The (r)evolution of gene regulatory networks controlling Arabidopsis plant reproduction: a two-decade history

Alice Pajoro1,2,*, Sandra Biewers3,*, Evangelia Dougal4,5,*, Felipe Leal Valentim1,*, Marta Adelina Mendes6,*, Aimone Porr7,*, George Coupland7, Yves Van de Peer4,5,8, Aalt D. J. van Dijk1,9, Lucia Colombo4,5,8, Brendan Davies3 and Gerco C. Angenent1,2,†

1 Plant Research International (PRI) Droevendaalseweg 1, 6708 PB Wageningen, The Netherlands
2 Laboratory of Molecular Biology, Wageningen University, Droevendaalseweg 1, 6708 PB Wageningen, The Netherlands
3 Centre for Plant Sciences, University of Leeds, Leeds LS2 9JT, UK
4 Department of Plant Systems Biology, VIB, Technologiepark 927, 9052 Ghent, Belgium
5 Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052 Ghent, Belgium
6 Dipartimento di BioScienze, Università degli Studi di Milano, Via Celoria 26, 20133, Milan, Italy
7 Max Planck Institute for Plant Breeding Research, Carl von Linne Weg 10, D-50829 Cologne, Germany
8 Genomics Research Institute (GRI), University of Pretoria, Private bag X20, Pretoria, 0028, South Africa
9 Biometris, Wageningen University, Droevendaalseweg 1, 6708 PB Wageningen, The Netherlands

* These authors contributed equally to the work.
† To whom correspondence should be addressed. E-mail: gerco.angenent@wur.nl

Received 16 March 2014; Revised 16 March 2014; Accepted 28 April 2014

Abstract

Successful plant reproduction relies on the perfect orchestration of singular processes that culminate in the product of reproduction: the seed. The floral transition, floral organ development, and fertilization are well-studied processes and the genetic regulation of the various steps is being increasingly unveiled. Initially, based predominantly on genetic studies, the regulatory pathways were considered to be linear, but recent genome-wide analyses, using high-throughput technologies, have begun to reveal a different scenario. Complex gene regulatory networks underlie these processes, including transcription factors, microRNAs, movable factors, hormones, and chromatin-modifying proteins. Here we review recent progress in understanding the networks that control the major steps in plant reproduction, showing how new advances in experimental and computational technologies have been instrumental. As these recent discoveries were obtained using the model species Arabidopsis thaliana, we will restrict this review to regulatory networks in this important model species. However, more fragmentary information obtained from other species reveals that both the developmental processes and the underlying regulatory networks are largely conserved, making this review also of interest to those studying other plant species.

Key words: Flowering time, floral organ development, gene regulation, molecular interactions, networks, transcription factors.

Plant sexual reproduction

Angiosperms such as Arabidopsis thaliana have lifecycles characterized by the alternation of diploid sporophytic and haploid gametophytic stages. Before the sporophyte produces the spores that develop into male and female gametophytes, the plant undergoes several developmental changes that can be considered as phase transitions. The initial vegetative stage can be subdivided into juvenile and adult phases. Plants in the juvenile phase are able to produce leaves and axillary buds, whereas the initiation of reproductive structures only occurs in the adult vegetative phase. The next phase transition is the switch from vegetative to reproductive growth, when the vegetative shoot apical meristem (SAM) acquires the identity of an inflorescence...
meristem (IM) that will then produce floral meristems (FM). Subsequently, floral organs are produced according to a canonical pattern with sepals, petals, stamens, and pistils being formed in sequence, from the periphery to the centre of the floral meristem, to make whorls of organs. Male reproductive development takes place in the third whorl, stamens, which are composed of anthers supported on filaments. Inside the anthers the male gametophytes develop by two sequential processes: microsporogenesis and microgametogenesis. Ovules, which form in the fourth whorl, provide structural support to the female gametes and enclose them until seed development, which follows fertilization. The development of the ovule can also be divided in two steps: the specification of the functional megaspore (megasporeogenesis) and the formation of the embryo sac (megagametogenesis) (Reiser and Fischer, 1993; Shi and Yang, 2011). Sexual reproduction requires delivery of the sperm nuclei, via the pollen, to the embryo sac, where a double fertilization occurs and the new diploid sporophyte is formed.

**Gene regulatory networks controlling reproduction**

Developmental processes are controlled by tightly coordinated networks of regulators, known as gene regulatory networks (GRNs) that activate and repress gene expression within a spatial and temporal context. In *Arabidopsis thaliana*, the key components of the GRNs controlling major processes in plant reproduction, such as the floral transition and floral organ identity specification, were first identified in loss of function mutants that affect these processes (Blazquez et al., 2006). The interactions between these regulators later began to be revealed through genetic analyses, resulting in the first, mostly linear, GRN maps. These were augmented by reverse genetics, analysis of protein–protein interactions and expression studies in wild-type and mutant plants, resulting in a hierarchical GRN in which master regulators target a subset of genes directing downstream processes (Blazquez et al., 2006) (Fig. 1). Most of these master regulator genes encode transcription factors (TFs), often members of the type II MADS-box family of TFs.

Foreshadowing the complexity that is now becoming very apparent, the early GRN studies already identified several examples of redundancy, where phenotypic consequences that are obvious in mutant combinations, are weak or absent in the corresponding single mutants. For example, individual mutants affecting any of the four SEPALLATA genes (*SEPI–4*) have no, or only subtle effects on flowers, whereas the indeterminate flowers of the *sep1sep2sep3* triple mutant are comprised only of sepals (Pelaz et al., 2000) and *sep1sep2sep3sep4* quadruple mutant flowers contain only multiple whorls of leaf-like organs (Ditta et al., 2004).

**GRN for floral transition**

Endogenous and environmental signals have their input into the decision to initiate flowering. Several pathways, including photoperiod, autonomous, vernalization, and gibberellin (GA) (Baurle and Dean, 2006; Simpson and Dean, 2002) converge on a small set of central flowering regulators, including *CONSTANS* (*CO*) and *FLOWERING LOCUS C* (*FLC*), which antagonistically regulate flowering (Fig. 1) (Samach et al., 2000). *FLC* acts as a repressor of flowering and mediates the vernalization and autonomous pathways, whereas *CO* is a floral activator and mediates the photoperiodic pathway. Both genes together regulate the expression of the downstream floral pathway integrator genes *FLOWERING LOCUS T* (*FT*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), and *LEAFY* (*LFY*) (Blazquez and Weigel, 2000; Simpson and Dean, 2002). The gibberellin pathway influences the phase transition at the SAM by promoting the expression of *SOC1* and *LFY* (Blazquez et al., 1998; Moon et al., 2003) and also acts upstream of *FT* in the leaf, providing evidence for crosstalk between the photoperiod and GA pathways (Porri et al., 2012).

The *FT* protein stimulates flowering by moving from the leaf (where the light is perceived) to the shoot apical meristem (where inflorescence and floral meristems form). Genetic and molecular studies indicate that the *FT* protein comprises part of the inductive signal known as florigen, which promotes flowering in response to photoperiod (Corbesier et al., 2007; Jaeger and Wigge, 2007). *FT* interacts at the SAM with the bZIP TF *FLOWERING LOCUS D* (*FD*) to activate the expression of the floral meristem identity genes *APETALA1* (*AP1*) and *SOC1*, which in turn activate *LFY*, promoting the floral transition (Wu and Gallagher, 2012). In addition to *LFY*, other genes such as AGAMOUS LIKE 24 (*AGL24*) also contribute to the up-regulation of *AP1* (Grandi et al., 2012; Pastore et al., 2011).

The microRNAs miR156 and miR172 have been identified as important regulators of this developmental phase change. Members of the *SQUAMOSA PROMOTER BINDING PROTEIN LIKE* (*SPL*) TF family are targeted by miR156, whereas miR172 targets 6 *APETALA2-LIKE* (*AP2-like*) transcription factors (Wu et al., 2009). Levels of miR156 are high in the juvenile vegetative phase and decrease before onset of the adult vegetative phase, allowing the production of a subset of SPL proteins (*SPL9*, *SPL10*). These SPLs induce the expression of miR172 genes, leading to a gradual increase in miR172 levels during the adult vegetative phase (Aukerman and Sakai, 2003; Jung et al., 2007; Schwarz et al., 2008; Wu and Poethig, 2006; Wu et al., 2009). Increased levels of miR172, in turn, cause a down-regulation of AP2-like genes, which otherwise repress adult traits and flowering (Fig. 1) making the plant competent to flower (Wu et al., 2009).

**GRN for floral organ development**

After successful initiation of the floral meristems, *AP1* and *LFY* activate the organ identity genes (Fig. 1). The way in which this limited set of genes acts to direct formation of the correct floral organs at the appropriate place in the flower is described by the classical ABC model (for reviews see: Alvarez-Buylla et al., 2010; Causier et al., 2010; Immink et al., 2010; O’Maoileidigh et al., 2014). The A-class genes are the MADS-box TF-encoding *AP1* (Mandel and Yanofsky, 1995) and *APETALA2* (*AP2*), which encodes a member of the AP2/ERF (ethylene
responsive factor) TF family (Jofuku et al., 1994; Okamura et al., 1997). The B-class genes are APETALA3 (AP3) (Jack et al., 1992) and PISTILLATA (PI) (Goto and Meyerowitz, 1994), both of which encode MADS-box TFs. AGAMOUS (AG), another MADS-box gene, is the only member of the C-class (Yanofsky et al., 1990). The original ABC model was extended by the addition of D- and E-classes, which specify ovule identity (Pinyopich et al., 2003) and floral identity, respectively. The D-class genes include the MADS-box genes SEEDSTICK (STK) and SHATTERPROOF1 and 2 (SHP1/2), which are redundantly required for the specification of ovule identity (Pinyopich et al., 2003). Four SEPALATA genes (SEP1–4), all of which encode MADS-box TFs, comprise the E-class (Ditta et al., 2004; Pelaz et al., 2000; Rounsley et al., 1995) (Fig. 1). Protein–protein interaction studies demonstrated that the ABCDE MADS-box TFs physically interact and bind DNA as tetramers, possibly establishing DNA loops in the promoters of their target genes to activate or repress their transcription (Egea-Cortines et al., 1999; Melzer and Theissen, 2009; Smaczniak et al., 2012b). The E-class SEP proteins are necessary for the formation of higher order complexes involving the A-, B- and C-class TFs and hence mediate their organ identity functions (Honma and Goto, 2001; Pelaz et al., 2001). The resultant hierarchical GRN formed the basis of our understanding of flower development until the advent of new technologies described below.

**The modern version of a GRN**

Recently, genome-wide molecular approaches, such as protein assays, expression profiling and chromatin immunoprecipitation (ChIP) have connected the nodes in the GRNs and added novel regulators, interactors, and downstream targets (Hawkins et al., 2010). In contrast to the earlier genetic networks described above, more recent versions show molecular interactions that are independent of mutant phenotypes and are potentially more dynamic in time and space. The
molecular interactions that were added are, for example, protein–protein interactions, which allow combinatorial actions of multiple regulatory factors. In the case of TF complexes, their composition can be an important determinant of DNA binding specificity and affinity and thus interactions will affect target gene regulation. Another layer of molecular interactions recently added to the networks comprises transcription factor binding profiles, i.e. protein–DNA interactions. Large-scale expression analyses, preferably performed with inducible systems (Kaufmann et al., 2010c; Wellmer et al., 2006), add transcriptional activity to the connections (‘edges’) between nodes in the network. These large-scale data sets have substantially increased the network connectivity and revealed the shortcomings of the classical hierarchical networks. Current GRNs are composed of TFs regulating subsets of genes in a combinatorial fashion and contain multiple regulatory feedback loops, which blur the hierarchical structure (Kaufmann et al., 2010a).

TFs act as process integrators and connect other developmental processes as hubs in the network (Pose et al., 2012). The TF target gene analyses have revealed unexpected connections between processes previously considered to be unrelated, which could not have been predicted by classical genetic approaches. For example, target gene analysis of the SEP3 floral organ identity transcription factor revealed a connection with auxin signalling (Kaufmann et al., 2009). As another striking example, a recent study addressing the role of LFY revealed that in addition to its function in reproductive transition, it is also involved in pathogen responses (Winter et al., 2011).

DNA–protein interaction studies provided novel insights into the wiring of GRNs

In the 90s, binding of TFs to DNA sequences was studied by in vitro methods, such as electrophoretic mobility shift assays (EMSA) and yeast 1-hybrid studies. Schwarz-Sommer’s research group (Tröbner et al., 1992) showed that the Antirrhinum class B homeotic proteins DEFICIENS and GLOBOSA interact with each other and bind in vitro to their own promoters, thereby forming an autoregulatory loop. More such autoregulatory loops have been recently identified in GRNs, particularly associated with transcriptional regulation of MADS-box genes (Gomez-Mena et al., 2005). For instance, expression of the MADS-box gene SOC1 is controlled by the SOC1 protein, which forms heterodimers with API that suppress SOC1 expression following the successful transition to floral meristem identity (Imarkin et al., 2012).

SELEX (systematic evolution of ligands by exponential enrichment) is a powerful method to characterise TF binding sites and has been applied to factors in the flowering GRN. For example, Moyroud et al. (Moyroud et al., 2011) applied SELEX coupled to next generation sequencing (NGS) to determine the preferred binding sites of LFY.

The development of ChIP was a major breakthrough in the study of DNA–protein interactions, as it allowed the identification of in vivo physical interactions between a TF and its target DNA (Gomez-Mena et al., 2005; Wang et al., 2002). The subsequent introduction of genome-wide arrays (ChIP–ChIP) (Zheng et al., 2009) or large-scale sequencing (ChIP-seq) (Kaufmann et al., 2010b) led to the identification of thousands of target genes for specific TFs. In the past 5 years many genome-wide TF-DNA binding profiles have been generated for TFs involved in plant reproduction (Table 1). The high numbers of interactions that were detected in ChIP-seq experiments revealed a much higher network complexity than expected and demonstrated that these major regulators not only control another layer of regulators, but also genes encoding structural proteins, enzymes and signalling proteins (Fig. 2; Kaufmann et al., 2010a). The ChIP-seq data also confirmed that most TFs involved in plant reproduction bind to their own locus, reinforcing the concept that auto-regulatory loops are a common mechanism of regulation in GRNs. For example, identification of the direct targets of LFY (Winter et al., 2011) and API (Kaufmann et al., 2010c) revealed that LFY is able to promote AP1 transcription through direct regulation, and API binds to LFY to form a positive feedback loop, leading to a strong and rapid up-regulation of both genes.

Through these genome-wide studies, in particular in combination with transcriptome analysis (Kaufmann et al., 2010c; Wuest et al., 2012), it also became apparent that some TFs act as both activators and repressors. For example, at early stages AP1 represses genes controlling flowering time, whereas at later stages it mainly acts as an activator of floral homeotic genes (Kaufmann et al., 2010c; Pajoro et al., 2014). A similar observation was made for the B-class genes AP3 and PI, which activate genes involved in organogenesis and repress those required for carpel development (Wuest et al., 2012). It is likely that the composition of the TF complexes and their ability to recruit co-factors act together to determine the DNA-binding specificity and the mode of transcriptional action. Detailed analyses of the DNA regions bound by several members of MADS-box TF family using the motif-based sequence analysis tool MEME (Machanick and Bailey, 2011) revealed distinct sequence specificities for each MADS-box TF. Although the consensus binding site for MADS-box proteins, the so-called CArG box, was found at the centre of virtually all binding peaks obtained in ChIP experiments, certain sequence motifs within and flanking the CArG box were preferentially enriched. This corroborates the idea that the DNA sequence specificity is determined to a large extent by the composition of the TF complex. Therefore, protein–protein studies will be required to enable us to understand TF–DNA interaction specificity and hence why specific target genes are recognised by specific transcription factors. Recent large-scale protein–protein and proteomics studies have elucidated the composition of many MADS-box protein complexes involved in flowering time control and floral organ development (Imarkin et al., 2009; Smaczniak et al., 2012a). The next challenge in these studies will be to unravel the binding specificity of these complexes and how this influences the dynamic control of target gene regulation.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Family</th>
<th>Function</th>
<th>Approach</th>
<th>Antibody</th>
<th>Tissue</th>
<th>Most relevant targets</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGAMOUS</td>
<td>AG</td>
<td>MADS-box transcription factor</td>
<td>ChIP-seq</td>
<td>Anti GFP</td>
<td>Flower buds stage 5</td>
<td>AG, CRC, SHP2, SPL, JAG, SEP3, AP1, AP3, SHP1, SUP, HEC1, HEC2, VDD</td>
<td>O’Maoileidigh et al., 2013</td>
</tr>
<tr>
<td>AGAMOUS-LIKE 15</td>
<td>AGL15</td>
<td>MADS-box transcription factor</td>
<td>ChIP–ChIP</td>
<td>Anti AGL15</td>
<td>Embryonic culture tissue</td>
<td>FLC, SVP, LEC2, FUS3, AB3, IAA30</td>
<td>Zheng et al., 2009</td>
</tr>
<tr>
<td>APETALA1</td>
<td>AP1</td>
<td>MADS-box transcription factor</td>
<td>ChIP-seq</td>
<td>Anti AP1</td>
<td>Inflorescence meristem; flowers at stage 2, 4, and 8</td>
<td>AG, CRC, AG, JAG, SEP3, AP1, AP3, SHP1, SUP, HEC1, HEC2, VDD</td>
<td>Kaufmann et al., 2010; Pajoro et al., 2014</td>
</tr>
<tr>
<td>APETALA2</td>
<td>AP2</td>
<td>AP2-like family</td>
<td>ChIP-seq</td>
<td>Anti AP2</td>
<td>Inflorescences</td>
<td>AG, SOC1, SEP3, AP1, TOE3, AGL15, ETT, SHP1, SHP2, AGL44, TOE1, RGA-like 1, miR156, miR172</td>
<td>Yant et al., 2010</td>
</tr>
<tr>
<td>APETALA3</td>
<td>AP3</td>
<td>MADS-box transcription factor</td>
<td>ChIP-seq</td>
<td>Anti GFP</td>
<td>Flower buds stage 5</td>
<td>CRC, SEP3, SPL, AP1, SUP, AG, Ufo, SHP2, RBE, HEC1, HEC2, ALC</td>
<td>Wuest et al., 2012</td>
</tr>
<tr>
<td>FLOWERERING LOCUS C</td>
<td>FLC</td>
<td>MADS-box transcription factor</td>
<td>ChIP-seq</td>
<td>Anti FLC</td>
<td>12-day-old seedlings</td>
<td>SOC1, FT, SEP3, CBF1, JA26, AGL16, SPL15, DIN10, SVP, SPL3, SMZ, TOE3, TEM1, FRI, CIR1, FIO1, LCL1, COL1</td>
<td>Deng et al., 2011</td>
</tr>
<tr>
<td>FLOWERERING LOCUS M</td>
<td>FLM</td>
<td>MADS-box transcription factor</td>
<td>ChIP-seq</td>
<td>Anti GFP</td>
<td>15-day-old seedlings</td>
<td>SOC1, ATC, TEM2, SMZ, SEP3, AP3, Pi, RVE2, FIO1, SHP2, Mir156, AP2, Mir172, AP1</td>
<td>Pose et al., 2013</td>
</tr>
<tr>
<td>LEAFY</td>
<td>LFY</td>
<td>Flcal transition</td>
<td>ChIP-seq</td>
<td>Anti LFY</td>
<td>Inflorescences</td>
<td>TFL1, AP1, AG, SEP4, LFY, SOC1, PRS, BB, GIS, GOA, STY2, AP3, GAXD2</td>
<td>Moyroud et al., 2011</td>
</tr>
<tr>
<td>PISTILLATA</td>
<td>PI</td>
<td>MADS-box transcription factor</td>
<td>ChIP-seq</td>
<td>Anti GFP</td>
<td>Flower buds stage 5</td>
<td>CRC, SEP3, SPL, AP1, SUP, AG, Ufo, SHP2, RBE, HEC1, HEC2, ALC</td>
<td>Wuest et al., 2012</td>
</tr>
<tr>
<td>SCHLAFMUTZE</td>
<td>SMZ</td>
<td>AP2-like family</td>
<td>ChIP–ChIP</td>
<td>Anti GFP</td>
<td>Seedlings</td>
<td>FT, SMZ, SNZ, AP2, TOE3, SOC1, AP1, TEM1, FRI</td>
<td>Mathieu et al., 2009</td>
</tr>
<tr>
<td>SEPALLATA3</td>
<td>SEP3</td>
<td>MADS-box transcription factor</td>
<td>ChIP-seq</td>
<td>Anti SEP3</td>
<td>Inflorescences (stage1–12); flowers at stage 2, 4, and 8</td>
<td>AP1, AP3, SEP1, SEP2, AP4, AG, SHP1, SHP2, GA1, PIN4, PID, ETT, ARG8, IAA4</td>
<td>Kaufmann et al., 2009; Pajoro et al., 2014</td>
</tr>
<tr>
<td>SHORT VEGETATIVE PHASE</td>
<td>SVP</td>
<td>MADS-box transcription factor</td>
<td>ChIP-seq</td>
<td>Anti GFP</td>
<td>2 weeks old seedlings</td>
<td>GI, PRR7, FLK, FLD, CLF, SWN, VNR2, PHYA, STIP, SVP, CLV1, CLV2, PHB, PHV, REV, ATBB8</td>
<td>Gregis et al., 2013</td>
</tr>
<tr>
<td>SHORT VEGETATIVE PHASE</td>
<td>SVP</td>
<td>MADS-box transcription factor</td>
<td>ChIP-seq</td>
<td>Anti GFP</td>
<td>Inflorescences (stage1–11)</td>
<td>SVP, CLV1, PHB, KAN1, ETT, PIN1, WDR55</td>
<td>Gregis et al., 2013</td>
</tr>
<tr>
<td>SUPPRESSOR OF OVEREXPRESSION</td>
<td>SOC1</td>
<td>MADS-box transcription factor</td>
<td>ChIP-seq</td>
<td>Anti GFP</td>
<td>Transition apices</td>
<td>SOC1, CBF1, CBF2, CBF3, mir156, SVP, AG15, AG18, TEM2, TOE3, SMZ, SNZ, SEP3, SHP2, AP3, Pi, SUP, ETT</td>
<td>Immink et al., 2012</td>
</tr>
</tbody>
</table>
SOC1 and SHORT VEGETATIVE PHASE (SVP), two MADS domain proteins acting as hubs in the GRNs

As an example of the progress made regarding the topology and complexity of the GRNs described above, we will highlight some genome-wide analysis of regulators and targets of the MADS domain proteins SVP and SOC1. Owing to its important role as a floral integrator, SOC1 is of particular interest, because it may act as a hub between the GRNs involved in vegetative and reproductive meristems. In this respect, both upstream regulators and downstream targets of SOC1 are relevant. SVP has a dual function, reflected in its expression pattern: in vegetative tissues SVP acts as a floral suppressor, whereas during the first stages of floral meristem development SVP prevents the precocious expression of the floral organ identity genes (Gregis et al., 2006). Recently, the genome-wide SVP binding sites in both tissues were determined and the difference reflects the dual role of this transcription factor (Gregis et al., 2013).

A small set of target genes that act downstream of SOC1 were identified by a microarray expression analysis (Seo et al., 2009) and genome-wide analysis of binding sites based on a ChIP-ChIP experiment using a SOC1 overexpression line (Tao et al., 2012). More recently, a genome-wide target gene analysis was performed by ChIP-seq on enriched apical meristem tissues expressing GFP-tagged SOC1 protein under the endogenous promoter during floral transition (Immink et al., 2012). This study revealed that floral timing and flower organogenesis networks are heavily interconnected via SOC1 acting as a transition hub and being involved in many positive and negative auto-regulatory feedback loops. Several other MADS-box genes were also found to be bound by SOC1. The floral repressors SVP, AGL15, and AGL18 were identified as direct targets of SOC1. Furthermore, SOC1 also binds to the regulatory regions of several floral homeotic genes, such as the MADS-box genes SEP3, AP3, PI, and SHP2 (Immink et al., 2012). Genes encoding AP2/EREBP transcription factors that are involved in the response to low temperature,
such as CRT/DRE-BINDING FACTOR1 (CBF1), CBF2, and CBF3, were bound by SOC1. Seo and collaborators (Seo et al., 2009) demonstrated that SOC1 down-regulates these CBFs and the ChIP-seq data confirmed that this down-regulation is direct (Immink et al., 2012). SOC1 also regulates a number of other AP2-like genes acting predominantly as suppressors of flowering, such as TEMPRANILLO2 (TEM2), AP2, TARGET OF EAT3 (TOE3), SCHLAFMUTZE (SNZ), and SCHNARCHZAPPEN (SNZ). Except AP2, these factors suppress SOC1 expression either directly or indirectly via FT.

Because multiple (auto)regulatory feedback loops exist and because SOC1, as a floral integrator, receives many signals from the different flowering pathways, upstream regulators of SOC1 were also determined (Immink et al., 2012). A comprehensive matrix-based yeast one-hybrid assay was performed using 135 type II MADS domain protein dimers, available from an Arabidopsis MADS dimer collection in yeast (de Folter et al., 2005; Immink et al., 2009; Immink et al., 2012). Although this approach was targeted to only MADS domain proteins, it contributed to a more defined GRN around SOC1. Dimers consisting of proteins involved in floral timing and floral transition, such as SVP–AGL15 and FUL–SOC1, were found to bind to the SOC1 regulatory sequences. Remarkably, a large number of DNA–TF interactions were also found with dimers containing “ABC-class” homeotic proteins, for instance AG–SEP3, AGL24–AP1, and SOC1–SEP3 (Immink et al., 2012). Previous studies (Kaufmann et al., 2010c; Liu et al., 2008) showed that AP1 is involved in the repression of SOC1 in the floral meristem, whereas the yeast one-hybrid data confirmed the binding of AP1, in combination with particular dimerization partners, to the SOC1 promoter. These yeast data were confirmed by SOC1 reporter lines, which showed down-regulation of the reporter when either AP1, SEP3, or AG expression was induced (Immink et al., 2012).

SVP is a key regulator of two developmental phases: during the vegetative phase it represses the floral transition and later it contributes to the specification of floral meristems. To maintain plants in the vegetative phase, SVP represses the expression of FT and TWIN SISTER OF FT (TSF) in the phloem and SOC1 in the SAM by directly binding to CaRg boxes in FT and SOC1 promoter regions (Jang et al., 2009; Lee et al., 2007; Li et al., 2008). SVP interacts with FLC and the FLM-β splicing variant, both dimers binding to the SOC1 promoter as repressor, whereas the SVP/FLM-δ dimer, which is not able to bind DNA, competes with the SVP/FLM-β dimer and acts as a promoter of flowering (Pose et al., 2013). Another competitor of SVP is its phylogenetically closest related MADS-box gene AGL24, which is a central promoter of flowering (Michaels et al., 2003; Parenicova et al., 2003; Yu et al., 2002). Both SVP and AGL24 directly regulate SOC1 by binding to the same binding sites in the promoter but they have an opposite effect on SOC1 expression (Liu et al., 2008). During the floral transition, SVP expression gradually decreases in the IM and becomes expressed again at the early stages of flower development (Gregis et al., 2009).

Two recent studies (Gregis et al., 2013; Tao et al., 2012) reported the targets of SVP at a genome-wide scale. A comparison between targets in the SAM and after the floral transition in the FM revealed targets common and specific to its developmental roles (Gregis et al., 2013). SVP binds directly to flowering-time genes involved in the different flowering pathways including the circadian clock and photoperiodic pathway represented by GIGANTEA (GI) and PSEUDO-RESPONSE REGULATOR 7 (PRR7), the autonomous pathway represented by genes such as FLOWERING LATE KH MOTIF (FLK) and FLOWERING LOCUS D (FD), the light signalling pathway represented by PHYTOCHROME A (PHYA), as well as to genes that encode components of chromatin-associated complexes such as CURLY LEAF (CLF), SWINGER (SWN), and VERNALIZATION2 (VRN2). During the early stages of flower development (stages 1 and 2) SVP controls the expression of homeotic genes to maintain the floral meristem in an undifferentiated state (Gregis et al., 2006). This role is executed together with API and together this dimer recruits the SEU-LUG repressor complex. Comparison of regions bound by API and SVP revealed many common targets, such as CLAVATA1 (CLV1) and AG, which are down-regulated by SVP (Gregis et al., 2013). CLV1 and AG are both suppressors of WUSCHEL (WUS), which maintains the meristematic activity in the FM and controls the size of the FM. At late stage 2, when SVP expression is switched off, API interacts with SEP3 and the repression of the homeotic genes is removed.

Connecting GRNs and hormonal pathways in plant reproduction

The ChIP-seq data published recently demonstrate that MADS-domain TFs are able to bind to thousands of different genomic regions, several of which belong to genes involved in different hormonal pathways (Table 1). To understand the connection between MADS-domain TFs and the different hormone pathways, we analysed each ChIP-seq dataset available for these proteins and grouped them into specific hormone-related subclasses. Among the targets of these transcription factors involved in flowering time and floral development are genes involved in auxin, gibberellin, jasmonic acid, ethylene, brassinosteroid, salicylic acid and abscisic acid pathways. Different functional sub-categories can be recognized for each hormone pathway (Table 2).

SEP3, AP1, and AG are the TFs that score the highest number of sub-categories, meaning that they are involved in almost all hormone pathways. This information is in accordance with previous findings, where SEP3 was found to integrate and modulate different growth-related and hormone pathway genes, particularly those connected with auxin signalling (Kaufmann et al., 2009). API directly controls the expression of genes with known functions in the control of organ growth such as genes involved in the biosynthesis and response to gibberellin (GA) (Kaufmann et al., 2010c). A recent publication about AG direct targets illustrated that approximately half of the 225 identified genes encode
proteins with regulatory functions, including many transcription factors, (receptor) kinases, putative ligands, and proteins involved in different phytohormone signalling pathways (O’Maoileidigh et al., 2013).

Genes belonging to “response to abscisic acid (ABA) stimulus”, and “ABA-mediated signalling pathway” categories are bound by almost all TFs analysed. ABA is known to play important roles in many aspects of seed development, such as accumulation of storage compounds, acquisition of desiccation tolerance, induction of seed dormancy, and suppression of precocious germination (Kanno et al., 2010). Most likely this hormone plays unknown functions during flower development. GA and ABA also play important roles in regulating the floral transition. For instance, Shan and colleagues (Shan et al., 2012) showed that an increase in sensitivity to ABA delays flowering time. Gibberellins promote flowering in Arabidopsis through the activation of the floral integrator genes SOC1, LFY, and FT in the inflorescence and floral meristems, and in leaves, respectively.

Among the many roles of auxin is its function in the outgrowth and development of lateral organs, including floral organs, which has been suggested based on mutant phenotypes (Immink et al., 2012) and SEP3 targets (Kaufmann et al., 2009). ETTIN (ETT) imparts regional identity in the FM by affecting perianth organ number spacing, stamen formation, and regional differentiation of stamens and gynoecium (Sessions et al., 1997). The ChIP-seq analysis with SVP (Gregis et al., 2013) also identified target genes involved in the auxin signalling pathway, such as BIG, which encodes a putative auxin transporter required for normal auxin efflux
and inflorescence development (Gil et al., 2001; Yamaguchi et al., 2007).

These examples show the integration of the flowering and flower development GRNs in hormone signalling, although it is far from understood how these TFs control the hormone pathways and lead to the initiation and differentiation of the flower.

**Tissue-specific characterization of gene regulatory networks: Spatial action and short/long distance movement**

The current GRNs are still relatively static and lack detailed information about the spatial and temporal expression dynamics of genes that are part of the network. The networks active in the leaf and the SAM are relevant to the floral transition and spatial expression patterns of a number of key regulators are available for these tissues. Examples include members of the SPL gene family, which encode transcription factors that regulate flowering in *Arabidopsis* (Wang et al., 2009; Wu et al., 2009). The temporal and spatial expression patterns of SPL genes were studied using laser capture microdissection coupled to transcriptome sequencing and in situ hybridization (Cardon et al., 1999; Porri et al., 2012; Torti et al., 2012; Wang et al., 2009). These studies showed that several SPL genes, including SPL3, SPL4, SPL5, and SPL9 are expressed in both leaf and shoot apical meristems of *Arabidopsis*. SPL4 and SPL5 expression rises in the centre of the meristem of wild-type plants during the transition to flowering, when plants are exposed to inductive long-day conditions through the *FT* pathway (Torti et al., 2012). In contrast, SPL9 mRNA is detectable at the flanks of the inflorescence meristem (Porri et al., 2012; Wang et al., 2009). Recently it was shown that SPL proteins are regulated by interaction with the transcriptional repressor DELLA proteins that is degraded in response to GA, indicating that these transcription factors function as important integrators of distinct flowering pathways (Yu et al., 2012). In the presence of active GA, DELLAAs are ubiquitinated and targeted for degradation by the 26S proteasome (Dill et al., 2004; Willige et al., 2007). As DELLAAs bind *in vitro* and *in vivo* to several members of the SPL family, such as SPL2, SPL9 and SPL10, it was suggested that the presence of active GA is an essential step to produce functional SPL proteins by releasing them from DELLA repressors (Yu et al., 2012). Downstream events of *SPL* gene activity include the activation of *FT* transcription through the module miR172/AP2-like, a process that occurs specifically in the leaf (Fig. 3A). SPL9 and SPL10 directly activate transcription of *MIR172*, which produces a mature miRNA that cleaves, or prevents translation of, transcripts encoding AP2-like transcription factors (Mathieu et al., 2009). One of these factors is SMZ, which binds the *FT* locus 1.5 kb downstream of the coding region to repress its expression (Mathieu et al., 2009). Therefore, in the presence of active GA, SPL proteins lead to *FT* activation by reducing AP2-like transcript levels by promoting *MIR172* expression (Fig. 3A). At the shoot apical meristem *SPLs* transcription is also dependent on GA and GA signalling (Galvao et al., 2012; Porri et al., 2012). Transgenic plants expressing a GA catabolic enzyme specifically at the SAM, showed significant delay in *SPL* gene transcriptional activation when plants were exposed to inductive LD (Porri et al., 2012).

Downstream events of *SPL* genes include the activation of key flowering genes. SPL9 and SPL3 bind To SOC1 and LFY loci respectively, providing a direct mechanism by which SPLs act at the meristem to promote the switch from the vegetative to reproductive phase (Wang et al., 2009; Yamaguchi et al., 2009) (Fig. 3B). On the contrary, SOC1 was found to bind SPL3, SPL4, and SPL5 promoters, suggesting the presence of a positive feedback loop between *SOC1* and *SPLs* in flowering time control (Jung et al., 2012). Furthermore, SOC1 and the related transcription factor FRUITFUL (FUL) were shown genetically to be required for SPL4 activation during floral transition (Torti et al., 2012). Thus, *SPL* genes have different genetic and molecular functions during floral transition, and these are spatially separated between leaf and shoot apex (Fig. 3A, B).

Advances in understanding GRNs also provide information about the local and long-distance movement of hormones, proteins, or RNAs, which may be produced in specific cell types and move to other cells to regulate pathways. In this way, mobile factors can act as a hub between spatially separated GRNs. A striking example is *FT*, which encodes a small protein that functions as a master regulator of *Arabidopsis* flowering (Kardailsky et al., 1999; Kobayashi et al., 1999). *FT* was initially identified through a late flowering mutant (Koornneef et al., 1991). Genetic interaction between *FT* and other flowering mutants identified important components of the *FT* pathway (Koornneef et al., 1991). These include the upstream circadian clock component *GI* and the downstream acting gene *CO*, whose product activates directly *FT* transcription in response to long-day (Sawa and Kay, 2011; Schaffer et al., 1998; Suarez-Lopez et al., 2001) (Fig. 3A). More recently, studies revealed the exact tissues where *FT* is expressed. A *GUS* reporter gene fused to the *FT* promoter sequence showed GUS activity specifically in the companion cells of the leaf throughout the vasculature (Adrian et al., 2010; Takada and Goto, 2003). Although *FT* is expressed in the vasculature, it promotes flowering at the shoot apex, because the *FT* protein moves from companion cells to the apex to trigger flowering (Corbesier et al., 2007; Jaeger and Wigge, 2007; Mathieu et al., 2007). Thus, *FT* is a long-distance signalling molecule that connects flowering promoting signals in the leaves with the GRN that is active in the SAM (Fig. 3). In addition, *FT* acts redundantly with a closely related gene *TSF*, but whether *TSF* protein also travels from leaf to the SAM is not yet understood (Jung et al., 2009; Yamaguchi et al., 2005).

Plant hormones are mostly small organic molecules with the ability to move from cell to cell through specific transmembrane transporters. Auxin, cytokinin and GA were shown to affect flowering of *Arabidopsis* through interconnected genetic pathways (Richter et al., 2013). Although specific cell-cell transporters have not yet been identified for GA, recent experimental evidence suggests that this hormone moves among plant tissues. Grafting experiments performed...
in Arabidopsis, suggested that GA is the mobile signal that induces thickness of the hypocotyl by increasing the production of xylem tissue (Ragni et al., 2011). Xylem formation is associated with the transition to flowering and with increasing expression of the GIBBERELLIN 3-OXIDASE 1 (GA3ox1) (which encodes an enzyme involved in GA biosynthesis) in the SAM. This generates increased levels of GA that migrate from the shoot to the hypocotyl to induce xylem formation, as demonstrated by a grafting experiment, in which a Landsberg scion rescued the defective xylem expansion of the
GA mutant ga1-3 (Ragni et al., 2011). Long distance movement of active GA may occur from the leaf to the shoot apex. Labelled GA$_4$ applied to the leaf could be detected at the shoot apex of Arabidopsis (Eriksson et al., 2006), implying that the hormone is capable of long-range movement, presumably through the vasculature.

Recently, fluorescein labelled GA (Fl–GA) was used to trace GA distribution in the root system (Shani et al., 2013). Application of Fl–GA$_4$ in Arabidopsis roots showed specific GA accumulation patterns in the endodermal cells of the elongation zone. In addition, REPRESSION OF GA (REGA) marker (GFP: RGA), which encodes a DELLA protein that is degraded in the presence of active GA, showed decreasing signal in the endodermal cells, indicating that GA accumulates in these cells in wild-type plants (Shani et al., 2013). These new experimental approaches could be useful to further investigate a possible role of GA movement in flowering time control of Arabidopsis.

**Modelling plant reproduction GRNs**

An important aspect of understanding GRNs is how perturbations in one part of the network are transmitted to other parts of the network, and ultimately how these results in changes in flowering time. The modern version of Arabidopsis GRNs as presented above involves highly-connected, non-linear networks. Given this complexity, it is not possible to predict the effect of gene perturbations on e.g. flowering time in an intuitive way. Therefore, mathematical modelling plays an important role in providing a quantitative understanding of GRNs.

One of the pioneering models for cell-fate determination during the formation of floral organ primordia in Arabidopsis provided insights into the semi-quantitative relationships between the genes in the reproduction GRN (Espinosa-Soto et al., 2004). Based on published data, regulatory relationship between 15 ABC and key non-ABC genes were translated into a discrete Boolean network model. The state of each gene is updated according to the states of the genes that directly regulate it, via a set of logical rules derived from a survey of molecular genetic experimental data. Model simulations for all possible starting states showed that the network converges to a few steady states that correspond to expression patterns observed in each of the primordial cell types (i.e. inflorescence, sepals, petals, stamens, or carpels) and are in agreement with the phenotypes predicted by the ABC genetic model for both wild type and mutants. Analysis of the simulation results and the logical rules derived from published data led the authors to speculate that AG is involved in a positive feedback loop to maintain its own expression; this prediction was experimentally confirmed by a later study (Gomez-Mena et al., 2005).

A quantitative model that captures not only the regulatory relationships but also the kinetics of MADS domain complex formation was later proposed to represent a GRN for organ-fate determination in Arabidopsis (van Mourik et al., 2010). For this work, transcriptional regulation of six genes representing the five gene classes in the ABCDE model were modelled using ordinary differential equations (ODEs). For each gene, there is an ODE describing how the gene expression level is influenced by the concentrations of its regulators. Based on gene expression data, whorl-specific concentrations were estimated, which were then used to estimate the various model parameters. These parameters describe the interaction affinity of the various MADS-domain protein dimers, the binding affinity of these dimers to target promoters and the decay of gene products into non-functional components. The model generates continuous time-course expressions for the involved genes in the different floral whorls that reasonably match experimental data. It has to be noted that such a model provides more detailed information about the network’s dynamics than the pioneering discrete-network model (Espinosa-Soto et al., 2004), but contains many more parameters that have to be fitted using experimental data.

Kaufmann et al. (2011) presented a model based on Petri nets that simulates floral quartet concentration dynamics. This model showed that complex formation attenuates stochastic fluctuations in gene expression thus enabling more robust organ-specific expression patterns.

With respect to the modelling of GRNs involved in flowering time control, some of the modelling approaches mentioned above, such as Boolean networks, in which time is not explicitly present, are less suitable. An ODE approach can serve as a framework for modelling the GRN underlying flowering time control. In particular, the control of flowering by photoperiod depends on an integration of external signals captured by photoreceptors and endogenous rhythm controlled by the circadian clock (Hayama and Coupland, 2003; Song et al., 2013). This makes mathematical models for the Arabidopsis clock gene circuit, as recently reviewed, very relevant (Bujdosó and Davis, 2013). One important aspect is the regulation of FT, which is a major direct target of the clock-regulated gene CO (Yanovsky and Kay, 2002). Salazar et al. (2009) used an ODE-based modelling to reconstruct the rhythmic regulation of CO by the circadian clock, and the subsequent effect of CO expression on the regulation of FT. To account for the effect of photoperiod on gene expression, the model assumed an explicit role of post-translational regulation: CO protein is stabilized during the daytime, but rapidly degrades during the night. Thus only the peak of CO mRNA that occurs in the light leads to CO protein accumulation and, therefore, FT activation. Interestingly, the expression levels of FT were simulated for different photoperiod cycles (light/dark), which indicated a non-linear relationship between observed flowering time and the amount of FT transcribed over a cycle-period. In the leaves, FT is not only regulated by the photoperiod sensors but also by temperature-related cues (Song et al., 2013) via an FLC-mediated mechanism. Recently, the regulatory relationship between FLC and FT under the influence of temperature and photoperiod was modelled for the perennial Arabidopsis halleri (Satake et al., 2013). Explicit information about temperature and photoperiod was used to simulate ODEs describing gene expression levels. Assessment of gene expression simulated for different temperatures allowed changes in the perennial flowering cycle to be forecast under a climate change scenario.
Finally, a recent ODE model for the transcriptional regulation of five key integrators of flowering time (Jaeger et al., 2013) was used to explore mechanistically how different feedback loops affect flowering time.

These examples show that GRN modelling enables various qualitative and quantitative predictions on network output and demonstrates the importance of such modelling approaches in understanding how the complex GRNs in plant reproduction fulfil their function.

Conclusion and perspective

The introduction of next-generation sequencing and genome-wide approaches has changed our view of gene regulation and GRNs. We have moved from linear genetic interactions towards global highly connected gene networks. Many genome-wide expression compendia have become available as well as protein–DNA binding profile data that rapidly increased our knowledge of transcriptional regulation and network wiring. Information on epigenetic regulation of gene expression remains poorly represented in the current GRN models and long-distance signalling also needs further consideration. Another complexity concerns spatial and temporal determinants. We know that genes can have different functions, depending on the tissue and stage at which they are expressed. For example, AP1 acts as a repressor of flowering time genes during flower initiation, whereas at a later stages it activates genes involved in organ formation. Information about tissue and stage-specific gene expression (Jiao and Meyerowitz, 2010; Wellmer et al., 2006) and protein–DNA binding (Pajoro et al., 2014) will generate a better understanding of gene function and network dynamics. Most likely, more tissue and stage-specific information about protein–DNA binding will become available in the near future and will allow reconstruction of stage-specific GRNs, highlighting the dynamics of the network topology and interactions. To achieve this resolution, novel technologies such as the INTACT system (Deal and Henikoff, 2011) and single cell approaches (Shapiro et al., 2013) will be important.

Recent studies of TF DNA-binding profiles and gene expression analyses have shown that there is a weak correlation between binding of a TF and changes in expression of its target genes (O’Maoileidigh et al., 2014). An explanation could be that multiple TF binding events or co-factors are needed for gene regulation. In such a scenario only a specific combination of TFs binding will trigger changes in expression. Sequential ChIP analysis (Oh et al., 2012) could be used to identify TF co-occupancy and obtain a better insight into the regulation of gene expression. Another explanation could be that a single binding event to a cis-regulatory element is not sufficient to drive expression, whereas binding of a TF to multiple sites, allowing a conformational change of the DNA, is needed to regulate gene expression. For instance, conformational changes are triggered by the binding of STK to multiple sites in the promoter of its target gene VERDANDI (Mendes et al., 2013). New techniques, such as chromatin capture (Stadhouders et al., 2013) and ChIA-Pet (Zhang et al., 2012) (Fullwood et al., 2009), can be used to characterise cis-regulatory element interactions and their role in gene regulation.

The use of mathematical modelling will facilitate a better insight into GRN complexity, allowing the consequences of network perturbations to be predicted. Current models are primarily based on gene expression profiles and information about non-coding RNAs, protein–DNA, and protein–protein interactions is increasing and becoming incorporated into GRNs. Nevertheless, the experimentally validated protein–protein interactome is still quite sparse (Chatr-Aryamontri et al., 2013; Smaczniak et al., 2012b) and therefore we still rely on predicting the Arabidopsis interactome using orthology relationships, gene ontology, and co-expression (De Bodt et al., 2009; Lee et al., 2010; Lin et al., 2011).

In conclusion, advances in genome-wide studies resulted in the confirmation of previously known genetic interactions at the molecular level, but also identified many novel regulatory interactions. We are just at the beginning of the genome-wide characterization of GRNs and new factors and interactions are undoubtedly waiting to be discovered. A challenge for the near future will be to unravel the spatial and temporal regulation of the genes in the current networks. This will allow further development and application of mathematical models, which will contribute to our understanding of the role and importance of genes during plant reproduction.

Acknowledgements

The authors wish to thank the Marie-Curie Actions for support via the Marie-Curie-ITN network grant SYSFLO (FP7/2007–2011, grant agreement no. 237909).

References


Kaufmann K, Muino JM, Osteras M, Farinelli L, Krajelewski P, Angenent GC. 2010b. Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP–SEQ) or hybridization to whole genome arrays (ChIP–ChIP). Nature Protocols 5, 457–472.


