previous studies have focused on the characterization of their primary DNA sequence, omitting the importance of structure properties of the DNA. Since the first structural characterization of the DNA-binding

domain of an animal MADS-domain TF in 1995 (Pellegrini et al., 1995), it has been suggested that this family of TFs directly binds the DNA through interactions of their amino acids mainly with the minor groove side of the DNA. This type of recognition usually relies more on structural properties of DNA rather than on the specific bp sequence (Pellegrini et al., 1995). Here, we studied the importance of the DNA structure as a determinant of the DNA recognition and binding specificity of MADS-domain TFs. By combining bioinformatics and experimental approaches we were able to detect a positive effect of periodically distributed A-tracts, which are associated with particular DNA curvature, on the MADS-domain DNA binding.

We have studied a set of 110 DNA properties as potential factors that can influence the binding of plant MADS-domain TF complexes containing SEP3. Among the most significant proprieties associated with functional CArG-boxes, as determined by ChIP-seq analysis, were properties related with the minor groove and curvature of the DNA. Genomic regions bound by SEP3 complexes were also found to be associated with the presence of periodically distributed A-tract elements. These elements are known to confer a particularly high level of curvature and minor groove properties to the DNA regions where they are located. Interestingly, previous *in vitro* studies have shown that upon MADS-domain TF binding, the DNA is bent with a high degree (e.g. 53° by AP1, 70° by AG) (Riechmann et al., 1996b), and its minor groove width is narrowed. We hypothesize that the affinity of MADS-domain TFs could be related with the energy needed to modify the DNA bending angle, and therefore, DNA-binding affinity will depend on *a priori* properties of the DNA. Indeed, we have shown a positive association of A-tracts inside CArG-box sequences and MADS-domain TF binding. Our bioinformatics analysis also supports this hypothesis, showing that SEP3 bind regions with higher predicted curvature with higher affinity than to other regions. MADS-domain TFs can form quaternary protein complexes that loop the DNA around two CArG-box elements (Egea-Cortines et al., 1999; Honma and Goto, 2001; Melzer and Theissen, 2009; Melzer et al., 2009; Smaczniak et al., 2012b). We hypothesize that the periodicity of A-tracts in regions as long as 300-600 bp could be associated with the need of looping the DNA by higher order complexes *in vivo*. However, this hypothesis remains to be experimentally validated.

Because several MADS-domain protein dimers are able to bend the DNA at different degrees, this structural property can play a role in determining the DNA-binding specificity of different dimers. Our EMSA experiment supports this idea, since the relative gel mobility of the oligonucleotide bound by different dimers is slightly different. Leveraging the temperature-dependent curvature of the A-tract elements confirmed this hypothesis. Changing the temperature will not modify the primary DNA sequence, but it is known that it will affect the curvature of the DNA

containing an A-tract (Koo et al., 1986; Prosseda et al., 2010). We observed that the relative *in vitro* affinity of the SEP3 and SEP3-AG dimers change with the temperature, supporting the influence of the DNA curvature in the *in vitro* specificity of these two dimers. Additionally, we found that DNA regions bound by different SEP3 dimers *in vivo* show an overrepresentation of A-tracts of different length which supports the hypothesis that the DNA curvature-dependent specificity of MADS-domain TFs may be also important *in vivo*. The fact that the length of A-tracts is conserved among the *Arabidopsis* ecotypes for regions bound by MADS-domain TFs also indicates the importance of this structural property.

The hypothesis that MADS-domain TFs recognize a special DNA structure that can be modified by external factors (e.g. temperature) opens new possibilities for its mechanism of DNA-binding specificity and therefore gene regulation. Several MADS-domain TFs act in processes that are temperature-dependent, such as floral transition, flower maturation and fruit ripening (reviewed by Smaczniak et al., 2012a). A similar mechanism of temperature-sensing has been observed in bacteria, where temperature-dependent changes in DNA curvature that are associated to promoter regions containing A-tract elements play an important role in temperature-controlled gene expression (Prosseda et al., 2010). Future studies need to reveal the biological importance of this temperature-dependent regulation *in vivo*.

Funding

This work was supported by funding from the NWO-NGI Horizon Breakthrough project 93519015 to AJvD, by the NWO-VIDI project to KK, and by the NGI-Netherlands Proteomics Centre (NPC) project to GCA and CS.

Chapter 7

Concluding remarks

Cezary Smaczniak, Gerco C. Angenent, Kerstin Kaufmann

MADS-domain proteins form one of the largest transcription factor family in plants and their function is essential for almost every developmental process. To date, more than 100 MADS-box genes have been identified in *Arabidopsis thaliana* (Parenicová et al., 2003) and for nearly half of them the biological function have been ascribed. Interestingly, many MADS-box transcription factors have more than one function in apparently different developmental processes (Smaczniak et al., 2012a). The emerging functional characterization of MADS-box transcription factors in many plant species provides useful information on the origin and diversification of plant morphologies and life history traits. By systematic screenings of MADS-box gene mutants, novel roles of these important family members are being uncovered. Until recently, only type II MADS-domain proteins have been considered as major players in the developmental processes of *Arabidopsis*. However, nowadays type I MADS-box genes are drawing more attention and their functional characterization shows their important roles in various plant developmental processes, in particular embryo and female gametophyte development (Bemer et al., 2010; Masiero et al., 2011). How MADS-box genes acquired their functional specificity still remained an unresolved question.

It is a general concept that the functional specificity of a transcription factor is determined by the set of target genes that it regulates. Therefore, by unraveling the 'molecular code' of DNA-binding site recognition of MADS-domain proteins and their complexes can help to understand how these factors acquire their functional specificity. *In vivo* genome-wide DNA-binding studies of the MADS-domain proteins SEPALLATA3, FLOWERING LOCUS C, APETALA1, AGAMOUS-LIKE15, SHORT VEGETATIVE PHASE, and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (Kaufmann et al., 2009; Zheng et al., 2009; Kaufmann et al., 2010c; Deng et al., 2011; Immink et al., 2012; Tao et al., 2012), revealed a high number of their potential target genes. Some of these target genes are common for multiple MADS-domain proteins (e.g. for SEPALLATA3 and APETALA1), which then also show an overlap in DNA binding sites. One of the reasons for the common binding sites is the formation of protein complexes. The well-known 'floral quartet' model suggests such a possibility (Theissen and Saedler, 2001). MADS-domain protein-protein interaction studies in heterologous systems unraveled a complex protein interaction network within the MADS-domain transcription factor family (de Folter et al., 2005). Moreover, higher-order MADS-domain protein complex formation was reported in numerous *in vitro* as well as in yeast-based studies (Egea-Cortines et al., 1999; Immink et al., 2009; Melzer and Theissen, 2009; Melzer et al., 2009; Smaczniak et al., 2012b). However, until recently, tools for studying the presence and composition of *in planta* MADS-domain protein complexes were lacking. Therefore, we optimized and applied the protein immunoprecipitation technique combined with

mass spectrometry to isolate and characterize MADS-domain protein complexes from native plant tissues (Chapters 2 and 4). In our method we made use of quantification of differential protein abundances between immunoprecipitated samples and controls, to distinguish specifically interacting proteins from background for the selected transcription factors. Using this optimized approach we were able, for the first time, to confirm the interactions in the 'core' protein complexes on which the 'floral quartet' model is based (Theissen and Saedler, 2001). Additionally, we were able to establish a MADS-domain protein interactome that supports a mechanistic link between MADS-domain proteins, chromatin remodeling factors and other transcriptional coregulators (Smaczniak et al., 2012b).

The interactions between MADS-domain proteins and other transcription factors, such as homeodomain proteins, may recruit other proteins to target gene promoters. This could in turn explain overlapping *in vivo* DNA-binding patterns and the formation of large molecular machineries that regulate target gene expression. The presence of SEUSS and LEUNIG HOMOLOG, the homolog of LEUNIG (all transcriptional coregulators) (Sitaraman et al., 2008), in the APETALA1 immunoprecipitate suggests an active role of these protein complexes in the regulation of *AGAMOUS* and maybe other targets. Recent chromatin accessibility studies by DNase I-seq revealed a secondary footprint in very close proximity to the well characterized MADS-domain binding footprints of SEPALLATA3, suggesting that SEPALLATA3 target genes are possibly regulated or coregulated also by other transcription factors (Zhang et al., 2012). Moreover, the identification of chromatin remodeling factors as interaction partners of MADS-domain proteins suggest more active regulation of the chromatin structure upon MADS-domain protein DNA-binding. This idea is especially supported by the interaction of MADS-domain proteins with the SWItch2/Sucrose NonFermentable (SWI2/SNF2) chromatin-remodeling ATPases SPLAYED and BRAHMA (Bezhani et al., 2007), which have been shown to be redundantly required for floral patterning and for the activation of *APETALA3* and *AGAMOUS* (Wu et al., 2012). The SPLAYED and BRAHMA factors have been independently characterized as interaction partners of MADS-domain protein SEPALLATA3 (Wu et al., 2012), which further supports our suggestions on chromatin state regulation by MADS-domain transcription factors. It is tempting to speculate that the expression of the target genes is regulated by MADS-domain proteins through changes of the chromatin context after transcription factor binding to the gene regulatory elements and recruitment of the chromatin remodeling machinery (and not the other way around). Analysis of the chromatin in a more dynamic manner would advance our understanding on how chromatin re-organization is linked to the MADS-domain protein activity.

Technological progress in the proteomics field, such as development of new generation, high-resolution mass spectrometers, allowed us to perform targeted MADS-domain protein interaction studies in native plant tissues and detect very low abundant MADS-domain protein complexes. Advanced proteomic approaches could be used to systematically study transcriptional regulation for example by detecting post-translational modifications or characterizing transcription factor protein complexes (transcriptional machineries). Improved biochemical procedures for tissue prefractionation allow separating different subcompartments of plant cells, which ultimately can be used to study entire protein signaling cascades throughout the plant cell (Kaufmann et al., 2011). In the future, the development of protein quantification methods, both relative and absolute, in native plant tissues should allow not only for protein complex partner characterization but also for deciphering the exact stoichiometry of protein complexes.

The ability of MADS-domain proteins to form different higher-order protein complexes that specify different types of floral organs, suggest that by modulating the complex composition these factors acquire different DNA-binding specificities. By performing *in vitro* protein-DNA binding site enrichment studies called SELEX (for Systematic Evolution of Ligands by Exponential Enrichment) and applying high-throughput DNA sequencing of the evolved DNA sequence libraries, we were able to detect minor but apparently significant differences in the DNA-binding specificities of several MADS-domain dimers, which are supposed to act by regulating different subsets of target genes (Chapter 5). Genome-wide mapping of the DNA sequences identified by SELEX-seq and compared with ChIP-seq data suggested that MADS-domain complexes can bind partly different sites throughout the *Arabidopsis* genome, therefore providing clues to their target gene specificity. Moreover, our results showed that each DNA binding motif can have different affinity towards a specific protein complex and that the DNA-binding affinity levels can distinguish some of the MADS-domain protein complexes from each other. This introduces another level in the transcriptional regulation, where different protein complexes could compete for the same binding sites in the regulatory regions of their target genes. What is the exact source of different DNA-binding characteristics of MADS-domain protein complexes *in vivo* still needs to be elucidated, but the results shown in this thesis suggest an important role for the transcription factor dimer composition in DNA binding specificity. In addition, cofactors or other transcription factors present in the complexes could influence the binding characteristics. In chapter 6 we have shown that not only the primary DNA sequence, but also the DNA structure may play a role in the DNA bending capacity and thereby binding capacity. How the different MADS domain dimers recognizes slightly different sites needs further studies. For this, protein

crystal structure of plant MADS domain dimers bound to DNA, as has been determined by SRF, would be extremely helpful.

The majority of studies on DNA-binding and protein interactions of transcription factors are static and therefore provide only a generalized overview on the regulatory networks. The ultimate goal in the characterization of protein and gene interaction networks, especially in the developmental biology field, should be the elucidation of the dynamic changes that occur over time: at different developmental stages and in different tissues. By introducing time-series experiments performed at different stages of plant development and by *in vivo* live imaging we may get a better insight into the dynamics of transcriptional regulation at different biological time-scales. Additionally, by combining high-throughput sequencing methods such as ChIP-seq (for genome-wide transcription factors binding characterization), DNase I-seq and MNase-seq (for DNA accessibility - nucleosome occupancy studies), RNA-seq (for gene expression and splicing variants description) we may get the full picture of the dynamics of transcription factor regulation during flower development. Alternatively, novel or improved computational modeling tools can also be used to predict gene and protein regulatory networks (e.g. by unraveling active protein DNA-binding domains or genome-wide transcription factor binding sites) (van Dijk et al., 2010; Ding et al., 2012).

References

Adamczyk, B.J., Lehti-Shiu, M.D., Fernandez, D.E. (2007). The MADS domain factors AGL15 and AGL18 act redundantly as repressors of the floral transition in Arabidopsis. Plant J. **50,** 1007-1019.

Adamczyk, B.J., Fernandez, D.E. (2009). MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in *Arabidopsis*. Plant Physiol. **149,** 1713-1723.

Agrawal, G.K., Jwa, N.S., Lebrun, M.H., Job, D., Rakwal, R. (2010). Plant secretome: unlocking secrets of the secreted proteins. Proteomics **10,** 799-827.

Aichinger, E., Villar, C.B., Farrona, S., Reyes, J.C., Hennig, L., Kohler, C. (2009). CHD3 proteins and polycomb group proteins antagonistically determine cell identity in *Arabidopsis*. PLoS Genet. **5,** e1000605.

Airoldi, C.A., Bergonzi, S., Davies, B. (2010). Single amino acid change alters the ability to specify male or female organ identity. Proc. Natl. Acad. Sci. USA **107,** 18898-18902.

Aki, T., Shigyo, M., Nakano, R., Yoneyama, T., Yanagisawa, S. (2008). Nano scale proteomics revealed the presence of regulatory proteins including three FT-Like proteins in phloem and xylem saps from rice. Plant Cell Physiol. **49,** 767-790.

Alam, I., Lee, D.G., Kim, K.H., Park, C.H., Sharmin, S.A., Lee, H., Oh, K.W., Yun, B.W., Lee, B.H. (2010). Proteome analysis of soybean roots under waterlogging stress at an early vegetative stage. Journal of Biosciences **35,** 49-62.

Alexandersson, E., Saalbach, G., Larsson, C., Kjellbom, P. (2004). *Arabidopsis* plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. Plant Cell Physiol. **45,** 1543-1556.

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., de Pouplana, L.R., Martinez-Castilla, L., Yanofsky, M.F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. USA **97,** 5328-5333.

Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M., Matsubayashi, Y. (2007). Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **104,** 18333-18338.

America, A.H., Cordewener, J.H. (2008). Comparative LC-MS: a landscape of peaks and valleys. Proteomics **8,** 731-749.

Amey, R., Schleicher, T., Macdonald, H., Neill, S., Spencer-Phillips, P. (2005). Global analysis of stress responses: the pea proteome following biotic and abiotic stresses. Comp. Biochem. Physiol. Part A Mol. Integr. Physiol. **141,** S255-S255.

Ampomah-Dwamena, C., Morris, B.A., Sutherland, P., Veit, B., Yao, J.L. (2002). Down-regulation of *TM29*, a tomato *SEPALLATA* homolog, causes parthenocarpic fruit development and floral reversion. Plant Physiol. **130,** 605-617.

Arora, R., Agarwal, P., Ray, S., Singh, A.K., Singh, V.P., Tyagi, A.K., Kapoor, S. (2007). MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. BMC Genomics **8,** 242.

Backstrom, S., Elfving, N., Nilsson, R., Wingsle, G., Bjorklund, S. (2007). Purification of a plant mediator from *Arabidopsis thaliana* identifies PFT1 as the Med25 subunit. Mol. Cell **26,** 717-729.

Baerenfaller, K., Grossmann, J., Grobei, M.A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., Zimmermann, P., Grossniklaus, U., Gruissem, W., Baginsky, S. (2008). Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. Science **320,** 938-941.

Baginsky, S., Hennig, L., Zimmermann, P., Gruissem, W. (2010). Gene expression analysis, proteomics, and network discovery. Plant Physiol. **152,** 402-410.

Bao, X., Franks, R.G., Levin, J.Z., Liu, Z. (2004). Repression of *AGAMOUS* by *BELLRINGER* in floral and inflorescence meristems. Plant Cell **16,** 1478-1489.

Barnes, A., Bale, J., Constantinidou, C., Ashton, P., Jones, A., Pritchard, J. (2004). Determining protein identity from sieve element sap in *Ricinus communis* L. by quadrupole time of flight (Q-TOF) mass spectrometry. J. Exp. Bot. **55,** 1473-1481.

Bateman, R.M., Hilton, J., Rudall, P.J. (2006). Morphological and molecular phylogenetic context of the angiosperms: contrasting the 'top-down' and 'bottom-up' approaches used to infer the likely characteristics of the first flowers. J. Exp. Bot. **57,** 3471-3503.

Bayer, E.M., Bottrill, A.R., Walshaw, J., Vigouroux, M., Naldrett, M.J., Thomas, C.L., Maule, A.J. (2006). *Arabidopsis* cell wall proteome defined using multidimensional protein identification technology. Proteomics **6,** 301-311.

Becker, A., Kaufmann, K., Freialdenhoven, A., Vincent, C., Li, M.A., Saedler, H., Theissen, G. (2002). A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. Mol. Genet. Genomics **266,** 942-950.

Becker, A., Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol. Phylogenet. Evol. **29,** 464-489.

Becker, A., Lange, M. (2010). VIGS - genomics goes functional. Trends Plant Sci. **15,** 1-4.

Bemer, M., Wolters-Arts, M., Grossniklaus, U., Angenent, G.C. (2008). The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in *Arabidopsis* ovules. Plant Cell **20,** 2088-2101.

Bemer, M., Heijmans, K., Airoldi, C., Davies, B., Angenent, G.C. (2010). An atlas of type I MADS box gene expression during female gametophyte and seed development in *Arabidopsis*. Plant Physiol. **154,** 287-300.

Benlloch, R., Roque, E., Ferrandiz, C., Cosson, V., Caballero, T., Penmetsa, R.V., Beltran, J.P., Canas, L.A., Ratet, P., Madueno, F. (2009). Analysis of B function in legumes: PISTILLATA proteins do not require the PI motif for floral organ development in *Medicago truncatula*. Plant J. **60,** 102-111.

Benschop, J.J., Mohammed, S., O'Flaherty, M., Heck, A.J., Slijper, M., Menke, F.L. (2007). Quantitative phosphoproteomics of early elicitor signaling in *Arabidopsis*. Mol. Cell. Proteomics **6,** 1198-1214.

Berbel, A., Ferrandiz, C., Hecht, V., Dalmais, M., Lund, O.S., Sussmilch, F.C., Taylor, S.A., Bendahmane, A., Ellis, T.H., Beltran, J.P., Weller, J.L., Madueno, F. (2012). *VEGETATIVE1* is essential for development of the compound inflorescence in pea. Nat. Commun. **3,** 797.

Berggard, T., Linse, S., James, P. (2007). Methods for the detection and analysis of protein-protein interactions. Proteomics **7,** 2833-2842.

Bezhani, S., Winter, C., Hershman, S., Wagner, J.D., Kennedy, J.F., Kwon, C.S., Pfluger, J., Su, Y., Wagner, D. (2007). Unique, shared, and redundant roles for the *Arabidopsis* SWI/SNF chromatin remodeling ATPases BRAHMA and SPLAYED. Plant Cell **19,** 403-416.

Bindschedler, L.V., Palmblad, M., Cramer, R. (2008). Hydroponic isotope labelling of entire plants (HILEP) for quantitative plant proteomics; an oxidative stress case study. Phytochemistry **69,** 1962-1972.

Birkenbihl, R.P., Jach, G., Saedler, H., Huijser, P. (2005). Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. J. Mol. Biol. **352,** 585-596.

Blagoev, B., Kratchmarova, I., Ong, S.E., Nielsen, M., Foster, L.J., Mann, M. (2003). A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. Nat. Biotechnol. **21,** 315-318.

Blazquez, M. (2000). Flower development pathways. J. Cell Sci. **113,** 3547-3548.

Bonardi, V., Pesaresi, P., Becker, T., Schleiff, E., Wagner, R., Pfannschmidt, T., Jahns, P., Leister, D. (2005). Photosystem II core phosphorylation and photosynthetic acclimation require two different protein kinases. Nature **437,** 1179-1182.

Bradley, D., Carpenter, R., Sommer, H., Hartley, N., Coen, E. (1993). Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of Antirrhinum. Cell **72,** 85-95.

Brambilla, V., Battaglia, R., Colombo, M., Masiero, S., Bencivenga, S., Kater, M.M., Colombo, L. (2007). Genetic and molecular interactions between BELL1 and MADS box factors support ovule development in *Arabidopsis*. Plant Cell **19,** 2544-2556.

Browse, J. (2009). Jasmonate passes muster: a receptor and targets for the defense hormone. Annu. Rev. Plant Biol. **60,** 183-205.

Burgeff, C., Liljegren, S.J., Tapia-Lopez, R., Yanofsky, M.F., Alvarez-Buylla, E.R. (2002). MADS-box gene expression in lateral primordia, meristems and differentiated tissues of *Arabidopsis thaliana* roots. Planta **214,** 365-372.

Buzgo, M., Soltis, P.S., Soltis, D.E. (2004). Floral developmental morphology of *Amborella trichopoda* (Amborellaceae). Int. J. Plant Sci. **165,** 925-947.

Calonje, M., Cubas, P., Martinez-Zapater, J.M., Carmona, M.J. (2004). Floral meristem identity genes are expressed during tendril development in grapevine. Plant Physiol. **135,** 1491-1501.

Calonje, M., Sung, Z. (2006). Complexity beneath the silence. Curr. Opin. Plant Biol. **9,** 530-537.

Carter, C., Pan, S.Q., Jan, Z.H., Avila, E.L., Girke, T., Raikhel, N.V. (2004). The vegetative vacuole proteorne of *Arabidopsis thaliana* reveals predicted and unexpected proteins. Plant Cell **16,** 3285-3303.

Cartolano, M., Castillo, R., Efremova, N., Kuckenberg, M., Zethof, J., Gerats, T., Schwarz-Sommer, Z., Vandenbussche, M. (2007). A conserved microRNA module exerts homeotic control over *Petunia hybrida* and *Antirrhinum majus* floral organ identity. Nat. Genet. **39,** 901-905.

Castellana, N.E., Payne, S.H., Shen, Z.X., Stanke, M., Bafna, V., Briggs, S.P. (2008). Discovery and revision of *Arabidopsis* genes by proteogenomics. Proc. Natl. Acad. Sci. USA **105,** 21034-21038.

Catusse, J., Strub, J.M., Job, C., Van Dorsselaer, A., Job, D. (2008). Proteome-wide characterization of sugarbeet seed vigor and its tissue specific expression. Proc. Natl. Acad. Sci. USA **105,** 10262-10267.

Causier, B., Castillo, R., Zhou, J., Ingram, R., Xue, Y., Schwarz-Sommer, Z., Davies, B. (2005). Evolution in action: following function in duplicated floral homeotic genes. Curr. Biol. **15,** 1508-1512.

Causier, B., Bradley, D., Cook, H., Davies, B. (2009). Conserved intragenic elements were critical for the evolution of the floral C-function. Plant J. **58,** 41-52.

Causier, B., Schwarz-Sommer, Z., Davies, B. (2010). Floral organ identity: 20 years of ABCs. Semin. Cell Dev. Biol. **21,** 73-79.

Cegan, R., Marais, G.A., Kubekova, H., Blavet, N., Widmer, A., Vyskot, B., Dolezel, J., Safar, J., Hobza, R. (2010). Structure and evolution of *Apetala3*, a sex-linked gene in *Silene latifolia*. BMC Plant Biol. **10,** 180.

Cevik, V., Ryder, C.D., Popovich, A., Manning, K., King, G.J., Seymour, G.B. (2010). A *FRUITFULL-like* gene is associated with genetic variation for fruit flesh firmness in apple (*Malus domestica* Borkh.). Tree Genet. Genomes **6,** 271-279.

Chang, I.F. (2006). Mass spectrometry-based proteomic analysis of the epitope-tag affinity purified protein complexes in eukaryotes. Proteomics **6,** 6158-6166.

Chang, I.F., Curran, A., Woolsey, R., Quilici, D., Cushman, J.C., Mittler, R., Harmon, A., Harper, J.F. (2009). Proteomic profiling of tandem affinity purified 14-3-3 protein complexes in *Arabidopsis thaliana*. Proteomics **9,** 2967-2985.

Chang, Y.Y., Kao, N.H., Li, J.Y., Hsu, W.H., Liang, Y.L., Wu, J.W., Yang, C.H. (2010). Characterization of the possible roles for B class MADS box genes in regulation of perianth formation in orchid. Plant Physiol. **152,** 837-853.

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.H., Sung, Z.R., Goodrich, J. (2004). Interaction of polycomb-group proteins controlling flowering in *Arabidopsis*. Development **131,** 5263-5276.

Chen, M.K., Hsu, W.H., Lee, P.F., Thiruvengadam, M., Chen, H.I., Yang, C.H. (2011). The MADS box gene, *FOREVER YOUNG FLOWER*, acts as a repressor controlling floral organ senescence and abscission in *Arabidopsis*. Plant J. **68,** 168-185.

Chen, Y., Hoehenwarter, W., Weckwerth, W. (2010). Comparative analysis of phytohormone-responsive phosphoproteins in *Arabidopsis thaliana* using TiO_2-phosphopeptide enrichment and mass accuracy precursor alignment. Plant J. **63,** 1-17.

Chiang, G.C., Barua, D., Kramer, E.M., Amasino, R.M., Donohue, K. (2009). Major flowering time gene, *FLOWERING LOCUS C*, regulates seed germination in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **106,** 11661-11666.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nurnberger, T., Jones, J.D., Felix, G., Boller, T. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature **448,** 497-500.

Chitteti, B.R., Peng, Z.H. (2007). Proteome and phosphoproteome differential expression under salinity stress in rice (*Oryza sativa*) roots. J. Proteome Res. **6,** 1718-1727.

Cho, W.K., Chen, X.Y., Chu, H., Rim, Y., Kim, S., Kim, S.T., Kim, S.W., Park, Z.Y., Kim, J.Y. (2009). Proteomic analysis of the secretome of rice calli. Physiol. Plant. **135,** 331-341.

Ciaffi, M., Paolacci, A.R., Tanzarella, O.A., Porceddu, E. (2011). Molecular aspects of flower development in grasses. Sex. Plant Reprod. **24,** 247-282.

Ciannamea, S., Kaufmann, K., Frau, M., Tonaco, I.A., Petersen, K., Nielsen, K.K., Angenent, G.C., Immink, R.G. (2006). Protein interactions of MADS box transcription factors involved in flowering in *Lolium perenne*. J. Exp. Bot. **57,** 3419-3431.

Citri, A., Yarden, Y. (2006). EGF-ERBB signalling: towards the systems level. Nat. Rev. Mol. Cell Biol. **7,** 505-516.

Clarke, J.H., Dean, C. (1994). Mapping FRI, a locus controlling flowering time and vernalization response in Arabidopsis thaliana. Mol. Gen. Genet. **242,** 81-89.

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

Cock, J.M., McCormick, S. (2001). A large family of genes that share homology with CLAVATA3. Plant Physiol. **126,** 939-942.

Coen, E.S., Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. Nature **353,** 31-37.

Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H.J., Angenent, G.C., van Tunen, A.J. (1995). The petunia MADS box gene *FBP11* determines ovule identity. Plant Cell **7,** 1859-1868.

Colombo, M., Masiero, S., Vanzulli, S., Lardelli, P., Kater, M.M., Colombo, L. (2008). *AGL23*, a type I MADS-box gene that controls female gametophyte and embryo development in *Arabidopsis*. Plant J. **54,** 1037-1048.

Colombo, M., Brambilla, V., Marcheselli, R., Caporali, E., Kater, M.M., Colombo, L. (2010). A new role for the *SHATTERPROOF* genes during *Arabidopsis* gynoecium development. Dev. Biol. **337,** 294-302.

Conner, J., Liu, Z. (2000). *LEUNIG*, a putative transcriptional corepressor that regulates *AGAMOUS* expression during flower development. Proc. Natl. Acad. Sci. USA **97,** 12902-12907.

Corbesier, L., Vincent, C., Jang, S.H., Fornara, F., Fan, Q.Z., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., Coupland, G. (2007). FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. Science **316,** 1030-1033.

Cox, J., Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. **26,** 1367-1372.

Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J.V., Mann, M. (2009). A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nat. Protoc. **4,** 698-705.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. **10,** 1794-1805.

Cui, R., Han, J., Zhao, S., Su, K., Wu, F., Du, X., Xu, Q., Chong, K., Theissen, G., Meng, Z. (2010). Functional conservation and diversification of class E floral homeotic genes in rice (*Oryza sativa*). Plant J. **61,** 767-781.

Dam, S., Laursen, B.S., Ornfelt, J.H., Jochimsen, B., Staerfeldt, H.H., Friis, C., Nielsen, K., Goffard, N., Besenbacher, S., Krusell, L., Sato, S., Tabata, S., Thogersen, I.B., Enghild, J.J., Stougaard, J. (2009). The proteome of seed development in the model legume *Lotus japonicus*. Plant Physiol. **149,** 1325-1340.

Davies, B., Egea-Cortines, M., de Andrade Silva, E., Saedler, H., Sommer, H. (1996). Multiple interactions amongst floral homeotic MADS box proteins. EMBO J. **15,** 4330-4343.

Davies, B., Motte, P., Keck, E., Saedler, H., Sommer, H., Schwarz-Sommer, Z. (1999). *PLENA* and *FARINELLI*: redundancy and regulatory interactions between two *Antirrhinum* MADS-box factors controlling flower development. EMBO J. **18,** 4023-4034.

De Bodt, S., Raes, J., Florquin, K., Rombauts, S., Rouze, P., Theissen, G., Van de Peer, Y. (2003). Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J. Mol. Evol. **56,** 573-586.

De Bodt, S., Maere, S., Van de Peer, Y. (2005). Genome duplication and the origin of angiosperms. Trends Ecol. Evol. **20,** 591-597.

de Folter, S., Immink, R.G., Kieffer, M., Parenicova, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M., Davies, B., Angenent, G.C. (2005). Comprehensive interaction map of the Arabidopsis MADS box transcription factors. Plant Cell **17,** 1424-1433.

de Folter, S., Angenent, G.C. (2006). trans meets cis in MADS science. Trends Plant Sci. **11,** 224-231.

de Folter, S., Shchennikova, A.V., Franken, J., Busscher, M., Baskar, R., Grossniklaus, U., Angenent, G.C., Immink, R.G. (2006). A B_{sister} MADS-box gene involved in ovule and seed development in petunia and Arabidopsis. Plant J. **47,** 934-946.

de Folter, S., Urbanus, S.L., van Zuijlen, L.G., Kaufmann, K., Angenent, G.C. (2007). Tagging of MADS domain proteins for chromatin immunoprecipitation. BMC Plant Biol. **7,** 47.

de la Fuente van Bentem, S., Hirt, H. (2007). Using phosphoproteomics to reveal signalling dynamics in plants. Trends Plant Sci. **12,** 404-411.

De Lucia, F., Crevillen, P., Jones, A.M.E., Greb, T., Dean, C. (2008). A PHD-Polycomb Repressive Complex 2 triggers the epigenetic silencing of *FLC* during vernalization. Proc. Natl. Acad. Sci. USA **105,** 16831-16836.

Deal, R.B., Henikoff, S. (2011). The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat. Protoc. **6,** 56-68.

Deng, W., Ying, H., Helliwell, C.A., Taylor, J.M., Peacock, W.J., Dennis, E.S. (2011). FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of *Arabidopsis*. Proc. Natl. Acad. Sci. USA **108,** 6680-6685.

Deng, Z., Zhang, X., Tang, W., Oses-Prieto, J.A., Suzuki, N., Gendron, J.M., Chen, H., Guan, S., Chalkley, R.J., Peterman, T.K., Burlingame, A.L., Wang, Z.Y. (2007). A proteomics study of brassinosteroid response in *Arabidopsis*. Mol. Cell. Proteomics **6,** 2058-2071.

Di Stilio, V.S., Kramer, E.M., Baum, D.A. (2005). Floral MADS box genes and homeotic gender dimorphism in *Thalictrum dioicum* (Ranunculaceae) - a new model for the study of dioecy. Plant J. **41,** 755-766.

Diaz-Riquelme, J., Lijavetzky, D., Martinez-Zapater, J.M., Carmona, M.J. (2009). Genome-wide analysis of MIKC^{C}-type MADS box genes in grapevine. Plant Physiol. **149,** 354-369.

Diggle, P.K., Di Stilio, V.S., Gschwend, A.R., Golenberg, E.M., Moore, R.C., Russell, J.R., Sinclair, J.P. (2011). Multiple developmental processes underlie sex differentiation in angiosperms. Trends Genet. **27,** 368-376.

Ding, J., Hu, H., Li, X. (2012). Thousands of cis-regulatory sequence combinations are shared by Arabidopsis and poplar. Plant Physiol. **158,** 145-155.

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., Yanofsky, M.F. (2004). The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. Curr. Biol. **14,** 1935-1940.

Dorca-Fornell, C., Gregis, V., Grandi, V., Coupland, G., Colombo, L., Kater, M.M. (2011). The Arabidopsis *SOC1*-like genes *AGL42*, *AGL71* and *AGL72* promote flowering in the shoot apical and axillary meristems. Plant J. **67,** 1006-1017.

Dornelas, M.C., Patreze, C.M., Angenent, G.C., Immink, R.G. (2011). MADS: the missing link between identity and growth? Trends Plant Sci. **16,** 89-97.

Drea, S., Hileman, L.C., de Martino, G., Irish, V.F. (2007). Functional analyses of genetic pathways controlling petal specification in poppy. Development **134,** 4157-4166.

Dreni, L., Pilatone, A., Yun, D., Erreni, S., Pajoro, A., Caporali, E., Zhang, D., Kater, M.M. (2011). Functional analysis of all AGAMOUS subfamily members in rice reveals their roles in reproductive organ identity determination and meristem determinacy. Plant Cell **23,** 2850-2863.

Du, X., Xiao, Q., Zhao, R., Wu, F., Xu, Q., Chong, K., Meng, Z. (2008). *TrMADS3*, a new MADS-box gene, from a perennial species *Taihangia rupestris* (Rosaceae) is upregulated by cold and experiences seasonal fluctuation in expression level. Dev. Genes Evol. **218,** 281-292.

Dunham, W.H., Mullin, M., Gingras, A.C. (2012). Affinity-purification coupled to mass spectrometry: Basic principles and strategies. Proteomics **12,** 1576-1590.

Dunkley, T.P., Hester, S., Shadforth, I.P., Runions, J., Weimar, T., Hanton, S.L., Griffin, J.L., Bessant, C., Brandizzi, F., Hawes, C., Watson, R.B., Dupree, P., Lilley, K.S. (2006). Mapping the *Arabidopsis* organelle proteome. Proc. Natl. Acad. Sci. USA **103,** 6518-6523.

Dunkley, T.P.J., Watson, R., Griffin, J.L., Dupree, P., Lilley, K.S. (2004). Localization of organelle proteins by isotope tagging (LOPIT). Mol. Cell. Proteomics **3,** 1128-1134.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32,** 1792-1797.

Egea-Cortines, M., Saedler, H., Sommer, H. (1999). Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. EMBO J. **18,** 5370-5379.

El-Khatib, R.T., Good, A.G., Muench, D.G. (2007). Analysis of the *Arabidopsis* cell suspension phosphoproteome in response to short-term low temperature and abscisic acid treatment. Physiol. Plant. **129,** 687-697.

Elo, A., Lemmetyinen, J., Novak, A., Keinonen, K., Porali, I., Hassinen, M., Sopanen, T. (2007). *BpMADS4* has a central role in inflorescence initiation in silver birch (*Betula pendula*). Physiol. Plant. **131,** 149-158.

Erhardt, M., Adamska, I., Franco, O.L. (2010). Plant nuclear proteomics - inside the cell maestro. FEBS J. **277,** 3295-3307.

Eshed, Y., Baum, S.F., Bowman, J.L. (1999). Distinct mechanisms promote polarity establishment in carpels of *Arabidopsis*. Cell **99,** 199-209.

Espinosa-Soto, C., Padilla-Longoria, P., Alvarez-Buylla, E.R. (2004). A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. Plant Cell **16,** 2923-2939.

Farrona, S., Hurtado, L., Bowman, J.L., Reyes, J.C. (2004). The *Arabidopsis thaliana* SNF2 homolog AtBRM controls shoot development and flowering. Development **131,** 4965-4975.

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M.F., Kater, M.M., Colombo, L. (2003). MADS-box protein complexes control carpel and ovule development in Arabidopsis. Plant Cell **15,** 2603-2611.

Fernandez, D.E., Heck, G.R., Perry, S.E., Patterson, S.E., Bleecker, A.B., Fang, S.C. (2000). The embryo MADS domain factor AGL15 acts postembryonically: Inhibition of perianth senescence and abscission via constitutive expression. Plant Cell **12,** 183-198.

Ferrandiz, C., Gu, Q., Martienssen, R., Yanofsky, M.F. (2000a). Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. Development **127,** 725-734.

Ferrandiz, C., Liljegren, S.J., Yanofsky, M.F. (2000b). Negative regulation of the *SHATTERPROOF* genes by FRUITFULL during *Arabidopsis* fruit development. Science **289,** 436-438.

Franks, R.G., Wang, C., Levin, J.Z., Liu, Z. (2002). *SEUSS*, a member of a novel family of plant regulatory proteins, represses floral homeotic gene expression with *LEUNIG*. Development **129,** 253-263.

Friedel, M., Nikolajewa, S., Suhnel, J., Wilhelm, T. (2009). DiProGB: the dinucleotide properties genome browser. Bioinformatics **25,** 2603-2604.

Fujisawa, M., Nakano, T., Ito, Y. (2011). Identification of potential target genes for the tomato fruit-ripening regulator RIN by chromatin immunoprecipitation. BMC Plant Biol. **11,** 26.

Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R.D., Burstin, J. (2003). Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol. **133,** 664-682.

Gan, Y., Filleur, S., Rahman, A., Gotensparre, S., Forde, B.G. (2005). Nutritional regulation of *ANR1* and other root-expressed MADS-box genes in *Arabidopsis thaliana*. Planta **222,** 730-742.

Gao, X., Liang, W., Yin, C., Ji, S., Wang, H., Su, X., Guo, C., Kong, H., Xue, H., Zhang, D. (2010). The *SEPALLATA*-like gene *OsMADS34* is required for rice inflorescence and spikelet development. Plant Physiol. **153,** 728-740.

Gerats, T., Vandenbussche, M. (2005). A model system for comparative research: *Petunia*. Trends Plant Sci. **10,** 251-256.

Geuten, K., Irish, V. (2010). Hidden variability of floral homeotic B genes in Solanaceae provides a molecular basis for the evolution of novel functions. Plant Cell **22,** 2562-2578.

Giavalisco, P., Kapitza, K., Kolasa, A., Buhtz, A., Kehr, J. (2006). Towards the proteome of *Brassica napus* phloem sap. Proteomics **6,** 896-909.

Gimenez, E., Pineda, B., Capel, J., Anton, M.T., Atares, A., Perez-Martin, F., Garcia-Sogo, B., Angosto, T., Moreno, V., Lozano, R. (2010). Functional analysis of the *Arlequin* mutant corroborates the essential role of the *Arlequin/TAGL1* gene during reproductive development of tomato. PLoS One **5,** e14427.

Gomez, A., Lopez, J.A., Pintos, B., Camafeita, E., Bueno, M.A. (2009). Proteomic analysis from haploid and diploid embryos of *Quercus suber* L. identifies qualitative and quantitative differential expression patterns. Proteomics **9,** 4355-4367.

Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M., Coupland, G. (1997). A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. Nature **386,** 44-51.

Goto, K., Meyerowitz, E.M. (1994). Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. Genes Dev. **8,** 1548-1560.

Gramzow, L., Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. Genome Biol. **11,** 214.

Gregis, V., Sessa, A., Dorca-Fornell, C., Kater, M.M. (2009). The *Arabidopsis* floral meristem identity genes *AP1*, *AGL24* and *SVP* directly repress class B and C floral homeotic genes. Plant J. **60,** 626-637.

Gruhler, A., Schulze, W.X., Matthiesen, R., Mann, M., Jensen, O.N. (2005). Stable isotope labeling of Arabidopsis thaliana cells and quantitative proteomics by mass spectrometry. Mol. Cell. Proteomics **4,** 1697-1709.

Gu, Q., Ferrandiz, C., Yanofsky, M.F., Martienssen, R. (1998). The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. Development **125,** 1509-1517.

Gudesblat, G.E., Schneider-Pizon, J., Betti, C., Mayerhofer, J., Vanhoutte, I., van Dongen, W., Boeren, S., Zhiponova, M., de Vries, S., Jonak, C., Russinova, E. (2012). SPEECHLESS integrates brassinosteroid and stomata signalling pathways. Nat. Cell Biol. **14,** 548–554.

Hackbusch, J., Richter, K., Muller, J., Salamini, F., Uhrig, J.F. (2005). A central role of *Arabidopsis thaliana* ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. Proc. Natl. Acad. Sci. USA **102,** 4908-4912.

Haebel, S., Kehr, J. (2001). Matrix-assisted laser desorption/ionization time of flight mass spectrometry peptide mass fingerprints and post source decay: a tool for the identification and analysis of phloem proteins from *Cacurbita maxima* Duch. separated by two-dimensional polyacrylamide gel electrophoresis. Planta **213,** 586-593.

Han, P., Garcia-Ponce, B., Fonseca-Salazar, G., Alvarez-Buylla, E.R., Yu, H. (2008). *AGAMOUS-LIKE 17*, a novel flowering promoter, acts in a *FT*-independent photoperiod pathway. Plant J. **55,** 253-265.

Hara, K., Kajita, R., Torii, K.U., Bergmann, D.C., Kakimoto, T. (2007). The secretory peptide gene *EPF1* enforces the stomatal one-cell-spacing rule. Genes Dev. **21,** 1720-1725.

Haran, T.E., Mohanty, U. (2009). The unique structure of A-tracts and intrinsic DNA bending. Q. Rev. Biophys. **42,** 41-81.

Harding, E.W., Tang, W., Nichols, K.W., Fernandez, D.E., Perry, S.E. (2003). Expression and maintenance of embryogenic potential is enhanced through constitutive expression of *AGAMOUS-Like 15*. Plant Physiol. **133,** 653-663.

Hartmann, U., Hohmann, S., Nettesheim, K., Wisman, E., Saedler, H., Huijser, P. (2000). Molecular cloning of *SVP*: a negative regulator of the floral transition in *Arabidopsis*. Plant J. **21,** 351-360.

Haughn, G.W., Somerville, C.R. (1988). Genetic control of morphogenesis in *Arabidopsis*. Dev. Genet. **9,** 73-89.

He, C., Saedler, H. (2005). Heterotopic expression of *MPF2* is the key to the evolution of the Chinese lantern of *Physalis*, a morphological novelty in Solanaceae. Proc. Natl. Acad. Sci. USA **102,** 5779-5784.

He, C., Saedler, H. (2007). Hormonal control of the inflated calyx syndrome, a morphological novelty, in *Physalis*. Plant J. **49,** 935-946.

Hebeler, R., Oeljeklaus, S., Reidegeld, K.E., Eisenacher, M., Stephan, C., Sitek, B., Stuhler, K., Meyer, H.E., Sturre, M.J.G., Dijkwel, P.P., Warscheid, B. (2008). Study of early leaf senescence in *Arabidopsis thaliana* by quantitative proteomics using reciprocal N-14/N-15 Labeling and difference gel electrophoresis. Mol. Cell. Proteomics **7,** 108-120.

Hecht, V., Foucher, F., Ferrandiz, C., Macknight, R., Navarro, C., Morin, J., Vardy, M.E., Ellis, N., Beltran, J.P., Rameau, C., Weller, J.L. (2005). Conservation of Arabidopsis flowering genes in model legumes. Plant Physiol. **137,** 1420-1434.

Heck, G.R., Perry, S.E., Nichols, K.W., Fernandez, D.E. (1995). AGL15, a MADS domain protein expressed in developing embryos. Plant Cell **7,** 1271-1282.

Hennig, L., Derkacheva, M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? Trends Genet. **25,** 414-423.

Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Munster, T., Theissen, G. (2002). Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. Mol. Biol. Evol. **19,** 801-814.

Hileman, L.C., Sundstrom, J.F., Litt, A., Chen, M., Shumba, T., Irish, V.F. (2006). Molecular and phylogenetic analyses of the MADS-box gene family in tomato. Mol. Biol. Evol. **23,** 2245-2258.

Hill, K., Wang, H., Perry, S.E. (2008). A transcriptional repression motif in the MADS factor AGL15 is involved in recruitment of histone deacetylase complex components. Plant J. **53,** 172-185.

Hintz, M., Bartholmes, C., Nutt, P., Ziermann, J., Hameister, S., Neuffer, B., Theissen, G. (2006). Catching a 'hopeful monster': shepherd's purse (*Capsella bursa-pastoris*) as a model system to study the evolution of flower development. J. Exp. Bot. **57,** 3531-3542.

Hirakawa, Y., Shinohara, H., Kondo, Y., Inoue, A., Nakanomyo, I., Ogawa, M., Sawa, S., Ohashi-Ito, K., Matsubayashi, Y., Fukuda, H. (2008). Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system. Proc. Natl. Acad. Sci. USA **105,** 15208-15213.

Hochholdinger, F., Sauer, M., Dembinsky, D., Hoecker, N., Muthreich, N., Saleem, M., Liu, Y. (2006). Proteomic dissection of plant development. Proteomics **6,** 4076-4083.

Holmes-Davis, R., Tanaka, C.K., Vensel, W.H., Hurkman, W.J., McCormick, S. (2005). Proteome mapping of mature pollen of *Arabidopsis thaliana*. Proteomics **5,** 4864-4884.

Hong, R.L., Hamaguchi, L., Busch, M.A., Weigel, D. (2003). Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. Plant Cell **15,** 1296-1309.

Honma, T., Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature **409,** 525-529.

Hsieh, T.F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., Zilberman, D. (2009). Genome-wide demethylation of *Arabidopsis* endosperm. Science **324,** 1451-1454.

Hu, L., Liu, S. (2012). Genome-wide analysis of the MADS-box gene family in cucumber. Genome **55,** 245-256.

Huang, H., Mizukami, Y., Hu, Y., Ma, H. (1993). Isolation and characterization of the binding sequences for the product of the Arabidopsis floral homeotic gene AGAMOUS. Nucleic Acids Res. **21,** 4769-4776.

Huang, H., Tudor, M., Weiss, C.A., Hu, Y., Ma, H. (1995). The Arabidopsis MADS-box gene AGL3 is widely expressed and encodes a sequence-specific DNA-binding protein. Plant Mol. Biol. **28,** 549-567.

Huang, H., Tudor, M., Su, T., Zhang, Y., Hu, Y., Ma, H. (1996). DNA binding properties of two Arabidopsis MADS domain proteins: binding consensus and dimer formation. Plant Cell **8,** 81-94.

Huang, K., Louis, J.M., Donaldson, L., Lim, F.L., Sharrocks, A.D., Clore, G.M. (2000). Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. EMBO J. **19,** 2615-2628.

Huang, S., Taylor, N.L., Narsai, R., Eubel, H., Whelan, J., Millar, A.H. (2009). Experimental analysis of the rice mitochondrial proteome, its biogenesis, and heterogeneity. Plant Physiol. **149,** 719-734.

Hubbard, K.E., Nishimura, N., Hitomi, K., Getzoff, E.D., Schroeder, J.I. (2010). Early abscisic acid signal transduction mechanisms: newly discovered components and newly emerging questions. Genes Dev. **24,** 1695-1708.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J. Cell Biol. **189,** 739-754.

Hubner, N.C., Mann, M. (2011). Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). Methods **53,** 453-459.

Hurtado, L., Farrona, S., Reyes, J.C. (2006). The putative SWI/SNF complex subunit BRAHMA activates flower homeotic genes in *Arabidopsis thaliana*. Plant Mol. Biol. **62,** 291-304.

Igawa, T., Fujiwara, M., Takahashi, H., Sawasaki, T., Endo, Y., Seki, M., Shinozaki, K., Fukao, Y., Yanagawa, Y. (2009). Isolation and identification of ubiquitin-related proteins from *Arabidopsis* seedlings. J. Exp. Bot. **60,** 3067-3073.

Immink, R.G., Ferrario, S., Busscher-Lange, J., Kooiker, M., Busscher, M., Angenent, G.C. (2003). Analysis of the petunia MADS-box transcription factor family. Mol. Genet. Genomics **268,** 598-606.

Immink, R.G., Tonaco, I.A., de Folter, S., Shchennikova, A., van Dijk, A.D., Busscher-Lange, J., Borst, J.W., Angenent, G.C. (2009). SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. Genome Biol. **10,** R24.

Immink, R.G., Kaufmann, K., Angenent, G.C. (2010). The 'ABC' of MADS domain protein behaviour and interactions. Semin. Cell Dev. Biol. **21,** 87-93.

Immink, R.G., Pose, D., Ferrario, S., Ott, F., Kaufmann, K., Valentim, F.L., de Folter, S., van der Wal, F., van Dijk, A.D., Schmid, M., Angenent, G.C. (2012). Characterization of SOC1's Central Role in Flowering by the Identification of Its Upstream and Downstream Regulators. Plant Physiol. **160,** 433-449.

Ippel, J.H., Pouvreau, L., Kroef, T., Gruppen, H., Versteeg, G., van den Putten, P., Struik, P.C., van Mierlo, C.P.M. (2004). In vivo uniform N-15-isotope labelling of plants: Using the greenhouse for structural proteomics. Proteomics **4,** 226-234.

Itkin, M., Seybold, H., Breitel, D., Rogachev, I., Meir, S., Aharoni, A. (2009). TOMATO AGAMOUS-LIKE 1 is a component of the fruit ripening regulatory network. Plant J. **60,** 1081-1095.

Ito, T. (2011). Coordination of flower development by homeotic master regulators. Curr. Opin. Plant Biol. **14,** 53-59.

Jaakola, L., Poole, M., Jones, M.O., Kamarainen-Karppinen, T., Koskimaki, J.J., Hohtola, A., Haggman, H., Fraser, P.D., Manning, K., King, G.J., Thomson, H., Seymour, G.B. (2010). A SQUAMOSA MADS box gene involved in the regulation of anthocyanin accumulation in bilberry fruits. Plant Physiol. **153,** 1619-1629.

Jack, T., Brockman, L.L., Meyerowitz, E.M. (1992). The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. Cell **68,** 683-697.

Jack, T., Fox, G.L., Meyerowitz, E.M. (1994). Arabidopsis homeotic gene *APETALA3* ectopic expression: transcriptional and posttranscriptional regulation determine floral organ identity. Cell **76,** 703-716.

James, P. (1997). Protein identification in the post-genome era: the rapid rise of proteomics. Q. Rev. Biophys. **30,** 279-331.

Jamet, E., Albenne, C., Boudart, G., Irshad, M., Canut, H., Pont-Lezica, R. (2008). Recent advances in plant cell wall proteomics. Proteomics **8,** 893-908.

Jang, S., Torti, S., Coupland, G. (2009). Genetic and spatial interactions between *FT*, *TSF* and *SVP* during the early stages of floral induction in Arabidopsis. Plant J. **60,** 614-625.

Jaquinod, M., Villiers, F., Kieffer-Jaquinod, S., Hugouvieu, V., Bruley, C., Garin, J., Bourguignon, J. (2007). A proteomics dissection of *Arabidopsis thaliana* vacuoles isolated from cell culture. Mol. Cell. Proteomics **6,** 394-412.

Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M., Wong, W.H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat. Biotechnol. **26,** 1293-1300.

Jiao, Y., Meyerowitz, E.M. (2010). Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. Mol. Syst. Biol. **6,** 419.

Jimenez, S., Lawton-Rauh, A.L., Reighard, G.L., Abbott, A.G., Bielenberg, D.G. (2009). Phylogenetic analysis and molecular evolution of the dormancy associated MADS-box genes from peach. BMC Plant Biol. **9,** 81.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., Dean, C. (2000). Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science **290,** 344-347.

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpaa, M.J., Bonke, M., Palin, K., Talukder, S., Hughes, T.R., Luscombe, N.M., Ukkonen, E., Taipale, J. (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. **20,** 861-873.

Kang, I.H., Steffen, J.G., Portereiko, M.F., Lloyd, A., Drews, G.N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. Plant Cell **20,** 635-647.

Kanno, A., Saeki, H., Kameya, T., Saedler, H., Theissen, G. (2003). Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). Plant Mol. Biol. **52,** 831-841.

Kanrar, S., Bhattacharya, M., Arthur, B., Courtier, J., Smith, H.M. (2008). Regulatory networks that function to specify flower meristems require the function of homeobox genes *PENNYWISE* and *POUND-FOOLISH* in *Arabidopsis*. Plant J. **54,** 924-937.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature **458,** 362-366.

Karlova, R., Boeren, S., Russinova, E., Aker, J., Vervoort, J., de Vries, S. (2006). The *Arabidopsis* SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 protein complex includes BRASSINOSTEROID-INSENSITIVE1. Plant Cell **18,** 626-638.

Karlova, R., de Vries, S.C. (2006). Advances in understanding brassinosteroid signaling. Sci. STKE **2006,** pe36.

Karlova, R. (2008). The SERK1 Protein Complexes. In Laboratory of Biochemistry (Wageningen: Wageningen University), pp. 142

Karlova, R., Boeren, S., van Dongen, W., Kwaaitaal, M., Aker, J., Vervoort, J., de Vries, S. (2009). Identification of in vitro phosphorylation sites in the *Arabidopsis thaliana* somatic embryogenesis receptor-like kinases. Proteomics **9,** 368-379.

Kaufmann, K., Anfang, N., Saedler, H., Theissen, G. (2005a). Mutant analysis, protein-protein interactions and subcellular localization of the Arabidopsis B_{sister} (ABS) protein. Mol. Genet. Genomics **274,** 103-118.

Kaufmann, K., Melzer, R., Theissen, G. (2005b). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene **347,** 183-198.

Kaufmann, K., Muino, J.M., Jauregui, R., Airoldi, C.A., Smaczniak, C., Krajewski, P., Angenent, G.C. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the *Arabidopsis* flower. PLoS Biol. **7,** e1000090.

Kaufmann, K., Muino, J.M., Osteras, M., Farinelli, L., Krajewski, P., Angenent, G.C. (2010a). Chromatin immunoprecipitation (ChIP) of plant transcription factors

followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nat. Protoc. **5,** 457-472.

Kaufmann, K., Pajoro, A., Angenent, G.C. (2010b). Regulation of transcription in plants: mechanisms controlling developmental switches. Nat. Rev. Genet. **11,** 830-842.

Kaufmann, K., Wellmer, F., Muino, J.M., Ferrier, T., Wuest, S.E., Kumar, V., Serrano-Mislata, A., Madueno, F., Krajewski, P., Meyerowitz, E.M., Angenent, G.C., Riechmann, J.L. (2010c). Orchestration of floral initiation by APETALA1. Science **328,** 85-89.

Kaufmann, K., Smaczniak, C., de Vries, S., Angenent, G.C., Karlova, R. (2011). Proteomics insights into plant signaling and development. Proteomics **11,** 744-755.

Kempin, S.A., Savidge, B., Yanofsky, M.F. (1995). Molecular basis of the *cauliflower* phenotype in *Arabidopsis*. Science **267,** 522-525.

Kersten, B., Agrawal, G.K., Durek, P., Neigenfind, J., Schulze, W., Walther, D., Rakwal, R. (2009). Plant phosphoproteomics: an update. Proteomics **9,** 964-988.

Khan, M.R., Hu, J.Y., Riss, S., He, C., Saedler, H. (2009). *MPF2-like-A* MADS-box genes control the inflated Calyx syndrome in *Withania* (Solanaceae): roles of Darwinian selection. Mol. Biol. Evol. **26,** 2463-2473.

Kim, D.H., Sung, S. (2012). Environmentally coordinated epigenetic silencing of *FLC* by protein and long noncoding RNA components. Curr. Opin. Plant Biol. **15,** 51-56.

Kim, S., Koh, J., Yoo, M.J., Kong, H., Hu, Y., Ma, H., Soltis, P.S., Soltis, D.E. (2005). Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. Plant J. **43,** 724-744.

Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjolander, K., Gruissem, W., Baginsky, S. (2004). The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. Curr. Biol. **14,** 354-362.

Kobayashi, K., Maekawa, M., Miyao, A., Hirochika, H., Kyozuka, J. (2010). *PANICLE PHYTOMER2* (*PAP2*), encoding a SEPALLATA subfamily MADS-box protein, positively controls spikelet meristem identity in rice. Plant Cell Physiol. **51,** 47-57.

Kofuji, R., Sumikawa, N., Yamasaki, M., Kondo, K., Ueda, K., Ito, M., Hasebe, M. (2003). Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. Mol. Biol. Evol. **20,** 1963-1977.

Kohler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., Grossniklaus, U. (2003). The *Polycomb*-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. Genes Dev. **17,** 1540-1553.

Kohler, C., Page, D.R., Gagliardini, V., Grossniklaus, U. (2005). The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nat. Genet. **37,** 28-30.

Koizumi, K., Wu, S., MacRae-Crerar, A., Gallagher, K.L. (2011). An essential protein that interacts with endosomes and promotes movement of the SHORT-ROOT transcription factor. Curr. Biol. **21,** 1559-1564.

Koller, A., Washburn, M.P., Lange, B.M., Andon, N.L., Deciu, C., Haynes, P.A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J., Wolters, D., Yates, J.R. (2002). Proteomic survey of metabolic pathways in rice. Proc. Natl. Acad. Sci. USA **99,** 11969-11974.

Komatsu, S. (2008). Plasma membrane proteome in *Arabidopsis* and rice. Proteomics **8,** 4137-4145.

Koo, H.S., Wu, H.M., Crothers, D.M. (1986). DNA bending at adenine-thymine tracts. Nature **320,** 501-506.

Koo, S.C., Bracko, O., Park, M.S., Schwab, R., Chun, H.J., Park, K.M., Seo, J.S., Grbic, V., Balasubramanian, S., Schmid, M., Godard, F., Yun, D.J., Lee, S.Y., Cho, M.J., Weigel, D., Kim, M.C. (2010). Control of lateral organ development and flowering time by the *Arabidopsis thaliana* MADS-box gene *AGAMOUS-LIKE6*. Plant J. **62,** 807-816.

Kramer, E.M., Hall, J.C. (2005). Evolutionary dynamics of genes controlling floral development. Curr. Opin. Plant Biol. **8,** 13-18.

Krizek, B.A., Meyerowitz, E.M. (1996). Mapping the protein regions responsible for the functional specificities of the Arabidopsis MADS domain organ-identity proteins. Proc. Natl. Acad. Sci. USA **93,** 4063-4070.

Krizek, B.A., Riechmann, J.L., Meyerowitz, E.M. (1999). Use of the *APETALA1* promoter to assay the *in vivo* function of chimeric MADS box genes. Sex. Plant Reprod. **12,** 14-26.

Krizek, B.A., Fletcher, J.C. (2005). Molecular mechanisms of flower development: an armchair guide. Nat. Rev. Genet. **6,** 688-698.

Kutter, C., Schob, H., Stadler, M., Meins, F., Jr., Si-Ammour, A. (2007). MicroRNA-mediated regulation of stomatal development in *Arabidopsis*. Plant Cell **19,** 2417-2429.

Kwantes, M., Liebsch, D., Verelst, W. (2012). How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. Mol. Biol. Evol. **29,** 293-302.

Lal, S., Pacis, L.B., Smith, H.M. (2011). Regulation of the *SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE genes/microRNA156* module by the homeodomain proteins PENNYWISE and POUND-FOOLISH in *Arabidopsis*. Mol. Plant **4,** 1123-1132.

Landis, J.B., Barnett, L.L., Hileman, L.C. (2012). Evolution of petaloid sepals independent of shifts in B-class MADS box gene expression. Dev. Genes Evol. **222,** 19-28.

Lee, H., Suh, S.S., Park, E., Cho, E., Ahn, J.H., Kim, S.G., Lee, J.S., Kwon, Y.M., Lee, I. (2000). The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. Genes Dev. **14,** 2366-2376.

Lee, H.L., Irish, V.F. (2011). Gene duplication and loss in a MADS box gene transcription factor circuit. Mol. Biol. Evol. **28,** 3367-3380.

Lee, J., Oh, M., Park, H., Lee, I. (2008). SOC1 translocated to the nucleus by interaction with AGL24 directly regulates *LEAFY*. Plant J. **55,** 832-843.

Lehti-Shiu, M.D., Adamczyk, B.J., Fernandez, D.E. (2005). Expression of MADS-box genes during the embryonic phase in *Arabidopsis*. Plant Mol. Biol. **58,** 89-107.

Lenser, T., Theissen, G., Dittrich, P. (2009). Developmental robustness by obligate interaction of class B floral homeotic genes and proteins. PLoS Comput. Biol. **5,** e1000264.

Leseberg, C.H., Li, A., Kang, H., Duvall, M., Mao, L. (2006). Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. Gene **378,** 84-94.

Leseberg, C.H., Eissler, C.L., Wang, X., Johns, M.A., Duvall, M.R., Mao, L. (2008). Interaction study of MADS-domain proteins in tomato. J. Exp. Bot. **59,** 2253-2265.

Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E., Yu, H. (2008). A repressor complex governs the integration of flowering signals in *Arabidopsis*. Dev. Cell **15,** 110-120.

Li, H., Wong, W.S., Zhu, L., Guo, H.W., Ecker, J., Li, N. (2009). Phosphoproteomic analysis of ethylene-regulated protein phosphorylation in etiolated seedlings of *Arabidopsis* mutant *ein2* using two-dimensional separations coupled with a hybrid quadrupole time-of-flight mass spectrometer. Proteomics **9,** 1646-1661.

Li, H., Liang, W., Hu, Y., Zhu, L., Yin, C., Xu, J., Dreni, L., Kater, M.M., Zhang, D. (2011). Rice *MADS6* interacts with the floral homeotic genes *SUPERWOMAN1*, *MADS3*, *MADS58*, *MADS13*, and *DROOPING LEAF* in specifying floral organ identities and meristem fate. Plant Cell **23,** 2536-2552.

Li, J., Wen, J., Lease, K.A., Doke, J.T., Tax, F.E., Walker, J.C. (2002). BAK1, an *Arabidopsis* LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. Cell **110,** 213-222.

Li, L. (2009). GADEM: a genetic algorithm guided formation of spaced dyads coupled with an EM algorithm for motif discovery. J. Comput. Biol. **16,** 317-329.

Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics **25,** 1966-1967.

Li, Z., Reighard, G.L., Abbott, A.G., Bielenberg, D.G. (2009). Dormancy-associated MADS genes from the *EVG* locus of peach [*Prunus persica* (L.) Batsch] have distinct seasonal and photoperiodic expression patterns. J. Exp. Bot. **60,** 3521-3530.

Liljegren, S.J., Ditta, G.S., Eshed, Y., Savidge, B., Bowman, J.L., Yanofsky, M.F. (2000). *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. Nature **404,** 766-770.

Lilley, K.S., Dupree, P. (2007). Plant organelle proteomics. Curr. Opin. Plant Biol. **10,** 594-599.

Lin, M.K., Lee, Y.J., Lough, T.J., Phinney, B.S., Lucas, W.J. (2009). Analysis of the pumpkin phloem proteome provides Insights into angiosperm sieve tube function. Mol. Cell. Proteomics **8,** 343-356.

Liu, C., Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T., Yu, H. (2007). Specification of *Arabidopsis* floral meristem identity by repression of flowering time genes. Development **134,** 1901-1910.

Liu, C., Chen, H., Er, H.L., Soo, H.M., Kumar, P.P., Han, J.H., Liou, Y.C., Yu, H. (2008). Direct interaction of *AGL24* and *SOC1* integrates flowering signals in *Arabidopsis*. Development **135,** 1481-1491.

Liu, C., Xi, W., Shen, L., Tan, C., Yu, H. (2009). Regulation of floral patterning by flowering time genes. Dev. Cell **16,** 711-722.

Liu, C.J., Zhang, J.A., Zhang, N., Shan, H.Y., Su, K.M., Zhang, J.S., Meng, Z., Kong, H.Z., Chen, Z.D. (2010). Interactions among proteins of floral MADS-box genes in

basal eudicots: implications for evolution of the regulatory network for flower development. Mol. Biol. Evol. **27,** 1598-1611.

Lu, F., Cui, X., Zhang, S., Jenuwein, T., Cao, X. (2011). *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase. Nat. Genet. **43,** 715-719.

Lu, X.J., Olson, W.K. (2008). 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nat. Protoc. **3,** 1213-1227.

Lucas, W.J., Lee, J.Y. (2004). Plant cell biology - Plasmodesmata as a supracellular control network in plants. Nat. Rev. Mol. Cell Biol. **5,** 712-726.

Lyngved, R., Renaut, J., Hausman, J.F., Iversen, T.H., Hvoslef-Eide, A.K. (2008). Embryo-specific proteins in *Cyclamen persicum* analyzed with 2-D DIGE. J. Plant Growth Regul. **27,** 353-369.

MacAlister, C.A., Ohashi-Ito, K., Bergmann, D.C. (2007). Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. Nature **445,** 537-540.

Mak, Y., Willows, R.D., Roberts, T.H., Wrigley, C.W., Sharp, P.J., Copeland, L. (2009). Germination of wheat: a functional proteomics analysis of the embryo. Cereal Chem. **86,** 281-289.

Makarevich, G., Villar, C.B., Erilova, A., Kohler, C. (2008). Mechanism of *PHERES1* imprinting in *Arabidopsis*. J. Cell Sci. **121,** 906-912.

Malcomber, S.T., Kellogg, E.A. (2004). Heterogeneous expression patterns and separate roles of the *SEPALLATA* gene *LEAFY HULL STERILE1* in grasses. Plant Cell **16,** 1692-1706.

Mandel, M.A., Gustafson-Brown, C., Savidge, B., Yanofsky, M.F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature **360,** 273-277.

Mandel, M.A., Yanofsky, M.F. (1998). The *Arabidopsis AGL9* MADS box gene is expressed in young flower primordia. Sex. Plant Reprod. **11,** 22-28.

Manzano, C., Abraham, Z., Lopez-Torrejon, G., Del Pozo, J.C. (2008). Identification of ubiquitinated proteins in *Arabidopsis*. Plant Mol. Biol. **68,** 145-158.

Mao, L., Begum, D., Chuang, H.W., Budiman, M.A., Szymkowiak, E.J., Irish, E.E., Wing, R.A. (2000). *JOINTLESS* is a MADS-box gene controlling tomato flower abscission zone development. Nature **406,** 910-913.

Maor, R., Jones, A., Nuhse, T.S., Studholme, D.J., Peck, S.C., Shirasu, K. (2007). Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. Mol. Cell. Proteomics **6,** 601-610.

Martel, C., Vrebalov, J., Tafelmeyer, P., Giovannoni, J.J. (2011). The tomato MADS-box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening processes in a COLORLESS NONRIPENING-dependent manner. Plant Physiol. **157,** 1568-1579.

Masiero, S., Li, M.A., Will, I., Hartmann, U., Saedler, H., Huijser, P., Schwarz-Sommer, Z., Sommer, H. (2004). *INCOMPOSITA*: a MADS-box gene controlling prophyll development and floral meristem identity in *Antirrhinum*. Development **131,** 5981-5990.

Masiero, S., Colombo, L., Grini, P.E., Schnittger, A., Kater, M.M. (2011). The emerging importance of type I MADS box transcription factors for plant reproduction. Plant Cell **23,** 865-872.

Mathieu, J., Yant, L.J., Murdter, F., Kuttner, F., Schmid, M. (2009). Repression of flowering by the miR172 target SMZ. PLoS Biol. **7,** e1000148.

Matias-Hernandez, L., Battaglia, R., Galbiati, F., Rubes, M., Eichenberger, C., Grossniklaus, U., Kater, M.M., Colombo, L. (2010). *VERDANDI* is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in *Arabidopsis*. Plant Cell **22,** 1702-1715.

Matsubayashi, Y., Sakagami, Y. (2006). Peptide hormones in plants. Annu. Rev. Plant Biol. **57,** 649-674.

Matsumoto, N., Okada, K. (2001). A homeobox gene, PRESSED FLOWER, regulates lateral axis-dependent development of Arabidopsis flowers. Genes Dev. **15,** 3355-3364.

McGonigle, B., Bouhidel, K., Irish, V.F. (1996). Nuclear localization of the Arabidopsis *APETALA3* and *PISTILLATA* homeotic gene products depends on their simultaneous expression. Genes Dev. **10,** 1812-1821.

Melzer, R., Theissen, G. (2009). Reconstitution of 'floral quartets' *in vitro* involving class B and class E floral homeotic proteins. Nucleic Acids Res. **37,** 2723-2736.

Melzer, R., Verelst, W., Theissen, G. (2009). The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes *in vitro*. Nucleic Acids Res. **37,** 144-157.

Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A., Beeckman, T. (2008). Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. Nat. Genet. **40,** 1489-1492.

Michaels, S.D., Amasino, R.M. (1999). *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell **11,** 949-956.

Michaels, S.D., Ditta, G., Gustafson-Brown, C., Pelaz, S., Yanofsky, M., Amasino, R.M. (2003). *AGL24* acts as a promoter of flowering in *Arabidopsis* and is positively regulated by vernalization. Plant J. **33,** 867-874.

Michniewicz, M., Zago, M.K., Abas, L., Weijers, D., Schweighofer, A., Meskiene, I., Heisler, M.G., Ohno, C., Zhang, J., Huang, F., Schwab, R., Weigel, D., Meyerowitz, E.M., Luschnig, C., Offringa, R., Friml, J. (2007). Antagonistic regulation of PIN phosphorylation by PP2A and PINOID directs auxin flux. Cell **130,** 1044-1056.

Miura, K., Hasegawa, P.M. (2010). Sumoylation and other ubiquitin-like post-translational modifications in plants. Trends Cell Biol. **20,** 223-232.

Mizzotti, C., Mendes, M.A., Caporali, E., Schnittger, A., Kater, M.M., Battaglia, R., Colombo, L. (2012). The MADS box genes *SEEDSTICK* and *ARABIDOPSIS* B_{sister} play a maternal role in fertilization and seed development. Plant J. **70,** 409-420.

Mondragon-Palomino, M., Theissen, G. (2008). MADS about the evolution of orchid flowers. Trends Plant Sci. **13,** 51-59.

Mondragon-Palomino, M., Theissen, G. (2011). Conserved differential expression of paralogous *DEFICIENS*- and *GLOBOSA*-like MADS-box genes in the flowers of Orchidaceae: refining the 'orchid code'. Plant J. **66,** 1008-1019.

Montoya, T., Nomura, T., Farrar, K., Kaneta, T., Yokota, T., Bishop, G.J. (2002). Cloning the tomato curl3 gene highlights the putative dual role of the leucine-rich repeat receptor kinase tBRI1/SR160 in plant steroid hormone and peptide hormone signaling. Plant Cell **14,** 3163-3176.

Montrichard, F., Alkhalfioui, F., Yano, H., Vensel, W.H., Hurkman, W.J., Buchanan, B.B. (2009). Thioredoxin targets in plants: the first 30 years. Journal of Proteomics **72,** 452-474.

Mooney, B.P., Miernyk, J.A., Greenlief, C.M., Thelen, J.J. (2006). Using quantitative proteomics of *Arabidopsis* roots and leaves to predict metabolic activity. Physiol. Plant. **128,** 237-250.

Moreno-Risueno, M.A., Van Norman, J.M., Moreno, A., Zhang, J., Ahnert, S.E., Benfey, P.N. (2010). Oscillating gene expression determines competence for periodic *Arabidopsis* root branching. Science **329,** 1306-1311.

Moyroud, E., Minguet, E.G., Ott, F., Yant, L., Pose, D., Monniaux, M., Blanchet, S., Bastien, O., Thevenon, E., Weigel, D., Schmid, M., Parcy, F. (2011). Prediction of regulatory interactions from genome sequences using a biophysical model for the *Arabidopsis* LEAFY transcription factor. Plant Cell **23,** 1293-1306.

Mravec, J., Petrasek, J., Li, N., Boeren, S., Karlova, R., Kitakura, S., Parezova, M., Naramoto, S., Nodzynski, T., Dhonukshe, P., Bednarek, S.Y., Zazimalova, E., de Vries, S., Friml, J. (2011). Cell plate restricted association of DRP1A and PIN proteins is required for cell polarity establishment in Arabidopsis. Curr. Biol. **21,** 1055-1060.

Muino, J.M., Kaufmann, K., van Ham, R.C., Angenent, G.C., Krajewski, P. (2011). ChIP-seq Analysis in R (CSAR): An R package for the statistical detection of protein-bound genomic regions. Plant Methods **7,** 11.

Nah, G., Jeffrey Chen, Z. (2010). Tandem duplication of the FLC locus and the origin of a new gene in *Arabidopsis* related species and their functional implications in allopolyploids. New Phytol. **186,** 228-238.

Nakano, T., Kimbara, J., Fujisawa, M., Kitagawa, M., Ihashi, N., Maeda, H., Kasumi, T., Ito, Y. (2012). MACROCALYX and JOINTLESS interact in the transcriptional regulation of tomato fruit abscission zone development. Plant Physiol. **158,** 439-450.

Nam, K.H., Li, J. (2002). BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. Cell **110,** 203-212.

Nawy, T., Lee, J.Y., Colinas, J., Wang, J.Y., Thongrod, S.C., Malamy, J.E., Birnbaum, K., Benfey, P.N. (2005). Transcriptional profile of the *Arabidopsis* root quiescent center. Plant Cell **17,** 1908-1925.

Ndimba, B.K., Chivasa, S., Simon, W.J., Slabas, A.R. (2005). Identification of *Arabidopsis* salt and osmotic stress responsive proteins using two-dimensional difference gel electrophoresis and mass spectrometry. Proteomics **5,** 4185-4196.

Nesi, N., Debeaujon, I., Jond, C., Stewart, A.J., Jenkins, G.I., Caboche, M., Lepiniec, L. (2002). The *TRANSPARENT TESTA16* locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell **14,** 2463-2479.

Ng, M., Yanofsky, M.F. (2001). Activation of the Arabidopsis B class homeotic genes by *APETALA1*. Plant Cell **13,** 739-753.

Ning, K., Fermin, D., Nesvizhskii, A.I. (2012). Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. J. Proteome Res. **11,** 2261-2271.

Nishimura, N., Hitomi, K., Arvai, A.S., Rambo, R.P., Hitomi, C., Cutler, S.R., Schroeder, J.I., Getzoff, E.D. (2009). Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. Science **326,** 1373-1379.

Nishimura, N., Sarkeshik, A., Nito, K., Park, S.Y., Wang, A., Carvalho, P.C., Lee, S., Caddell, D.F., Cutler, S.R., Chory, J., Yates, J.R., Schroeder, J.I. (2010). PYR/PYL/RCAR family members are major in-vivo ABI1 protein phosphatase 2C-interacting proteins in Arabidopsis. Plant J. **61,** 290-299.

Norman, C., Runswick, M., Pollock, R., Treisman, R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the *c-fos* serum response element. Cell **55,** 989-1003.

Nuhse, T.S., Bottrill, A.R., Jones, A.M., Peck, S.C. (2007). Quantitative phosphoproteomic analysis of plasma membrane proteins reveals regulatory mechanisms of plant innate immune responses. Plant J. **51,** 931-940.

Nurrish, S.J., Treisman, R. (1995). DNA binding specificity determinants in MADS-box transcription factors. Mol. Cell. Biol. **15,** 4076-4085.

Oh, M.H., Wang, X., Kota, U., Goshe, M.B., Clouse, S.D., Huber, S.C. (2009). Tyrosine phosphorylation of the BRI1 receptor kinase emerges as a component of brassinosteroid signaling in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **106,** 658-663.

Ohmori, S., Kimizu, M., Sugita, M., Miyao, A., Hirochika, H., Uchida, E., Nagato, Y., Yoshida, H. (2009). *MOSAIC FLORAL ORGANS1*, an *AGL6*-like MADS box gene, regulates floral organ identity and meristem fate in rice. Plant Cell **21,** 3008-3025.

Ohyama, K., Ogawa, M., Matsubayashi, Y. (2008). Identification of a biologically active, small, secreted peptide in *Arabidopsis* by in silico gene screening, followed by LC-MS-based structure analysis. Plant J. **55,** 152-160.

Ohyama, K., Shinohara, H., Ogawa-Ohnishi, M., Matsubayashi, Y. (2009). A glycopeptide regulating stem cell fate in *Arabidopsis thaliana*. Nat. Chem. Biol. **5,** 578-580.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., Ahn, N.G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol. Cell. Proteomics **4,** 1487-1502.

Olinares, P.D.B., Ponnala, L., van Wijk, K.J. (2010). Megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* characterized by size exclusion

chromatography, mass spectrometry, and hierarchical clustering. Mol. Cell. Proteomics **9,** 1594-1615.

Pabon-Mora, N., Ambrose, B., Litt, A. (2012). Poppy *APETALA1/FRUITFULL* orthologs control flowering time, branching, perianth identity and fruit development. Plant Physiol. **158,** 1685-1704.

Parenicová, L., de Folter, S., Kieffer, M., Horner, D.S., Favalli, C., Busscher, J., Cook, H.E., Ingram, R.M., Kater, M.M., Davies, B., Angenent, G.C., Colombo, L. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. Plant Cell **15,** 1538-1551.

Passmore, S., Maine, G.T., Elble, R., Christ, C., Tye, B.K. (1988). *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of *MAT alpha* cells. J. Mol. Biol. **204,** 593-606.

Pauwels, L., Barbero, G.F., Geerinck, J., Tilleman, S., Grunewald, W., Perez, A.C., Chico, J.M., Bossche, R.V., Sewell, J., Gil, E., Garcia-Casado, G., Witters, E., Inze, D., Long, J.A., De Jaeger, G., Solano, R., Goossens, A. (2010). NINJA connects the co-repressor TOPLESS to jasmonate signalling. Nature **464,** 788-791.

Pawlowski, T.A. (2009). Proteome analysis of Norway maple (*Acer platanoides* L.) seeds dormancy breaking and germination: influence of abscisic and gibberellic acids. BMC Plant Biol. **9,** 48.

Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., Yanofsky, M.F. (2000). B and C floral organ identity functions require *SEPALLATA* MADS-box genes. Nature **405,** 200-203.

Pellegrini, L., Tan, S., Richmond, T.J. (1995). Structure of serum response factor core bound to DNA. Nature **376,** 490-498.

Peng, J.M., Schwartz, D., Elias, J.E., Thoreen, C.C., Cheng, D.M., Marsischky, G., Roelofs, J., Finley, D., Gygi, S.P. (2003). A proteomics approach to understanding protein ubiquitination. Nat. Biotechnol. **21,** 921-926.

Perbal, M.C., Haughn, G., Saedler, H., Schwarz-Sommer, Z. (1996). Non-cell-autonomous function of the *Antirrhinum* floral homeotic proteins *DEFICIENS* and *GLOBOSA* is exerted by their polar cell-to-cell trafficking. Development **122,** 3433-3441.

Perry, S.E., Lehti, M.D., Fernandez, D.E. (1999). The MADS-domain protein AGAMOUS-like 15 accumulates in embryonic tissues with diverse origins. Plant Physiol. **120,** 121-130.

Petrasek, J., Mravec, J., Bouchard, R., Blakeslee, J.J., Abas, M., Seifertova, D., Wisniewska, J., Tadele, Z., Kubes, M., Covanova, M., Dhonukshe, P., Skupa, P., Benkova, E., Perry, L., Krecek, P., Lee, O.R., Fink, G.R., Geisler, M., Murphy, A.S., Luschnig, C., Zazimalova, E., Friml, J. (2006). PIN proteins perform a rate-limiting function in cellular auxin efflux. Science **312,** 914-918.

Pflieger, D., Bigeard, J., Hirt, H. (2011). Isolation and characterization of plant protein complexes by mass spectrometry. Proteomics **11,** 1824-1833.

Pinyopich, A., Ditta, G.S., Savidge, B., Liljegren, S.J., Baumann, E., Wisman, E., Yanofsky, M.F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature **424,** 85-88.

Piwarzyk, E., Yang, Y., Jack, T. (2007). Conserved C-terminal motifs of the Arabidopsis proteins APETALA3 and PISTILLATA are dispensable for floral organ identity function. Plant Physiol. **145,** 1495-1505.

Poethig, R.S. (2003). Phase change and the regulation of developmental timing in plants. Science **301,** 334-336.

Pollock, R., Treisman, R. (1990). A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res. **18,** 6197-6204.

Pollock, R., Treisman, R. (1991). Human SRF-related proteins: DNA-binding properties and potential regulatory targets. Genes Dev. **5,** 2327-2341.

Portereiko, M.F., Lloyd, A., Steffen, J.G., Punwani, J.A., Otsuga, D., Drews, G.N. (2006). *AGL80* is required for central cell and endosperm development in *Arabidopsis*. Plant Cell **18,** 1862-1872.

Pose, D., Yant, L., Schmid, M. (2012). The end of innocence: flowering networks explode in complexity. Curr. Opin. Plant Biol. **15,** 45-50.

Poser, I., Sarov, M., Hutchins, J.R., Heriche, J.K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M.M., Hofemeister, H., Zhang, Y., Nasmyth, K., White, K.P., Dietzel, S., Mechtler, K., Durbin, R., Stewart, A.F., Peters, J.M., Buchholz, F., Hyman, A.A. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat. Meth. **5,** 409-415.

Prasad, K., Zhang, X., Tobon, E., Ambrose, B.A. (2010). The Arabidopsis B-sister MADS-box protein, GORDITA, represses fruit growth and contributes to integument development. Plant J. **62,** 203-214.

Prosseda, G., Mazzola, A., Di Martino, M.L., Tielker, D., Micheli, G., Colonna, B. (2010). A temperature-induced narrow DNA curvature range sustains the maximum activity of a bacterial promoter in vitro. Biochemistry **49,** 2778-2785.

Quirino, B.F., Candido, E.S., Campos, P.F., Franco, O.L., Kruger, R.H. (2010). Proteomic approaches to study plant-pathogen interactions. Phytochemistry **71,** 351-362.

Ratcliffe, O.J., Nadzan, G.C., Reuber, T.L., Riechmann, J.L. (2001). Regulation of flowering in Arabidopsis by an *FLC* homologue. Plant Physiol. **126,** 122-132.

Ratcliffe, O.J., Kumimoto, R.W., Wong, B.J., Riechmann, J.L. (2003). Analysis of the Arabidopsis *MADS AFFECTING FLOWERING* gene family: *MAF2* prevents vernalization by short periods of cold. Plant Cell **15,** 1159-1169.

Reeves, P.A., He, Y., Schmitz, R.J., Amasino, R.M., Panella, L.W., Richards, C.M. (2007). Evolutionary conservation of the *FLOWERING LOCUS C*-mediated vernalization response: evidence from the sugar beet (*Beta vulgaris*). Genetics **176,** 295-307.

Reiland, S., Messerli, G., Baerenfaller, K., Gerrits, B., Endler, A., Grossmann, J., Gruissem, W., Baginsky, S. (2009). Large-scale *Arabidopsis* phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. Plant Physiol. **150,** 889-903.

Remmerie, N., Roef, L., Van De Slijke, E., Van Leene, J., Persiau, G., Eeckhout, D., Stals, H., Laukens, K., Lemiere, F., Esmans, E., Van Onckelen, H., Inze, D., De Jaeger, G., Witters, E. (2009). A bioanalytical method for the proteome wide display and analysis of protein complexes from whole plant cell lysates. Proteomics **9,** 598-609.

Reumann, S., Babujee, L., Ma, C.L., Wienkoop, S., Siemsen, T., Antonicelli, G.E., Rasche, N., Luder, F., Weckwerth, W., Jahn, O. (2007). Proteome analysis of *Arabidopsis* leaf peroxisomes reveals novel targeting peptides, metabolic pathways, and defense mechanisms. Plant Cell **19,** 3170-3193.

Reumann, S., Quan, S., Aung, K., Yang, P.F., Manandhar-Shrestha, K., Holbrook, D., Linka, N., Switzenberg, R., Wilkerson, C.G., Weber, A.P.M., Olsen, L.J., Hu, J.P. (2009). In-depth proteome analysis of *Arabidopsis* leaf peroxisomes combined with in vivo subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. Plant Physiol. **150,** 125-143.

Riechmann, J.L., Krizek, B.A., Meyerowitz, E.M. (1996a). Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. Proc. Natl. Acad. Sci. USA **93,** 4793-4798.

Riechmann, J.L., Wang, M., Meyerowitz, E.M. (1996b). DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Res. **24,** 3134-3141.

Riechmann, J.L., Meyerowitz, E.M. (1997). Determination of floral organ identity by Arabidopsis MADS domain homeotic proteins AP1, AP3, PI, and AG is independent of their DNA-binding specificity. Mol. Biol. Cell **8,** 1243-1259.

Rijpkema, A.S., Royaert, S., Zethof, J., van der Weerden, G., Gerats, T., Vandenbussche, M. (2006). Analysis of the *Petunia TM6* MADS box gene reveals functional divergence within the *DEF/AP3* lineage. Plant Cell **18,** 1819-1832.

Rijpkema, A.S., Zethof, J., Gerats, T., Vandenbussche, M. (2009). The petunia *AGL6* gene has a *SEPALLATA*-like function in floral patterning. Plant J. **60,** 1-9.

Rinner, O., Mueller, L.N., Hubalek, M., Muller, M., Gstaiger, M., Aebersold, R. (2007). An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. Nat. Biotechnol. **25,** 345-352.

Roeder, A.H., Ferrandiz, C., Yanofsky, M.F. (2003). The role of the REPLUMLESS homeodomain protein in patterning the *Arabidopsis* fruit. Curr. Biol. **13,** 1630-1635.

Rohila, J.S., Chen, M., Cerny, R., Fromm, M.E. (2004). Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. Plant J. **38,** 172-181.

Rohila, J.S., Chen, M., Chen, S., Chen, J., Cerny, R., Dardick, C., Canlas, P., Xu, X., Gribskov, M., Kanrar, S., Zhu, J.K., Ronald, P., Fromm, M.E. (2006). Protein-

protein interactions of tandem affinity purification-tagged protein kinases in rice. Plant J. **46,** 1-13.

Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., Honig, B. (2009). The role of DNA shape in protein-DNA recognition. Nature **461,** 1248-1253.

Rojo, E., Sharma, V.K., Kovaleva, V., Raikhel, N.V., Fletcher, J.C. (2002). CLV3 is localized to the extracellular space, where it activates the *Arabidopsis* CLAVATA stem cell signaling pathway. Plant Cell **14,** 969-977.

Rosin, F.M., Hart, J.K., Van Onckelen, H., Hannapel, D.J. (2003). Suppression of a vegetative MADS box gene of potato activates axillary meristem development. Plant Physiol. **131,** 1613-1622.

Rubio, V., Shen, Y., Saijo, Y., Liu, Y., Gusmaroli, G., Dinesh-Kumar, S.P., Deng, X.W. (2005). An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation. Plant J. **41,** 767-778.

Ruokolainen, S., Ng, Y.P., Albert, V.A., Elomaa, P., Teeri, T.H. (2010). Large scale interaction analysis predicts that the *Gerbera hybrida* floral E function is provided both by general and specialized proteins. BMC Plant Biol. **10,** 129.

Sadowski, P.G., Groen, A.J., Dupree, P., Lilley, K.S. (2008). Sub-cellular localization of membrane proteins. Proteomics **8,** 3991-4011.

Sakai, H., Medrano, L.J., Meyerowitz, E.M. (1995). Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries. Nature **378,** 199-203.

Salome, P.A., Bomblies, K., Laitinen, R.A., Yant, L., Mott, R., Weigel, D. (2011). Genetic architecture of flowering-time variation in *Arabidopsis thaliana*. Genetics **188,** 421-433.

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., Coupland, G. (2000). Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. Science **288,** 1613-1616.

Santner, A., Estelle, M. (2010). The ubiquitin-proteasome system regulates plant hormone signaling. Plant J. **61,** 1029-1040.

Saracco, S.A., Hansson, M., Scalf, M., Walker, J.M., Smith, L.M., Vierstra, R.D. (2009). Tandem affinity purification and mass spectrometric analysis of ubiquitylated proteins in *Arabidopsis*. Plant J. **59,** 344-358.

Sather, D.N., Jovanovic, M., Golenberg, E.M. (2010). Functional analysis of B and C class floral organ genes in spinach demonstrates their role in sexual dimorphism. BMC Plant Biol. **10,** 46.

Schnittger, A., Jurgens, G., Hulskamp, M. (1998). Tissue layer and organ specificity of trichome formation are regulated by GLABRA1 and TRIPTYCHON in Arabidopsis. Development **125,** 2283-2289.

Schonrock, N., Bouveret, R., Leroy, O., Borghi, L., Kohler, C., Gruissem, W., Hennig, L. (2006). Polycomb-group proteins repress the floral activator *AGL19* in the *FLC*-independent vernalization pathway. Genes Dev. **20,** 1667-1678.

Schranz, M.E., Quijada, P., Sung, S.B., Lukens, L., Amasino, R., Osborn, T.C. (2002). Characterization and effects of the replicated flowering time gene *FLC* in *Brassica rapa*. Genetics **162,** 1457-1468.

Schroter, Y., Steiner, S., Matthai, K., Pfannschmidt, T. (2010). Analysis of oligomeric protein complexes in the chloroplast sub-proteome of nucleic acid-binding proteins from mustard reveals potential redox regulators of plastid gene expression. Proteomics **10,** 2191-2204.

Schulze, B., Mentzel, T., Jehle, A.K., Mueller, K., Beeler, S., Boller, T., Felix, G., Chinchilla, D. (2010). Rapid heteromerization and phosphorylation of ligand-activated plant transmembrane receptors and their associated kinase BAK1. J. Biol. Chem. **285,** 9444-9451.

Schulze, W.X. (2010). Proteomics approaches to understand protein phosphorylation in pathway modulation. Curr. Opin. Plant Biol. **13,** 280-287.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature **473,** 337-342.

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. Science **250,** 931-936.

Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P.J., Hansen, R., Tetens, F., Lonnig, W.E., Saedler, H., Sommer, H. (1992). Characterization of the *Antirrhinum* floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. EMBO J. **11,** 251-263.

Schwarz-Sommer, Z., Davies, B., Hudson, A. (2003). An everlasting pioneer: the story of *Antirrhinum* research. Nat. Rev. Genet. **4,** 657-666.

Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasino, R.A., Coupland, G. (2006). The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. Genes Dev. **20,** 898-912.

Seo, E., Lee, H., Jeon, J., Park, H., Kim, J., Noh, Y.S., Lee, I. (2009). Crosstalk between cold response and flowering in *Arabidopsis* is mediated through the flowering-time gene *SOC1* and its upstream negative regulator *FLC*. Plant Cell **21,** 3185-3197.

Serikawa, K.A., Martinez-Laborda, A., Kim, H.S., Zambryski, P.C. (1997). Localization of expression of *KNAT3*, a class 2 *knotted1*-like gene. Plant J. **11,** 853-861.

Seymour, G.B., Ryder, C.D., Cevik, V., Hammond, J.P., Popovich, A., King, G.J., Vrebalov, J., Giovannoni, J.J., Manning, K. (2011). A *SEPALLATA* gene is involved in the development and ripening of strawberry (*Fragaria* x *ananassa* Duch.) fruit, a non-climacteric tissue. J. Exp. Bot. **62,** 1179-1188.

Sghaier-Hammami, B., Valledor, L., Drira, N., Jorrin-Novo, J.V. (2009). Proteomic analysis of the development and germination of date palm (*Phoenix dactylifera* L.) zygotic embryos. Proteomics **9,** 2543-2554.

Shan, H., Zahn, L., Guindon, S., Wall, P.K., Kong, H., Ma, H., DePamphilis, C.W., Leebens-Mack, J. (2009). Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. Mol. Biol. Evol. **26,** 2229-2244.

Sharrocks, A.D., Gille, H., Shaw, P.E. (1993). Identification of amino acids essential for DNA binding and dimerization in p67SRF: implications for a novel DNA-binding motif. Mol. Cell. Biol. **13,** 123-132.

Shikata, M., Koyama, T., Mitsuda, N., Ohme-Takagi, M. (2009). Arabidopsis SBP-box genes SPL10, SPL11 and SPL2 control morphological change in association with shoot maturation in the reproductive phase. Plant Cell Physiol. **50,** 2133-2145.

Shirzadi, R., Andersen, E.D., Bjerkan, K.N., Gloeckle, B.M., Heese, M., Ungru, A., Winge, P., Koncz, C., Aalen, R.B., Schnittger, A., Grini, P.E. (2011). Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of *AGAMOUS-LIKE36*. PLoS Genet. **7,** e1001303.

Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F., Li, W.H. (2004). Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. Plant Cell **16,** 1220-1234.

Sieburth, L.E., Drews, G.N., Meyerowitz, E.M. (1998). Non-autonomy of *AGAMOUS* function in flower development: use of a Cre/*loxP* method for mosaic analysis in *Arabidopsis*. Development **125,** 4303-4312.

Singer, S.D., Krogan, N.T., Ashton, N.W. (2007). Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. Plant Cell Rep. **26,** 1155-1169.

Sitaraman, J., Bui, M., Liu, Z. (2008). *LEUNIG_HOMOLOG* and *LEUNIG* perform partially redundant functions during *Arabidopsis* embryo and floral development. Plant Physiol. **147,** 672-681.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., Mann, R.S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell **147,** 1270-1282.

Smaczniak, C., Immink, R.G., Angenent, G.C., Kaufmann, K. (2012a). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development **139,** 3081-3098.

Smaczniak, C., Immink, R.G., Muino, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Parcy, F., Xu, L., Carles, C.C., Angenent, G.C., Kaufmann, K. (2012b). Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. Proc. Natl. Acad. Sci. USA **109,** 1560-1565.

Smyth, D.R., Bowman, J.L., Meyerowitz, E.M. (1990). Early flower development in Arabidopsis. Plant Cell **2,** 755-767.

Soares, N.C., Francisco, R., Ricardo, C.P., Jackson, P.A. (2007). Proteomics of ionically bound and soluble extracellular proteins in *Medicago truncatula* leaves. Proteomics **7,** 2070-2082.

Soltis, D.E., Soltis, P.S., Albert, V.A., Oppenheimer, D.G., dePamphilis, C.W., Ma, H., Frohlich, M.W., Theissen, G. (2002). Missing links: the genetic architecture of flower and floral diversification. Trends Plant Sci. **7,** 22-31; dicussion 31-24.

Sridhar, V.V., Surendrarao, A., Liu, Z. (2006). *APETALA1* and *SEPALLATA3* interact with *SEUSS* to mediate transcription repression during flower development. Development **133,** 3159-3166.

Steffen, J.G., Kang, I.H., Portereiko, M.F., Lloyd, A., Drews, G.N. (2008). AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. Plant Physiol. **148,** 259-268.

Stefl, R., Wu, H., Ravindranathan, S., Sklenar, V., Feigon, J. (2004). DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum. Proc. Natl. Acad. Sci. USA **101,** 1177-1182.

Sun, B., Xu, Y., Ng, K.H., Ito, T. (2009). A timing mechanism for stem cell maintenance and differentiation in the *Arabidopsis* floral meristem. Genes Dev. **23,** 1791-1804.

Tadege, M., Sheldon, C.C., Helliwell, C.A., Stoutjesdijk, P., Dennis, E.S., Peacock, W.J. (2001). Control of flowering time by *FLC* orthologues in *Brassica napus*. Plant J. **28,** 545-553.

Takayama, S., Shimosato, H., Shiba, H., Funato, M., Che, F.S., Watanabe, M., Iwano, M., Isogai, A. (2001). Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility. Nature **413,** 534-538.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28,** 2731-2739.

Tan, S., Richmond, T.J. (1998). Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. Nature **391,** 660-666.

Tanabe, Y., Hasebe, M., Sekimoto, H., Nishiyama, T., Kitani, M., Henschel, K., Munster, T., Theissen, G., Nozaki, H., Ito, M. (2005). Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. Proc. Natl. Acad. Sci. USA **102,** 2436-2441.

Tandre, K., Albert, V.A., Sundas, A., Engstrom, P. (1995). Conifer homologues to genes that control floral development in angiosperms. Plant Mol. Biol. **27,** 69-78.

Tang, W., Perry, S.E. (2003). Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. J. Biol. Chem. **278,** 28154-28159.

Tang, W., Deng, Z., Oses-Prieto, J.A., Suzuki, N., Zhu, S., Zhang, X., Burlingame, A.L., Wang, Z.Y. (2008a). Proteomics studies of brassinosteroid signal transduction using prefractionation and two-dimensional DIGE. Mol. Cell. Proteomics **7,** 728-738.

Tang, W., Kim, T.W., Oses-Prieto, J.A., Sun, Y., Deng, Z., Zhu, S., Wang, R., Burlingame, A.L., Wang, Z.Y. (2008b). BSKs mediate signal transduction from the receptor kinase BRI1 in *Arabidopsis*. Science **321,** 557-560.

Tang, W., Deng, Z., Wang, Z.Y. (2010). Proteomics shed light on the brassinosteroid signaling mechanisms. Curr. Opin. Plant Biol. **13,** 27-33.

Tao, Z., Shen, L., Liu, C., Liu, L., Yan, Y., Yu, H. (2012). Genome-wide identification of SOC1 and SVP targets during the floral transition in Arabidopsis. Plant J. **70,** 549-561.

Tapia-Lopez, R., Garcia-Ponce, B., Dubrovsky, J.G., Garay-Arroyo, A., Perez-Ruiz, R.V., Kim, S.H., Acevedo, F., Pelaz, S., Alvarez-Buylla, E.R. (2008). An AGAMOUS-related MADS-box gene, *XAL1* (*AGL12*), regulates root meristem cell proliferation and flowering transition in *Arabidopsis*. Plant Physiol. **146,** 1182-1192.

Teeri, T.H., Uimari, A., Kotilainen, M., Laitinen, R., Help, H., Elomaa, P., Albert, V.A. (2006). Reproductive meristem fates in *Gerbera*. J. Exp. Bot. **57,** 3445-3455.

The-Arabidopsis-Genome-Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408,** 796-815.

Theissen, G. (2001). Development of floral organ identity: stories from the MADS house. Curr. Opin. Plant Biol. **4,** 75-85.

Theissen, G., Saedler, H. (2001). Plant biology. Floral quartets. Nature **409,** 469-471.

Thelen, J.J., Peck, S.C. (2007). Quantitative proteomics in plants: choices in abundance. Plant Cell **19,** 3339-3346.

Thelen, J.J. (2009). Proteomics tools and resources for investigating protein allergens in oilseeds. Regul. Toxicol. Pharmacol. **54,** S41-S45.

Thompson, B.E., Bartling, L., Whipple, C., Hall, D.H., Sakai, H., Schmidt, R., Hake, S. (2009). *bearded-ear* encodes a MADS box transcription factor critical for maize floral development. Plant Cell **21,** 2578-2590.

Thompson, R., Burstin, J., Gallardo, K. (2009). Post-genomics studies of developmental processes in legume seeds. Plant Physiol. **151,** 1023-1029.

Torii, K.U. (2004). Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. Int. Rev. Cytol. **234,** 1-46.

Torti, S., Fornara, F., Vincent, C., Andres, F., Nordstrom, K., Gobel, U., Knoll, D., Schoof, H., Coupland, G. (2012). Analysis of the *Arabidopsis* shoot meristem transcriptome during floral transition dentifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. Plant Cell **24,** 444-462.

Truernit, E., Siemering, K.R., Hodge, S., Grbic, V., Haseloff, J. (2006). A map of KNAT gene expression in the Arabidopsis root. Plant Mol. Biol. **60,** 1-20.

Truernit, E., Haseloff, J. (2007). A Role for KNAT Class II Genes in Root Development. Plant Signal. Behav. **2,** 10-12.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., Colot, V. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genet. **3,** e86.

Turgeon, R., Wolf, S. (2009). Phloem transport: cellular pathways and molecular trafficking. Annu. Rev. Plant Biol. **60,** 207-221.

Uimari, A., Kotilainen, M., Elomaa, P., Yu, D., Albert, V.A., Teeri, T.H. (2004). Integration of reproductive meristem fates by a *SEPALLATA*-like MADS-box gene. Proc. Natl. Acad. Sci. USA **101,** 15817-15822.

Umezawa, T., Sugiyama, N., Mizoguchi, M., Hayashi, S., Myouga, F., Yamaguchi-Shinozaki, K., Ishihama, Y., Hirayama, T., Shinozaki, K. (2009). Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **106,** 17588-17593.

Unte, U.S., Sorensen, A.M., Pesaresi, P., Gandikota, M., Leister, D., Saedler, H., Huijser, P. (2003). SPL8, an SBP-box gene that affects pollen sac development in Arabidopsis. Plant Cell **15,** 1009-1019.

Urbanus, S.L., de Folter, S., Shchennikova, A.V., Kaufmann, K., Immink, R.G., Angenent, G.C. (2009). In planta localisation patterns of MADS domain proteins during floral development in *Arabidopsis thaliana*. BMC Plant Biol. **9,** 5.

Urbanus, S.L., Martinelli, A.P., Dinh, Q.D., Aizza, L.C., Dornelas, M.C., Angenent, G.C., Immink, R.G. (2010). Intercellular transport of epidermis-expressed MADS domain transcription factors and their effect on plant morphology and floral transition. Plant J. **63,** 60-72.

van Dijk, A.D., Morabito, G., Fiers, M., van Ham, R.C., Angenent, G.C., Immink, R.G. (2010). Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction. PLoS Comput Biol **6,** e1001017.

Van Leene, J., Stals, H., Eeckhout, D., Persiau, G., De Slijke, E.V., Van Isterdael, G., De Clercq, A., Bonnet, E., Laukens, K., Remmerie, N., Henderickx, K., De Vijlder, T., Abdelkrim, A., Pharazyn, A., Van Onckelen, H., Inze, D., Witters, E., De Jaeger, G. (2007). A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*. Mol. Cell. Proteomics **6,** 1226-1238.

Van Leene, J., Witters, E., Inze, D., De Jaeger, G. (2008). Boosting tandem affinity purification of plant protein complexes. Trends Plant Sci. **13,** 517-520.

Van Leene, J., Hollunder, J., Eeckhout, D., Persiau, G., Van De Slijke, E., Stals, H., Van Isterdael, G., Verkest, A., Neirynck, S., Buffel, Y., De Bodt, S., Maere, S., Laukens, K., Pharazyn, A., Ferreira, P.C., Eloy, N., Renne, C., Meyer, C., Faure, J.D., Steinbrenner, J., Beynon, J., Larkin, J.C., Van de Peer, Y., Hilson, P., Kuiper, M., De Veylder, L., Van Onckelen, H., Inze, D., Witters, E., De Jaeger, G. (2010). Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis thaliana*. Mol. Syst. Biol. **6,** 397.

van Mourik, S., van Dijk, A.D., de Gee, M., Immink, R.G., Kaufmann, K., Angenent, G.C., van Ham, R.C., Molenaar, J. (2010). Continuous-time modeling of cell fate determination in *Arabidopsis* flowers. BMC Syst. Biol. **4,** 101.

Van Tunen, A.J., Angenent, G.C. (1993). Floral organogenesis in Tulipa. Flow. Newsl. **16,** 33-38.

Vandenbussche, M., Theissen, G., Van de Peer, Y., Gerats, T. (2003a). Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. Nucleic Acids Res. **31,** 4401-4409.

Vandenbussche, M., Zethof, J., Souer, E., Koes, R., Tornielli, G.B., Pezzotti, M., Ferrario, S., Angenent, G.C., Gerats, T. (2003b). Toward the analysis of the petunia MADS box gene family by reverse and forward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require *SEPALLATA*-like MADS box genes in petunia. Plant Cell **15,** 2680-2693.

Vekemans, D., Viaene, T., Caris, P., Geuten, K. (2012). Transference of function shapes organ identity in the dove tree inflorescence. New Phytol. **193,** 216-228.

Verelst, W., Saedler, H., Munster, T. (2007a). MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters. Plant Physiol. **143,** 447-460.

Verelst, W., Twell, D., de Folter, S., Immink, R., Saedler, H., Munster, T. (2007b). MADS-complexes regulate transcriptome dynamics during pollen maturation. Genome Biol. **8,** R249.

Vierstra, R.D. (2003). The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. Trends Plant Sci. **8,** 135-142.

Vierstra, R.D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. Nat. Rev. Mol. Cell Biol. **10,** 385-397.

Villar, C.B., Erilova, A., Makarevich, G., Trosch, R., Kohler, C. (2009). Control of *PHERES1* imprinting in *Arabidopsis* by direct tandem repeats. Mol. Plant **2,** 654-660.

Viswanathan, S., Unlu, M., Minden, J.S. (2006). Two-dimensional difference gel electrophoresis. Nat. Protoc. **1,** 1351-1358.

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W., Giovannoni, J. (2002). A MADS-box gene necessary for fruit ripening at the tomato *ripening-inhibitor* (*rin*) locus. Science **296,** 343-346.

Vrebalov, J., Pan, I.L., Arroyo, A.J., McQuinn, R., Chung, M., Poole, M., Rose, J., Seymour, G., Grandillo, S., Giovannoni, J., Irish, V.F. (2009). Fleshy fruit expansion and ripening are regulated by the Tomato *SHATTERPROOF* gene *TAGL1*. Plant Cell **21,** 3041-3062.

Wagner, D., Meyerowitz, E.M. (2002). SPLAYED, a novel SWI/SNF ATPase homolog, controls reproductive development in *Arabidopsis*. Curr. Biol. **12,** 85-94.

Wagner, D. (2009). Flower morphogenesis: timing is key. Dev. Cell **16,** 621-622.

Walia, H., Josefsson, C., Dilkes, B., Kirkbride, R., Harada, J., Comai, L. (2009). Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. Curr. Biol. **19,** 1128-1132.

Walter, M., Chaban, C., Schutze, K., Batistic, O., Weckermann, K., Nake, C., Blazevic, D., Grefen, C., Schumacher, K., Oecking, C., Harter, K., Kudla, J. (2004). Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. Plant J. **40,** 428-438.

Walz, C., Giavalisco, P., Schad, M., Juenger, M., Klose, J., Kehr, J. (2004). Proteomics of curcurbit phloem exudate reveals a network of defence proteins. Phytochemistry **65,** 1795-1804.

Wang, G.D., Fiers, M. (2010). CLE peptide signaling during plant development. Protoplasma **240,** 33-43.

Wang, R., Farrona, S., Vincent, C., Joecker, A., Schoof, H., Turck, F., Alonso-Blanco, C., Coupland, G., Albani, M.C. (2009). *PEP1* regulates perennial flowering in *Arabis alpina*. Nature **459,** 423-427.

Wang, Y., Liu, C., Yang, D., Yu, H., Liou, Y.C. (2010). *Pin1At* encoding a peptidyl-prolyl *cis/trans* isomerase regulates flowering time in *Arabidopsis*. Mol. Cell **37,** 112-122.

Wang, Y.Q., Melzer, R., Theissen, G. (2010). Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon* provide a clue to the evolutionary origin of 'floral quartets'. Plant J. **64,** 177-190.

Wapinski, I., Pfeffer, A., Friedman, N., Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. Nature **449,** 54-61.

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F., Meyerowitz, E.M. (1992). LEAFY controls floral meristem identity in Arabidopsis. Cell **69,** 843-859.

Wellmer, F., Alves-Ferreira, M., Dubois, A., Riechmann, J.L., Meyerowitz, E.M. (2006). Genome-wide analysis of gene expression during early *Arabidopsis* flower development. PLoS Genet. **2,** 1012-1024.

West, A.G., Shore, P., Sharrocks, A.D. (1997). DNA binding by MADS-box transcription factors: a molecular mechanism for differential DNA bending. Mol. Cell. Biol. **17,** 2876-2887.

West, A.G., Causier, B.E., Davies, B., Sharrocks, A.D. (1998). DNA binding and dimerisation determinants of *Antirrhinum majus* MADS-box transcription factors. Nucleic Acids Res. **26,** 5277-5287.

Wienkoop, S., Weckwerth, W. (2006). Relative and absolute quantitative shotgun proteomics: targeting low-abundance proteins in *Arabidopsis thaliana*. J. Exp. Bot. **57,** 1529-1535.

Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., Williams, K.L. (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol. Genet. Eng. Rev. **13,** 19-50.

Winter, C.M., Austin, R.S., Blanvillain-Baufume, S., Reback, M.A., Monniaux, M., Wu, M.F., Sang, Y., Yamaguchi, A., Yamaguchi, N., Parker, J.E., Parcy, F., Jensen, S.T., Li, H., Wagner, D. (2011). LEAFY target genes reveal floral regulatory logic, *cis* motifs, and a link to biotic stimulus response. Dev. Cell **20,** 430-443.

Winter, K.U., Weiser, C., Kaufmann, K., Bohne, A., Kirchner, C., Kanno, A., Saedler, H., Theissen, G. (2002). Evolution of class B floral homeotic proteins: obligate heterodimerization originated from homodimerization. Mol. Biol. Evol. **19,** 587-596.

Witte, C.P., Noel, L.D., Gielbert, J., Parker, J.E., Romeis, T. (2004). Rapid one-step protein purification from plant material using the eight-amino acid StrepII epitope. Plant Mol. Biol. **55,** 135-147.

Wollmann, H., Mica, E., Todesco, M., Long, J.A., Weigel, D. (2010). On reconciling the interactions between *APETALA2*, miR172 and *AGAMOUS* with the ABC model of flower development. Development **137,** 3633-3642.

Wu, M.F., Sang, Y., Bezhani, S., Yamaguchi, N., Han, S.K., Li, Z., Su, Y., Slewinski, T.L., Wagner, D. (2012). SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors. Proc. Natl. Acad. Sci. USA **109,** 3576-3581.

Wu, X., Dinneny, J.R., Crawford, K.M., Rhee, Y., Citovsky, V., Zambryski, P.C., Weigel, D. (2003). Modes of intercellular transcription factor movement in the *Arabidopsis* apex. Development **130,** 3735-3745.

Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenfuhrer, J., von Mering, C., Grossniklaus, U. (2010). *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. Curr. Biol. **20,** 506-512.

Wuest, S.E., O'Maoileidigh, D.S., Rae, L., Kwasniewska, K., Raganelli, A., Hanczaryk, K., Lohan, A.J., Loftus, B., Graciet, E., Wellmer, F. (2012). Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. Proc. Natl. Acad. Sci. USA **109,** 13452-13457.

Xu, C.P., Huang, B.R. (2008). Root proteomic responses to heat stress in two *Agrostis* grass species contrasting in heat tolerance. J. Exp. Bot. **59,** 4183-4194.

Yamaguchi, Y., Huffaker, A., Bryan, A.C., Tax, F.E., Ryan, C.A. (2010). PEPR2 is a second receptor for the Pep1 and Pep2 peptides and contributes to defense responses in *Arabidopsis*. Plant Cell **22,** 508-522.

Yang, S.L., Xie, L.F., Mao, H.Z., Puah, C.S., Yang, W.C., Jiang, L., Sundaresan, V., Ye, D. (2003). *TAPETUM DETERMINANT1* is required for cell specialization in the *Arabidopsis* anther. Plant Cell **15,** 2792-2804.

Yang, Y., Fanning, L., Jack, T. (2003). The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. Plant J. **33,** 47-59.

Yang, Y., Jack, T. (2004). Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. Plant Mol. Biol. **55,** 45-59.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., Meyerowitz, E.M. (1990). The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature **346,** 35-39.

Yin, L.L., Xue, H.W. (2012). The *MADS29* transcription factor regulates the degradation of the nucellus and the nucellar projection during rice seed development. Plant Cell **24,** 1049-1065.

Yoo, S.K., Lee, J.S., Ahn, J.H. (2006). Overexpression of *AGAMOUS-LIKE 28* (*AGL28*) promotes flowering by upregulating expression of floral promoters within the autonomous pathway. Biochem. Biophys. Res. Commun. **348,** 929-936.

Yoo, S.K., Wu, X.L., Lee, J.S., Ahn, J.H. (2011). *AGAMOUS-LIKE 6* is a floral promoter that negatively regulates the *FLC/MAF* clade genes and positively regulates *FT* in *Arabidopsis*. Plant J. **65,** 62-76.

Yoshida, H., Nagato, Y. (2011). Flower development in rice. J. Exp. Bot. **62,** 4719-4730.

Yu, H., Ito, T., Wellmer, F., Meyerowitz, E.M. (2004). Repression of AGAMOUS-LIKE 24 is a crucial step in promoting flower development. Nat. Genet. **36,** 157-161.

Zacher, B., Kuan, P.F., Tresch, A. (2010). Starr: Simple Tiling ARRay analysis of Affymetrix ChIP-chip data. BMC Bioinformatics **11,** 194.

Zahn, L.M., Kong, H., Leebens-Mack, J.H., Kim, S., Soltis, P.S., Landherr, L.L., Soltis, D.E., Depamphilis, C.W., Ma, H. (2005). The evolution of the *SEPALLATA*

subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. Genetics **169,** 2209-2223.

Zahn, L.M., Ma, X., Altman, N.S., Zhang, Q., Wall, P.K., Tian, D., Gibas, C.J., Gharaibeh, R., Leebens-Mack, J.H., Depamphilis, C.W., Ma, H. (2010). Comparative transcriptomics among floral organs of the basal eudicot *Eschscholzia californica* as reference for floral evolutionary developmental studies. Genome Biol. **11,** R101.

Zhang, H.M., Forde, B.G. (1998). An *Arabidopsis* MADS box gene that controls nutrient-induced changes in root architecture. Science **279,** 407-409.

Zhang, J.Z., Li, Z.M., Mei, L., Yao, J.L., Hu, C.G. (2009). *PtFLC* homolog from trifoliate orange (*Poncirus trifoliata*) is regulated by alternative splicing and experiences seasonal fluctuation in expression level. Planta **229,** 847-859.

Zhang, W., Zhang, T., Wu, Y., Jiang, J. (2012). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. Plant Cell **24,** 2719-2731.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., Jacobsen, S.E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. PLoS Biol. **5,** e129.

Zhao, Y., Wei, T., Yin, K.Q., Chen, Z., Gu, H., Qu, L.J., Qin, G. (2012). Arabidopsis RAP2.2 plays an important role in plant resistance to *Botrytis cinerea* and ethylene responses. New Phytol. **195,** 450-460.

Zheng, Y., Ren, N., Wang, H., Stromberg, A.J., Perry, S.E. (2009). Global identification of targets of the *Arabidopsis* MADS domain protein AGAMOUS-Like15. Plant Cell **21,** 2563-2577.

Zhong, J.P., Haynes, P.A., Zhang, S.P., Yang, X.P., Andon, N.L., Eckert, D., Yates, J.R., Wang, X., Budworth, P. (2003). Development of a system for the study of protein-protein interactions in planta: characterization of a TATA-box binding protein complex in *Oryza sativa*. J. Proteome Res. **2,** 514-522.

Zhou, M., Veenstra, T.D. (2007). Proteomic analysis of protein complexes. Proteomics **7,** 2688-2697.

Zhu, J.M., Chen, S.X., Alvarez, S., Asirvatham, V.S., Schachtman, D.P., Wu, Y.J., Sharp, R.E. (2006). Cell wall proteome in the maize primary root elongation zone. I. Extraction and identification of water-soluble and lightly ionically bound proteins. Plant Physiol. **140,** 311-325.

Zhu, W., Smith, J.W., Huang, C.M. (2010). Mass spectrometry-based label-free quantitative proteomics. J. Biomed. Biotechnol. **2010,** 840518.

Zobell, O., Faigl, W., Saedler, H., Munster, T. (2010). MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. Mol. Biol. Evol. **27,** 1201-1211.

Samenvatting

Eiwit-eiwit en eiwit-DNA interacties zijn essentieel voor de moleculaire werking van transcriptiefactoren. Door in verschillende combinaties te binden aan promoters van targetgenen kunnen transcriptiefactoren de expressie van deze genen verhogen of verlagen. MADS-domein eiwitten vormen een grote familie van transcriptiefactoren die aanwezig is in alle eukaryoten. In planten, en vooral in zaadplanten, is deze familie enorm uitgebreid. Er zijn bijvoorbeeld meer dan 100 familieleden in het genoom van *Arabidopsis*. MADS-domein eiwitten zijn in eerste instantie bekend vanwege hun belangrijke rol in bloemontwikkeling, maar uit verdere functionele karakterisering blijkt dat ze betrokken zijn bij bijna alle ontwikkelingsprocessen gedurende de levenscyclus van de plant. Hoe MADS-domein eiwitten specifieke bindingsplaatsen op het DNA herkennen is niet bekend. Het doel van dit proefschrift was om de moleculaire mechanismen van het functioneren van MADS-domein eiwitten in *Arabidopsis* te karakteriseren.

Hoofdstuk 1 geeft een uitvoerige beschrijving van de functies van MADS-domein transcriptiefactoren in bloeiende planten, waarbij de focus ligt op *Arabidopsis*. De belangrijkste subfamilies van MADS-domein eiwitten worden geïntroduceerd en hun basisstructuur wordt beschreven. Daarnaast wordt getoond dat verscheidene subfamilies van MADS-box genen kunnen worden onderscheiden op basis van fylogenetische verwantschap. Vervolgens worden de MADS-box genfuncties globaal beschreven gedurende verschillende ontwikkelingsstadia van *Arabidopsis*, van gametofyt-, embryo-, en zaadontwikkeling, via de vegetatieve fase tot bloem- en vruchtontwikkeling. Hierbij wordt de nadruk gelegd op een aantal recente bevindingen. De functies van bijna de helft van de *Arabidopsis* MADS-box genen zijn beschreven. Vele van deze functies zijn geconserveerd in andere planten soorten, maar sommige MADS-domein genen hebben ook een andere functie dan hun *Arabidopsis* homoloog. Dit weerspiegelt de verschillende aspecten van MADS-box gen functie-evolutie, die hebben bijgedragen tot de huidige morfologische diversiteit. Tenslotte wordt informatie over MADS-domein eiwit-eiwit en eiwit-DNA interacties geïntegreerd tot een model over MADS-domein eiwit activiteit. Dit model beschrijft hogere-orde eiwit complexen en actieve chromatine modulaties als onderdeel van het transcriptie proces.

Hoofdstuk 2 beschrijft MADS-domein eiwitcomplexen die potentieel gevormd worden gedurende de bloemontwikkeling in *Arabidopsis*. Door het toepassen van een

gerichte proteomics aanpak zijn we in staat geweest om het eiwit interactoom van een aantal belangrijke homeotische MADS-domein eiwitten (APETALA1, APETALA3, PISTILLATA, AGAMOUS, SEPALLATA3 en FRUITFULL) afkomstig uit plantmateriaal te karakteriseren, waarmee we interacties zoals voorgesteld in het 'floral quartet' model hebben kunnen bevestigen. Daarnaast hebben we transcriptiefactoren van andere families en chromatine-geassocieerde eiwitten als mogelijke interactiepartners van MADS-domain eiwitten ontdekt.

De meest prominente basale transcriptie-coregulators die met MADS-domain eiwitten interacteren zijn LEUNIG en SEUSS. Chromatine-geassocieerde eiwitten gevonden in MADS-domein eiwitcomplexen omvatten SWI/SNF ATP-afhankelijke nucleosoom moduleerders zoals BRAHMA en SPLAYED, en de histon 3 lysine 27 demethylase RELATIVE OF EARLY FLOWERING 6. Deze interacties verduidelijken de gecombineerde werking van MADS-domain transcriptiefactoren en suggereren dat ze werken door het opvangen en doorsturen van de chromatine moduleer-machine om de expressie van hun targetgenen te controleren.

In **hoofdstuk 3** worden de nieuwste methoden in het Proteomics gebied beschreven met een focus op signaal- en ontwikkelingsprocessen. De verbeterde technieken om proteoom-breed maar ook subcellulair de eiwitten te bestuderen, worden behandeld. Voorbeelden worden genoemd van componenten van volledige signaal routes in planten, startend van receptor-ligand interacties, gevolgd door fosforyleringscascades. Ook worden proteomics studies aan transcriptieregulatie-eiwitten belicht. Voor signaaloverdracht routes zijn eiwitmodificaties, zoals fosforylatie en ubiquitinylatie essentieel en daar zijn tegenwoordig ook vernieuwde methoden voor beschikbaar.

In **hoofdstuk 4** worden twee voorbeelden beschreven van biochemische procedures die gebruikt worden om complexen van membraangebonden receptoren en transcriptieregulatoren van celkernen te identificeren. In onze geoptimaliseerde methode maken we gebruik van een fluorophore-getagde éénstaps affiniteitsopzuivering van eiwitcomplexen, en ongelabelde massaspectrometrie-gebaseerde eiwitkwantificatie om ware complexpartners van niet-specifiek geprecipiteerde eiwitten te onderscheiden. In het kort worden de voor- en nadelen van verschillende methoden gebaseerd op immunoprecipitatie in vergelijking met onze methode beschreven. Het gedetailleerde protocol dat wordt beschreven in dit hoofdstuk is gebruikt om de MADS-domain eiwitcomplexen te karakteriseren, die worden beschreven in **Hoofdstuk 2**.

Hoe MADS domein eiwitten specifieke DNA volgorden herkennen en binden is nog onbekend. Interessant is om te weten of verschillende dimeercombinaties van MADS domeineiwitten andere affiniteiten hebben voor bepaalde DNA volgorden. Deze vragen worden behandeld in **Hoofdstuk 5**. Gebruik wordt gemaakt van

'systematic evolution of ligands by exponential enrichment' (SELEX) gevolgd door grootschalig sequencen (seq) om de verschillende DNA-affiniteiten MADS domein homo- en heterodimeren te onderscheiden. Door gebruik te maken van deze methode zijn we er in geslaagd om verschillen in affiniteit voor CArG varianten waar te nemen. De CArG-box is het DNA motief waar de MADS domeineiwitten aan binden en bestaat uit de consensus: $CC[A/T]_6GG$. Met SELEX is aangetoond dat er ook buiten de CArG-box nog basen van belang zijn. Vervolgens zijn de *in vitro* SELEX data vergeleken met *in vivo* ChIP-seq data om de relevantie na te gaan en uitspraken te doen welke dimeren binden aan bepaalde bindingsplaatsen.

In **hoofdstuk 6** tenslotte, was het doel om de moleculaire kenmerken van DNA-binding specificiteit van verschillende MADS-domein transcriptiefactoren te identificeren. Met behulp van bioinformatica en *in vitro* DNA-binding assays bleek dat de structuur van het DNA een belangrijke rol speelt in binding van MADS-domein eiwitten aan DNA. Een motief genaamd 'A-tract' speelt een centrale rol. A-tracts zijn onderdeel van de CArG-box, het motief dat herkend wordt door MADS-domein eiwitten. De resultaten suggereren dat de kromming van het DNA, wat beïnvloed wordt door verschillen in lengte van de A-tract en periodieke verdeling, een rol speelt in het bepalen van de DNA-binding specificiteit van de verschillende MADS-domein eiwit dimeren.

Alles bij elkaar, heeft het onderzoek dat in dit proefschrift beschreven wordt onze kennis over de moleculaire mechanismen van de werking van MADS-domein transcriptie factoren in planten uitgebreid. **Hoofdstuk 7** sluit dit proefschrift af en geeft de toekomst perspectieven voor onderzoek aan MADS-domein eiwitten. Eruit gelicht worden de voordelen van high-throughput (proteomics en genomics) technologieën, die niet alleen gebruikt kunnen worden om statische karakteristieken van transcriptionele regulatie te bestuderen, maar ook kunnen helpen om de dynamische en stoichiometrische veranderingen van complexe eiwit- en genregulatie netwerken in plantontwikkeling te ontrafelen.

Summary

Protein-protein and protein-DNA interactions are essential for the molecular action of transcription factors. By combinatorial binding to target gene promoters, transcription factors are able to up- or down-regulate the expression of these genes. MADS-domain proteins comprise a large family of transcription factors present in all eukaryotes. In plants, and especially in seed plants, this family has significantly expanded. For example, more than 100 representatives are found in the *Arabidopsis* genome. MADS-box genes have initially been shown to play major roles in flower development, however their emerging functional characterization revealed functions in almost all developmental processes throughout the plant life cycle. How MADS-domain transcription factors acquire their functional specificity remains unresolved. The goal of this thesis was to characterize some of the molecular mechanisms by which MADS-domain proteins act in *Arabidopsis*.

Chapter 1 comprehensively reviews functions of MADS-domain transcription factors in flowering plants, with a main focus on *Arabidopsis*. Major classes of MADS-domain proteins are introduced, and their modular structures are described. Additionally, it is shown that several distinctive subfamilies of MADS-box genes can be inferred from the phylogenetic analysis of the whole gene family. Next, we broadly describe MADS-box gene functions in developmental stages of *Arabidopsis*, from gametophyte, embryo and seed development, via sporophytic phase transitions to flower and fruit development, highlighting some of the recent findings. In *Arabidopsis*, functions for nearly half of the MADS-box gene family members have already been described. Many MADS-box gene functions are conserved in other species, but some deviate from their homologs in *Arabidopsis*, illustrating various aspects of MADS-box gene functional evolution which may contribute to morphological diversification. Finally, by compiling recent studies on MADS-domain protein-protein and protein-DNA interactions, we present a hypothetical model of MADS-domain protein action that combines higher-order protein complex formation and active chromatin remodeling by large transcriptional machineries.

Chapter 2 describes MADS-domain protein complexes that are potentially formed during *Arabidopsis* flower development. By using a targeted proteomics approach we were able to characterize the protein interactome of major floral homeotic MADS-domain proteins (APETALA1, APETALA3, PISTILLATA, AGAMOUS, SEPALLATA3 and FRUITFULL) in native plant tissues, confirming

interactions suggested in the 'floral quartet' model. Additionally, we discovered transcription factors from other families and chromatin-associated proteins as possible interaction partners of MADS-domain proteins. The most prominent general transcriptional coregulators that interact with MADS-domain proteins are LEUNIG and SEUSS. Chromatin-associated proteins found in MADS-domain protein complexes include SWI/SNF ATP-dependent nucleosome remodelers such as BRAHMA and SPLAYED, as well as the histone 3 lysine 27 demethylase RELATIVE OF EARLY FLOWERING 6. These interactions shed light on the combinatorial modes of action of MADS-domain transcription factors and suggest that they can act by recruiting or redirecting the chromatin remodeling machinery to control the expression of their target genes.

In **Chapter 3** we review recent advances in proteomics approaches used to study cellular signaling and developmental processes in plants. We mention the emerging tools for of whole plant proteome characterization as well as subcellular protein localization. The major focus, though, is on the description of complete cellular signaling cascades in plants, starting from the characterization of signaling mobile molecules (e.g. peptide or protein), through identification of receptors and receptor protein complexes, ending with identification of intermediate signaling pathway members. In addition, we highlight proteomics studies on transcriptional regulators, which is very well related to the study presented in this thesis. We present ways to study post-translational modifications, such as phosphorylation and ubiquitinylation, which are essential to understand signal transduction cascades in plants. Furthermore, quantitative proteomics methods used to identify new components in signaling pathways as well as to characterize the composition of protein complexes were presented.

Two examples of biochemical procedures used to identify complexes of membrane-bound receptors and transcriptional regulators from nuclei are described in **Chapter 4**. In our optimized method we make use of fluorophore-tagged single step affinity purification of protein complexes and label-free mass spectrometry-based protein quantification to distinguish true complex partners from non-specifically precipitated proteins. We briefly describe advantages and disadvantages of various immunoprecipitation-based protocols in comparison to our method. The detailed protocol presented in this chapter was used to characterize MADS-domain protein complexes described in **Chapter 2**.

The exact molecular mechanisms of DNA sequence recognition by MADS-domain transcription factors are still unknown. Particularly intriguing is the question whether various MADS-domain protein complexes possess different DNA-binding specificities. We address this question in **Chapter 5**. We used systematic evolution of ligands by exponential enrichment (SELEX) followed by high-throughput sequencing

(seq) approach to discriminate DNA-binding specificities of several MADS-domain protein homo- and heterodimers. Using this strategy, we were able to distinguish between different protein complexes based on their affinity to particular CArG-box variants. CArG-boxes are the MADS-domain binding motifs with the DNA consensus sequence of $CC[A/T]_6GG$. Additionally, by comparing the *in vitro* binding characteristics of various MADS-domain protein dimers with the available *in vivo* DNA-binding data we could identify complex-specific binding events at genome wide scale in *Arabidopsis*.

Finally, in **Chapter 6**, we aimed to identify the molecular features of different DNA-binding specificities of MADS-domain transcription factors. With help of bioinformatics tools and *in vitro* DNA-binding assays we found that structural characteristics of the DNA play an important role in DNA-binding of MADS-domain proteins. A central role has a motif called 'A-tract'. A-tracts are part of the CArG-box, the DNA binding motif of MADS-domain proteins. Our results suggest that the curvature of the DNA, which can be modulated by different A-tract length and their periodic distribution, may play a role in determining the DNA-binding specificity of different MADS-domain protein dimers.

Taken together, research described in this thesis advances our knowledge on the molecular mechanisms of MADS-domain transcription factor action in plants. **Chapter 7** concludes the thesis and describes future perspectives in MADS-domain protein research. Highlighted are the advances of high-throughput (proteomics and genomics) technologies that could be used to unravel not only the static characteristics of transcriptional regulation but also the dynamic and stoichiometric changes of complex protein and gene regulatory networks during plant development.

Streszczenie

Oddziaływania białko-białko i białko-DNA są niezbędne w prawidłowym funkcjonowaniu czynników transkrypcyjnych na poziomie molekularnym. Kombinatoryczne oddziaływania czynników transkrypcyjnych z promotorami specyficznych genów, pozwalają na aktywacje, bądź też hamowanie ekspresji tychże genów. Białka z rodziny MADS stanowią liczną rodzinę czynników transkrypcyjnych szeroko rozpowszechnionych u Eukaryota, a szczególnie u roślin nasiennych. Dla przykładu, u rzodkiewnika pospolitego (*Arabidopsis thaliana*), organizmu modelowego w badaniach biologii molekularnej, możemy wyróżnić ponad 100 genów kodujących białka MADS. Początkowa analiza funkcjonalna genów MADS wskazała ich główny udział w procesie rozwoju kwiatów. Jednakże, w okresie 20 lat badań nad genami MADS, ujawniono ich funkcje w prawie każdym procesie rozwojowym u roślin. Pytanie, jak czynniki MADS nabyły swoje charakterystyczne funkcje, pozostaje nierozwiązane. Celem tej dysertacji było zbadanie niektórych mechanizmów molekularnych wykorzystywanych przez białka MADS w regulacji ekspresji genów u *Arabidopsis*, które przyczyniają się do pełnionych przez nie funkcji.

W rozdziale pierwszym szczegółowo zrecenzowano funkcje czynników transkrypcyjnych z domeną MADS u roślin okrytonasiennych, a zwłaszcza u *Arabidopsis*. Opisano tutaj podział rodziny białek MADS na szereg podrodzin w oparciu o analizę filogenetyczną oraz wyróżniono główne klasy białek MADS wraz z ich schematem budowy. Następnie, szeroko opisano funkcje jakie czynniki MADS pełnią w cyklu rozwojowym rzodkiewnika, począwszy od rozwoju zarodka i nasiona, poprzez fazy wzrostu wegetatywnego, po kwitnienie i rozwój owocu. Do tej pory udało się scharakteryzować funkcje prawie połowy genów z rodziny białek MADS. Wiele funkcji jakie pełnią czynniki MADS w dużym stopniu zostały zachowane w ewolucji i nie różnią się znacznie u różnych gatunków roślin. Jednakże, niektóre z funkcji homologów białek MADS odbiegają od tych jakie pełnią one u *Arabidopsis*, ilustrując ewolucję funkcjonalną czynników MADS, która mogła przyczynić się do różnorodności w budowie morfologicznej roślin. Pod koniec rozdziału, w oparciu on najnowsze publikacje naukowe, przedstawiony został hipotetyczny model pracy białek MADS, który łączy formowanie się skomplikowanych kompleksów białkowych z aktywnym remodelingiem chromatyny przez maszynerie transkrypcyjne.

W rozdziale drugim scharakteryzowano kompleksy molekularne białek MADS, tworzące się w czasie rozwoju kwiatu u *Arabidopsis*. Wykorzystując metody

proteomiczne do badań kompleksów białkowych, udało się opisać interaktom (komplet oddziaływań białko-białko) głównych białek MADS (APETALA1, APETALA3, PISTILLATA, AGAMOUS, SEPALLATA3, i FRUITFULL) w izolowanych tkankach roślinnych *in vivo*. Opisane kompleksy białkowe potwierdziły oddziaływania molekularne zaproponowane w sławnym modelu 'kwiatowy kwartet' w 2001 roku. Ponadto, okazało się, że czynniki transkrypcyjne z innych rodzin białkowych oraz białka aktywnie remodelujące chromatynę także, potencjalnie, oddziałują z białkami MADS. Najważniejszymi współregulatorami białek MADS opisanymi tutaj były: ogólne czynniki regulujące transkrypcję genów LEUNIG i SEUSS; białka wchodzące w skład kompleksów białkowych remodelujących chromatynę z wykorzystaniem ATP, np. BRAHMA i SPLAYED; oraz białka wprowadzające modyfikacje histonów, np. demetylaza RELATIVE OF EARLY FLOWERING 6. Oddziaływania opisane w tym rozdziale rzucają nowe światło na molekularne mechanizmy kontrolujące regulację ekspresji genów przez czynniki MADS.

W rozdziale trzecim przedstawiono przegląd najnowszych osiągnięć w metodach proteomicznych stosowanych w badaniach sygnalizacji komórkowej oraz procesów rozwojowych u roślin. Podkreślono tutaj ważną rolę rozwoju metod molekularnych do badań całego proteomu roślinnego, a także metod wewnątrzkomórkowej lokalizacji białek. Główny nacisk położono na opisanie badań nad ścieżkami przekazywania sygnału wewnątrz- oraz zewnątrzkomórkowego, począwszy od charakteryzacji cząsteczek sygnalizacyjnych tzw. ligandów (np. białek lub peptydów), poprzez identyfikacje ich receptorów oraz kompleksów receptorowych, po identyfikacje elementów pośrednich. Dodatkowo, opisano tutaj badania modyfikacji potranslacyjnych białek, np. fosforylacji czy ubikwitynacji, które są niezbędne do pełnego zrozumienia kaskad sygnalizacyjnych w komórce roślinnej, oraz ilościowe metody proteomiczne wykorzystywane do identyficaji nowych elementów ścieżek przekazywania sygnału i do jakościowej charakteryzacji kompleksów białkowych.

W rozdziale czwartym opisano dwa przykłady metod biochemicznych do identyfikacji kompleksów czynników transkrypcyjnych oraz kompleksów receptorowych związanych z błoną komórkową. W obydwu metodach wykorzystano immunoprecipitację białek fluoryzujących połączonych z białkiem 'przynętą' oraz ilościową kwantyfikację białek 'ofiar' metodami spektrometrii masowej, w celu wyróżnienia rzeczywistych oddziaływań białko-białko. Pokrótce przedstawiono także zalety i wady różnych metod opartych o immunoprecypitację białek. Szczegółowy protokół laboratoryjny opisany w tym rozdziale wykorzystany został do scharakteryzowania kompleksów białkowych białek z domeną MADS opisanych w rozdziale drugim.

Szczegółowy mechanizm działania jaki czynniki MADS używają do rozpoznawania docelowej sekwencji DNA jest nadal nieznany. Szczególnie intrygujące jest pytanie, czy różne kompleksy czynników transkrypcyjnych MADS posiadają odmienną specyfikę wiązania DNA. Na to pytanie starano się odpowiedzieć w rozdziale piątym, gdzie wykorzystano metodę wysoko-wydajnego sekwencjonowania DNA (seq) poprzedzonego systematyczną ewolucją ligandów przez powielanie eksponencjalne (SELEX) w celu zbadania specyfiki wiązania się do DNA białkowych kompleksów z rodziny MADS. Wykorzystując tą strategię, udało się rozróżnić poszczególne kompleksy białkowe, w oparciu o ich powinowactwo do różnych wariantów sekwencji DNA, tzw. sekwencji CArG. Dodatkowo, porównując dane *in vivo* oddziaływań białko-DNA z danymi *in vitro* opartymi o metodę SELEX, precyzyjnie zidentyfikowano miejsca w genomie *Arabidopsis*, które wiązane są przez dimery czynników MADS.

W rozdziale szóstym starano się zidentyfikować czynniki molekularne, które mogą leżeć u podstaw odmiennej specyfiki wiązania DNA przez kompleksy białek z domeną MADS. Przy pomocy technik bioinformatyki wspartymi metodami laboratoryjnymi, wykazano, że cechy strukturalne docelowego fragmentu DNA wiązanego przez białka MADS odgrywają bardzo ważną rolę w tym procesie. Wyniki przedstawione w tym rozdziale sugerują, że krzywizna DNA, regulowana przez długość fragmentu AT w sekwencji CArG oraz jego okresowa dystrybucja w genomie, przyczynia się do modyfikacji specyfiki wiązania białek MADS do DNA.

Reasumując, badania opisane w tej dysertacji poszerzają naszą wiedzę na temat mechanizmów molekularnych w jaki czynniki MADS regulują transkrypcję genów docelowych w roślinach. Rozdział siódmy podsumowuje dysertację i przedstawia perspektywy przyszłych badań nad białkami MADS.

Acknowledgments

In these few words, I would like to express my full gratitude to all of the people who were scientifically and morally involved in the conception of this thesis.

I would like to thank Prof. Gerco Angenent, my thesis promoter and supervisor, for giving me the opportunity to perform my PhD research at the Plant Developmental Systems cluster. Your door was always open and you were always available for the scientific discussions and advices. You were a great support to my research, always questioning the most important parts of the chapters, especially the non-clear results or shaky conclusions. Thank you for your time spent on corrections of this thesis and very useful comments. Besides research, I learned many professional aspects of supervision and problem-solving capabilities from you. I also enjoyed all the lab-uitjes, where we could focus on other things than science.

I would most like to thank Dr. Kerstin Kaufmann, my thesis co-promoter and daily supervisor. As far as I know, I was your first full-time PhD student, and even if it sometimes felt a little bit like being a lab rabbit (for me), I can fairly say that you were the most thoughtful, helpful and understanding supervisor that I ever had. Thank you for your patience in the lab at the beginning of my internship, brainstorming discussions about EMSAs binding patterns, precise suggestions about the IP protocol and that 'little' push, which motivates me to work even harder. I am extremely grateful for your support in revising, advising on every chapter of this book and all the tiniest corrections, it would look a whole lot different without it. It was a great pleasure to work with you and I hope that our fruitful collaboration will continue in the future.

Many thanks go to Dr. Jose Muiño, the one and only bioinformatician who helped me a lot with making sense of the SELEX-seq data as well as with statistical analysis of the proteomics experiments; Dr. Richard Immink, who was actively involved in the first and the major part of this thesis; Sjef Boeren, the master of the mass spec, from whom I got a lot of input about technical aspects of proteomics sample preparation, analysis and data processing; Marco (aka Macro) Busscher for beautiful confocal pictures of my GFP plants and Jacqueline Busscher for successful yeast matting that confirmed some of my precious interactions of MADS-domain proteins.

Special thanks go to all my colleagues at the PDS cluster, with whom I shared my daily scientific and non-scientific life, for creating a perfect working environment. Thanks to Sela (Dr. Danisman), you were there when I arrived at the PDS as a rookie

student and, although it took some time for me to relax and start talking (I think the time when we were dressed as women standing alone in the woods broke the ice... Eee…, never mind...), we soon became good friends and neighbors in and outside the lab. Although, I would not take your advices on treating plants, the protoplasts work was going very smoothly with your instructions. And then there were girls, many girls… Whoever says about gender balance in science is wrong in case of the PhD-student section of the PDS cluster. Special thanks go to: Anneke (aka Hanneke), my direct office neighbor, on whom I could always count when looking for a scientific advice and discussion. Jenny, my fellow climber, we always aimed high (not only in climbing) and we never fell down (almost); thanks for all the adrenaline and emotional chit-chats. Alice, my dear paranymph, I still have some small problems with understanding your accent and extremely direct e-mails (but after all, maybe the least words the better); I promise, we will do something with this SEP3ΔCt and it will be the greatest Flower Power research. Violeta (+1), my in-front neighbor, thanks for understanding my sense of humor and the Rakia evenings. Hui, thanks for being 'occasionally' in the lab on Saturdays, my bacteria really appreciated not being grown for 72 h. I would also like to thank the new batch of my PhD fellows. Suzanne, by observing you I am constantly motivated to work harder on my assertiveness and persuasiveness. Hilda, my apprentice, thanks for being patience with me and for all the small-talks... talks... talks... (by the way, my name is pronounced: /ˈsmɑːtʃnɪɑːk/). Leonie, my beer buddy, I enjoyed a lot our EMSA discussions and double-beer during Fridays' borrels.

I would like to thank all the past and present members of the BU Bioscience at the PRI for all useful discussions during my yearly presentations, help and advices. Special thanks go to all members of the PDS cluster. Here, I would like to mention: Andrea B, Froukje, Guodong, Elio, Jan C, Jan K, Jannie, Jeroen P, Jeroen v A, Kim, Leo, Martijn (my snack buddy), Merche, Michiel (the EBS orders master), Mieke, Ronny, Rumyana, Ruud (thanks for accepting me for the internship, afterwards, it all begun then), Steven, Susan, Tjitske, Tom (Alejandro, Alejandro, Ale..., Ale...), Twan and Wilco.

Wageningen is a distinct place, attracting people from all around the world, including myself. It is also a great place to experience a mixture of different cultures, which I liked a lot. Special thanks I address here to all the international guest students and guest workers, many of whom became my good friends: Erica, Clarissa, Mari Carmen, Cédric, Giusy (I hope you are doing well in Rome), Quy (Peter), Huong, Martina, Michela, Andrea O, Pedro, Rosella, Juan, Francisco, Asmini, Ali, Tetty, Ania S (I know that you are struggling in Poland with you PhD but keep going and you will make it!), Priscilla and Camilla (we should go for a run one day in Brazil), Lukas and Juliana (I am missing you and your cheerfulness. I am working on my caipira

Portuguese and I hope I will test it someday; I have a good teacher...), Lilian and Livia (your Caipirinhas are great), Diego, Anna-Catharina and Verónica. I hope we will stay in touch as much as possible in the future.

Furthermore, I would like to thank my student that I had an opportunity to supervise in longer term, Giuseppe (aka Pepe). I learned a lot from this experience and I hope you did the same. Moreover, I would like to acknowledge here all the people with whom I had occasion to directly collaborate on various experimental procedures: Wei, Na, Ya-Fen, Mareike, Soenita and Maria. This allowed me to get a perspective on the research carried out in other laboratories. I would also like to thank my scientific collaborators, especially members of Dr. François Parcy team in Grenoble (France), members of Dr. Lin Xu team in Shanghai (China), members of Dr. Markus Schmid team in Tübingen (Germany), and members of Prof. Sacco de Vries team in Wageningen.

Als mijn Nederlands is niet zo goed, I sincerely thank Anneke, Hilda, Leonie, Suzanne and Gerco for translating the summary to Dutch. I am sure that you did a perfect job, although I am not able to check it.

I would like to thank my 'old' good friends spread around Europe and the 'Polish community' that I met here, during my stay in Wageningen, for all the spiritual support, awesome weekends and native talks that allowed me to separate myself from science a little bit: Ania and Łukasz (Ziom-North Holland), Aga and Wojtek (Ziom-Cambridge), Marta and Robert (Ziom-Edinburgh), Anka and Żuraw (Ziom-Kraków), Paweł and Stempel (we should go hiking soon), Agnieszka, Ania P, Ania S, Ania U, Dorota, Gosia, Kasia, Ola P, Monika, Grzesiek and Robert Cz (all the best in Gdańsk). I am also very grateful to Kristin and Marina, my neighbours in vein, who supported me during the culture shock, especially when I saw my first accommodation in Wageningen, which was not very different from the Harry Potter's cellar.

Pleasant evenings, relaxing parties and delicious dinners wouldn't be the same without Ana Carolina, Maria Cecília, Bart, Charles, Delyan, Felipe, Jordi, Julio, Luigi, Pádraic, and many more. Thank you for your company.

Special thanks to all members of the Wageningen Student Alpine Club 'IBEX', especially to my fellow climbers from the 6+ training and all the participants of the climbing weekends. I am grateful to all IBEX climbing instructors that allowed me to develop my passion for climbing without letting me fall hard.

Last but not the least; I would like to thank my dear Aline. We met almost one year before I handled my thesis and you supported me all the way from that time. Thank you for your understanding to my 'I need to finish my Thesis', I will try to make it up to you. Beijo!

Chciałbym serdecznie podziękować mojej całej rodzinie, a w szczególności mojej mamie Krystynie, tacie Feliksowi oraz bratu Jarkowi (wraz z żoną Danką oraz córką Gosią, która urodziła się w dniu w którym te podziękowania były pisane). Dziękuję za trud i wysiłek włożony w moje wychowanie oraz wsparcie w każdym okresie mojego życia, a w szczególności w czasie moich studiów uniwersyteckich w Krakowie oraz doktoranckich w Wageningen.

Thank you!

Cezary Smaczniak

Wageningen, 25 November 2012

Curriculum vitae

Cezary Dominik Smaczniak was born in Opole, Poland, on 26th December 1983 as a second son of Krystyna and Feliks. His fascination to natural sciences was first recognized in the primary school where he showed a strong interest for chemistry. Soon after that, he was accepted to the General High School No 3 in Opole, Poland, in the class with biology and chemistry as his major subjects. His outstanding performance in the 48th National Chemistry Competition in Warsaw, Poland, granted him an access to almost every Polish university. He slowly switched his interests towards bio- subjects and started his biotechnological studies in one of the oldest European universities, the Jagiellonian University in Cracow, Poland, in 2002. Towards the end of his studies he went abroad for a half-year internship at the Laboratory of Molecular Biophysics in Orléans, France, where he studied the stability of protein-DNA interactions upon radiation in bacteria. He graduated with the Master's degree in Biotechnology with specialization in Biochemistry in 2007. Soon after, he started his internship at the Plant Research International in Wageningen, The Netherlands, in Prof. Gerco Angenent group, Plant Developmental Systems, under supervision of Dr. Kerstin Kaufmann. The work with plants as model species was completely new to him. However, it soon became apparent to him that protein-protein and protein-DNA interaction studies – the hardcore interactomics – can as well be performed in plants as they are performed in other organisms. He started his PhD studies in 2008 in the same group, focusing on the molecular mechanisms controlling flower organ specification in the model species *Arabidopsis thaliana*. He handled in his doctoral thesis in 2012. He will continue the scientific research as a post-doc in the field of molecular biology and biochemistry.

In his free time he is usually hanging on the wall of a climbing gym or, during sunny days, on the rocks of Belgium, France and Germany. He also enjoys playing squash and watching a good movie.

Education Statement of the Graduate School
Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

Issued to: Cezary Smaczniak
Date: 16 January 2013
Group: Plant Developmental Systems, Wageningen University & Research Centre

1) Start-up phase	date
► **First presentation of your project**	
DNA binding by MADS box transcription complexes	Sep 27, 2008
► **Writing or rewriting a project proposal**	
► **Writing a review or book chapter**	
► **MSc courses**	
► **Laboratory use of isotopes**	
Subtotal Start-up Phase	*1,5 credits**

2) Scientific Exposure	date
► **EPS PhD student days**	
PhD student days 2009 (Leiden, The Netherlands)	Feb 26, 2009
2nd International Retreat of PhD Students in Plant Sciences (Cologne, Germany)	Apr 15-17, 2010
PhD student days 2011 (Wageningen, The Netherlands)	May 20, 2011
► **EPS theme symposia**	
2009 Theme 1 'Developmental Biology of Plants', Leiden University	Jan 30, 2009
2010 Theme 1 'Developmental Biology of Plants', Wageningen University	Jan 28, 2010
2011 Theme 1 'Developmental Biology of Plants', Leiden University	Jan 20, 2011
2012 Theme 1 'Developmental Biology of Plants', Wageningen University	Jan 19, 2012
► **NWO Lunteren days and other National Platforms**	
NWO-ALW meeting 'Experimental Plant Sciences' 2009 (Lunteren, The Netherlands)	Apr 06-07, 2009
NWO-ALW meeting 'Experimental Plant Sciences' 2010 (Lunteren, The Netherlands)	Apr 19-20, 2010
NWO-ALW meeting 'Experimental Plant Sciences' 2011 (Lunteren, The Netherlands)	Apr 04-05, 2011
NWO-ALW meeting 'Experimental Plant Sciences' 2012 (Lunteren, The Netherlands)	Apr 02-03, 2012
Netherlands Bioinformatics Conference 2012 (NBIC 2012) (Lunteren, The Netherlands)	Apr 24, 2012
NPC Progress Meeting 2009 (Utrecht, The Netherlands)	Jan 26, 2009
NPC Progress Meeting 2010 (Utrecht, The Netherlands)	Feb18, 2010
NPC Progress Meeting 2011 (Utrecht, The Netherlands)	Jan 13, 2011
NPC Progress Meeting 2012 (Utrecht, The Netherlands)	Feb 07, 2012
CBSG Meeting 2012, Proteomics Hotel Projects (Wageningen, The Netherlands)	Dec 14, 2011
8th Dutch Chromatin meeting 2010 (Leiden, The Netherlands)	Oct 26, 2010
► **Seminars (series), workshops and symposia**	
BU Bioscience Seminars	2008-2012
Kathryn S. Lilley, Mapping plant organelle proteomes using stable isotope labeling.	Nov 02, 2009
Daniel Schubert, Dynamic control of histone methylation and plant cell fate by Polycomb-group proteins.	May11, 2010
David Baulcombe, Mobile RNA silencing in plants	Sep 27, 2010
Peter Cook, Transcription factories as organizers of the genome: the role of fixed polymerases	Oct 27, 2010
Symposium 'Advances in Life Science Technology'	Nov 25, 2010
Seminar Series Plant Sciences: Chairs Plant Physiology and Molecular Biology	Sep 08, 2009
Seminar Series Plant Sciences: Chairs Mathematical & Statistical Methods and Phytopathology	Nov 10, 2009
Seminar Series Plant Sciences: Chairs Mathematics and Molecular Biology (Bioscience)	Feb 09, 2010
► **Seminar plus**	
► **International symposia and congresses**	
Workshop on Mechanisms Controlling Flower Development (Aiguablava, Spain)	Jun 08-12, 2009
COST action: Plant proteomics in Europe - Systems biology and Omic approaches (Namur, Belgium)	May 05-07, 2010
Workshop on Molecular Mechanisms Controlling Flower Development (Maratea, Italy)	June 13-17, 2011
Plant development and environmental interactions, EMBO Conference (Matera, Italy)	May 27-30, 2012
► **Presentations (highly recommended)**	
In vitro characterization of MADS-box transcription factor complexes (Aiguablava, Spain) - poster	Jun 08-12, 2009
In vitro and in vivo characterization of MADS-box transcription factor complexes (Brixen, Italy) - poster	Aug 02-08, 2009
Characterization of MADS-box transcription factor complexes (Utrecht, The Netherlands) - poster	Feb 16, 2010
In vitro characterization of MADS-box transcription factor complexes (Cologne, Germany) - talk	Apr 15-17, 2010
Characterization of MADS-box transcription factor complexes in Arabidopsis flower development - talk	Dec 14, 2011
Characterization of MADS-box transcription factor complexes in Arabidopsis flower development - talk	Jan 19, 2012
BU Bioscience Presentation (Wageningen, The Netherlands) - talk	Oct 13, 2009
BU Bioscience Presentation (Wageningen, The Netherlands) - talk	Sep 21, 2010
BU Bioscience Presentation (Wageningen, The Netherlands) - talk	Sep 20, 2011
BU Bioscience Presentation (Wageningen, The Netherlands) - talk	Sep18, 2012
► **IAB interview**	Feb 18, 2011
► **Excursions**	
Subtotal Scientific Exposure	*26,3 credits**

continued on next page

3) In-Depth Studies	*date*
► **EPS courses or other PhD courses**	
Modifications of Proteins“ (Brixen, Italy)	Aug 02-08, 2009
Systems Biology: Statistical Analysis of ~Omics Data (Wageningen, The Netherlands)	Dec 13-17, 2010
Transcription Factors and Transcription Regulation (Leiden, The Netherlands)	May 09-11, 2011
► **Journal club**	
Journal club of the PRI-PDS cluster	2008-2012
► **Individual research training**	
Subtotal In-Depth Studies	*7,2 credits**

4) Personal development	*date*
► **Skill training courses**	
Scientific Writing	Sep 06 - Nov 04, 2010
Career Perspectives	Sep 18 - Nov 15, 2011
► **Organisation of PhD students day, course or conference**	
► **Membership of Board, Committee or PhD council**	
Subtotal Personal Development	*3,4 credits**

TOTAL NUMBER OF CREDIT POINTS*	**38.4**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

** A credit represents a normative study load of 28 hours of study.*

Financial support from Wageningen University for printing this Thesis is gratefully acknowledged.

Printed and bound by GVO drukkers & vormgevers B.V. / Ponsen & Looijen, Ede

Thesis layout and cover: Cezary Smaczniak
Fonts used: Arial, Adobe Garamond Pro and Myriad Pro