MATHEMATICAL MODELLING OF THE FLORAL TRANSITION

Jean-Louis Thai Quang Dinh

2017

PhD in Biosciences

Acknowledgments

First, I would like to thank my Charlie Hodgman, Etienne Farcot and Graham Seymour for their supervision over the four years of this PhD, at the University of Nottingham. I would also like to thank Marie Skłodowska-Curie actions and the European Commission, who funded this PhD, as part of the EpiTRAITs project.

Outside of the University of Nottingham, over these four years, I have also had the privilege to work with great people from other institutions, who have generously shared the knowledge of their respective subjects with me.

In my first foray into the world of experimental biology, at the University of Wageningen, I was guided by Gerco Angenent, Richard Immink, Aalt-Jan van Dikj, Suraj Jamge, Froukje van der Waal, Martijn Fiers, Marco Busscher, Alice Pajoro, Leonie Verhage and Sam van Es. I am very grateful to them for the time they took to teach me lab work and discuss modelling ideas.

I also thank Christophe Godin, Eugenio Azpeitia, Frédéric Boudon, Christophe Pradal, Guillaume Cerutti, Sophie Ribes and Jérôme Chopard from the Virtual Plants team for sharing their knowledge of computational modelling and 3D tissue modelling with me.

I am also very grateful to Fabio Fornara for contributing the rice data used in Chapter 2.

I would like to thank all members of the EpiTRAITs consortium, in particular Maike Stam and Helen Bergman, who handled the management of this European project from start to finish.

Finally, I thank my family and friends for their support.

Contents

ABSTRACT

The floral transition is a developmental process through which some plants commit to flowering and stop producing leaves. This is controlled by changes in gene expression in the shoot apical meristem (SAM). Many of the genes involved are known, but their interactions are usually only studied one by one, or in small sets. While it might be necessary to properly ascertain the existence of regulatory interactions from a biological standpoint, it cannot really provide insight in the functioning of the floral-transition process as a whole. For this reason, a modelling approach has been used to integrate knowledge from multiple studies.

Several approaches were applied, starting with ordinary differential equation (ODE) models. It revealed in two cases – one on rice and one on *Arabidopsis thaliana* – that the currently available data were not sufficient to build data-driven ODE models. The main issues were the low temporal resolution of the time series, the low spatial resolution of the sampling methods used on meristematic tissue, and the lack of gene expression measurements in studies of factors affecting the floral transition. These issues made the available gene expression time series of little use to infer the regulatory mechanisms involved. Therefore, another approach based on qualitative data was investigated. It relies on data extracted from published *in situ* hybridization (ISH) studies, and Boolean modelling. The ISH data clearly showed that shoot apical meristems (SAM) are not homogeneous and contain multiple spatial

domains corresponding to coexisting steady-states of the same regulatory network. Using genetic programming, Boolean models with the right steady-states were successfully generated. Finally, the third modelling approach builds upon one of the generated Boolean models and implements its logic into a 3D tissue of SAM. As Boolean models cannot represent quantitative spatio-temporal phenomena such as passive transport, the model had to be translated into ODEs. This model successfully reproduced the patterning of SAM genes in a static tissue structure.

The main biological conclusions of this thesis are that the spatial organization of gene expression in the SAM is a crucial part of the floral transition and of the development of inflorescences, and it is mediated by the transport of mobile proteins and hormones. On the modelling front, this work shows that quantitative ODE models, despite their popularity, cannot be applied to all situations. When the data are insufficient, simpler approaches like Boolean models and ODE models with qualitatively selected parameters can provide suitable alternatives and facilitate large-scale explorations of the space of possible models, due to their low computational cost.

1. GENERAL INTRODUCTION

1.1. THE PROPER CONTROL OF THE FLORAL TRANSITION IS IMPORTANT BIOLOGICALLY AND AGRONOMICALLY

The floral transition is a developmental change whereby some plants switch from producing vegetative organs to producing reproductive organs (Reece and Campbell, 2011). This change arises in response to external (e.g. photoperiod, ambient temperature, prolonged exposure to cold) and internal cues (e.g. age, hormones) (Fornara et al., 2010).

In wild species, the timing of the floral transition determines reproductive success: plants flowering too early deprive themselves of the opportunity to grow more and produce more offspring, while plants flowering too late run the risk of being destroyed by harsh climatic conditions before they can reproduce (Engelmann and Purugganan, 2006). The control of the floral transition is also important in cultivated species, for agronomic and economic reasons. In species whose main product are derived from fruits (e.g. cereals), the timing of the floral transition affects yield (Cockram et al., 2007). As the floral transition marks the end of leaf production, an early floral transition would limit the amount of photosynthetic surface plants can deploy, and therefore, the amount of resources available to produce fruits. However, a late floral transition would expose cultivated species to the same risks as wild species. In species cultivated for vegetative parts (e.g. lettuce, cabbage), proper control of the floral transition is also relevant (Jung and Müller, 2009). Indeed, those

products are harvested before the floral transition occurs. If the floral transition were to occur before the harvest, those products would be substantially altered (Frugis et al., 2001) and would not be marketable. It is however still desirable that those plants are able to flower under some conditions, as it enables propagation through seeds instead of vegetative methods.

The next section presents a snapshot of what was known about the regulatory network of the floral transition at the beginning of this PhD project, which served as the foundation for the modelling projects presented in the following chapters.

1.2. MANY GENES OF THE REGULATORY NETWORK OF THE FLORAL TRANSITION ARE KNOWN

1.2.1. In *Arabidopsis thaliana* (*A. thaliana*)

In the model plant *A. thaliana*, the regulatory network of the floral transition was reviewed several times (Liu et al., 2009; Fornara et al., 2010). The information reported hereafter about the topology of the floral transition network is drawn from these reviews, unless mentioned otherwise. It is summarized in Figure 1.1.

Functionally, the network can be split into two parts: upstream, pathways dedicated to the perception of various environmental and internal cues, each contributing to the decision to flower or not; downstream, a network controlling the identities of the cells of the SAM (Adrian et al., 2009). At the

interface between those two sub-networks are *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), two genes known as floral integrators, because they integrate the signals of the cue-sensing pathways into a decision to flower or not (Simpson and Dean, 2002).

Upstream of the floral integrators, the cue-sensing pathways monitor signals as diverse as photoperiod, ambient temperature, prolonged exposure to cold (vernalization) and age.

The photoperiod pathway (Golembeski et al., 2014) includes the circadian clock, which is a module composed of genes forming three interlocking negative feedback loops, with a 24-hour-periodic expression pattern. The genes of the circadian clock are *PSEUDO-RESPONSE REGULATOR 7* (*PRR7*), *PRR9*, *LATE ELONGATED HYPOCOTYL* (*LHY*), *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), *TIMING OF CAB 1* (*TOC1*) and *GIGANTEA* (*GI*). Late in the day, the circadian clock activates *CONSTANS* (*CO*). CO is an activator of *FT*, however the CO protein is degraded in the dark by CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1), and in the morning by a pathway triggered by the photoreceptor Phytochrome B (PHYB). Therefore, *CO* can only accumulate and activate *FT* under long day (LD) conditions.

The ambient temperature pathway consists of *SHORT VEGETATIVE PHASE* (*SVP*). *SVP* is a repressor of *FT* and *SOC1* up-regulated by low temperatures (Lee et al., 2007).

The vernalization pathway consists of *FLOWERING LOCUS C* (*FLC*), a repressor of *FT* and *SOC1* silenced by prolonged exposure to cold, and *FRIGIDA* (*FRI*), its activator (Amasino, 2004).

The aging pathway consists of *miR156*, an indirect inhibitor of *SOC1*, *AP1* and *LFY* that gets down-regulated by aging (Wang et al., 2009).

Finally, *SOC1* is also upregulated by the gibberellic acid hormone (Jung et al., 2012).

Downstream of the floral integrators are the meristem identity genes, whose expression determine the architecture of the inflorescence and the fates of meristematic cells: vegetative, inflorescence or floral (Adrian et al., 2009; Simon et al., 1996). Vegetative identity genes include *TERMINAL FLOWER 1* (*TFL1*), which is expressed in the pre-transition SAM, but also in the inflorescence part of the post-transition SAM (Conti and Bradley, 2007; Liljegren et al., 1999). Inflorescence genes also include *SOC1* and *AGAMOUS-LIKE 24* (*AGL24*), which constitute a positive feedback loop (Liu et al., 2008, p. 1). *SOC1* is directly activated by the FT-FD dimer. Floral identity genes include *LEAFY* (*LFY*) and *APETALA* 1 (*AP1*), which also form a positive feedback loop (Liljegren et al., 1999; Mandel et al., 1992; Mandel and Yanofsky, 1995; Weigel et al., 1992; Weigel and Nilsson, 1995). The *LFY-AP1* loop can be activated on the *LFY* side by SOC1 if AGL24 is also present (Lee et al., 2008, p. 1). It might also be activated by FT from the *AP1* side, however this FT-*AP1* interaction is

controversial (Benlloch et al., 2011). Finally, both *LFY* and *AP1* are repressed by TFL1, and AP1 represses inflorescence genes (*TFL1*, *SOC1* and *AGL24*).

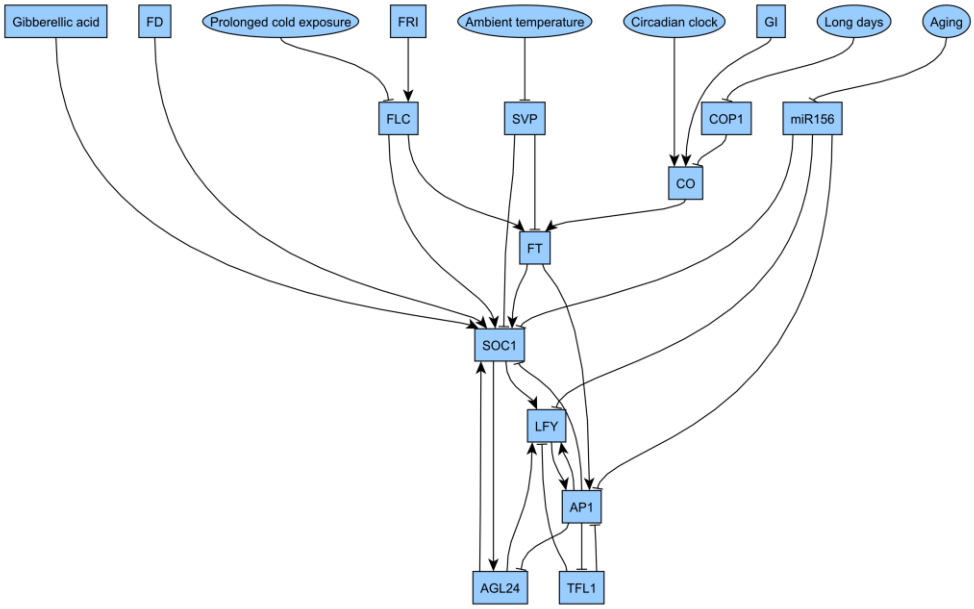


Figure 1.1. Simplified topology of the floral transition network. Rectangles are genes (with the exception of miR156, which is a micro RNA), while ellipses represent abstract concepts. V- and T-shaped arrowheads on edges indicate activating and repressing regulatory interactions, respectively.

The floral transition was studied most extensively in *A. thaliana* due to its status as a model plant. However, the floral transition has also been studied in other species.

1.2.2. In other species

Aside from *A. thaliana*, the network of the floral transition has mainly been studied in Poaceae (Colasanti and Coneva, 2009; Higgins et al., 2010), but some information is also available in other species. Most notably, *FT* is known to be

conserved in all sequenced angiosperms, and is even present in some gymnosperms (Klintenäs et al., 2012). It is thought to have acquired its function as a floral regulator early in the history of angiosperms (Ballerini and Kramer, 2011; Klintenäs et al., 2012).

Concerning the signal-sensing pathways upstream of *FT*, the photoperiod pathway seems conserved in cereals. Homologs of *GI* and *CO* have been identified in cereal species, including rice (*Oryza sativa*), barley (*Hordeum vulgare*) and wheat (*Triticum aestivum*) (Colasanti and Coneva, 2009). Photoreceptor and circadian clock genes also have known homologs in rice and barley (Higgins et al., 2010). Yet, despite the conservation of many components of the photoperiod pathway across species, the effect of the whole pathway on flowering varies drastically. Most notably, in rice, *Heading date 1* (*Hd1*), an ortholog of *CO*, has the effect opposite of *CO* in *A. thaliana* (Hayama et al., 2003).

Some varieties of wheat and barley have a vernalization pathway, functionally similar to that of *A. thaliana*. However, the genes involved do not seem related to *FRI* or *FLC* (Colasanti and Coneva, 2009).

Downstream of *FT*, some *A. thaliana* meristem identity genes also have homologs in cereal species. In rice, *SOC1* is homologous to *Oryza sativa MADS50* (*OsMADS50*) (Lee et al., 2004), *LFY* to *RICE FLORICULA/LEAFY* (*RFL*) (Kyozuka et al., 1998) and *AP1* to *OsMADS15* and *OsMADS14* (Kyozuka et al., 2000). In maize, *LFY* is homologous to *zea floricaula/leafy1* (*zfl1*) and *zfl2*

(Bomblies et al., 2003). Despite these homologies, inflorescence architectures in cereals differ quite strongly from that of *A. thaliana*. The *A. thaliana* inflorescence is simply composed of an inflorescence meristem (on the main shoot) and lateral floral meristems, while rice and maize have intermediate types of meristem between the inflorescence meristem and the floral meristems (branch meristems in rice, spikelet pair meristems in maize, and spikelet meristems in both) (Liu et al., 2009; Tanaka et al., 2013).

As shown above, the network of the floral transition is quite extensive. The aim of this thesis is therefore to model the regulatory network controlling the floral transition, in order to better understand how this process takes place and identify potential gaps in biological knowledge. The insights granted by modelling are expected to provide guidance for plant breeding, potentially resulting in yield improvements.

1.3. MATHEMATICAL MODELS OF THE FLORAL TRANSITION HAVE PREVIOUSLY BEEN DEVELOPED

Mathematical models have long been used in plant biology, and there were over 160 models of plant systems as of January 2015 (Hodgman and Ajmera, 2015). Since then, 6 models related to *Viridiplantae* (green algae and land plants) have been added to the curated BioModels database (Chelliah et al., 2015; Le Novère et al., 2006; Li et al., 2010), and 43 more have been updated. Mathematical models of plant systems describe systems as diverse as plant architecture (Prusinkiewicz, 2007), growth mechanics (Boudon et al., 2015),

intercellular exchanges (Jönsson et al., 2006) and gene regulatory networks (Espinosa-Soto et al., 2004), using a wide variety of formalisms. Even within regulatory network models, several kinds of formalisms can be used, such as Bayesian networks, ODE models, PDE models, Boolean networks and stochastic models (de Jong, 2002). There exist a few models describing the regulatory network of the floral transition. They are reviewed hereafter.

Welch and colleagues developed a model able to predict flowering time for various *A. thaliana* mutants, at several temperatures (Figure 1.2) (Welch et al., 2003). It is based on a neural network, where neurons represent genes from the photoperiod and autonomous pathways. The connections between nodes were derived from known gene regulatory interactions. This model is however focused on the timing of the floral transition and does not aim at predicting gene expression levels (none of the post-*FT* genes are represented in this model).

One of the authors also proposed an ODE model of the same pathways (photoperiod and autonomous), plus *SOC1* and *LFY* (Figure 1.3) (Dong, 2003). In addition to gene expression levels, this model also predicts flowering time, based on the expression of *LFY*. Parameters were fitted to flowering time data, with the constraint that predicted gene expression levels should qualitatively match the overall shape of published expression time series. The use of expression time series is an improvement over the previous model, but the model uses a now outdated regulatory network topology as its base.

Jaeger and colleagues have proposed an ODE model of the post-*FT* part of the network of the floral transition (Figure 1.4) (Jaeger et al., 2013). It involves five genes: *FT*, *FD*, *LFY*, *AP1* and *TFL1*, each of whom stands for a cluster of similarly regulated genes. Expression levels are not fitted to experimental data, only to flowering times, measured in number of leaves formed before and during the floral transition.

Finally, Dong and colleagues have proposed a simple ODE model of the floral transition in maize (Figure 1.5) (Dong et al., 2012). It involves four genes: *VGT1*, *ID1*, *DLF1* and *ZMM4*. However, *ZMM4* is the only gene modelled, the others are simply used as binary input variables. The model is fitted to *ZMM4* expression data and to flowering time measurements.

In conclusion, existing models of the floral transition are mainly models of flowering time, and are accordingly mostly based on flowering time data, even though they do model some gene expression levels internally. Those that do make use of gene expression data do it in a very limited way (only for a few genes, or with no actual parameter-fitting). It therefore seems that developing new models based on up-to-date biological knowledge and fitted to gene expression data would result in more accurate models from a mechanistic point of view, thereby furthering the current understanding of the floral transition.

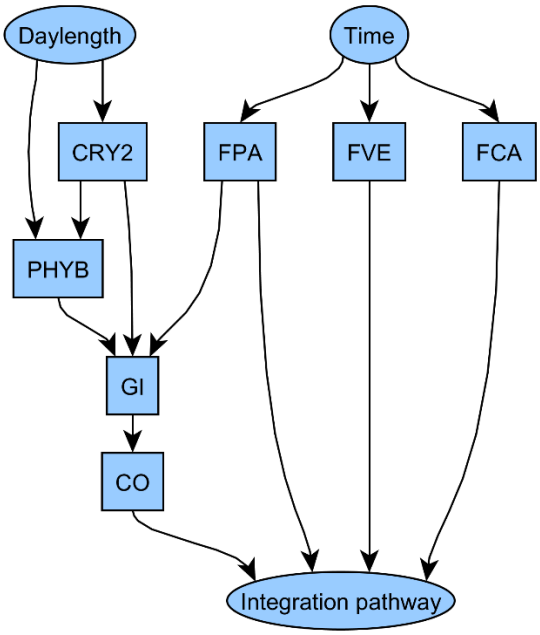


Figure 1.2. Topology of Welch and colleagues' model (Welch et al., 2003). Rectangular and elliptic nodes denote genes and abstract concepts, respectively. Arrows represent regulatory interactions.

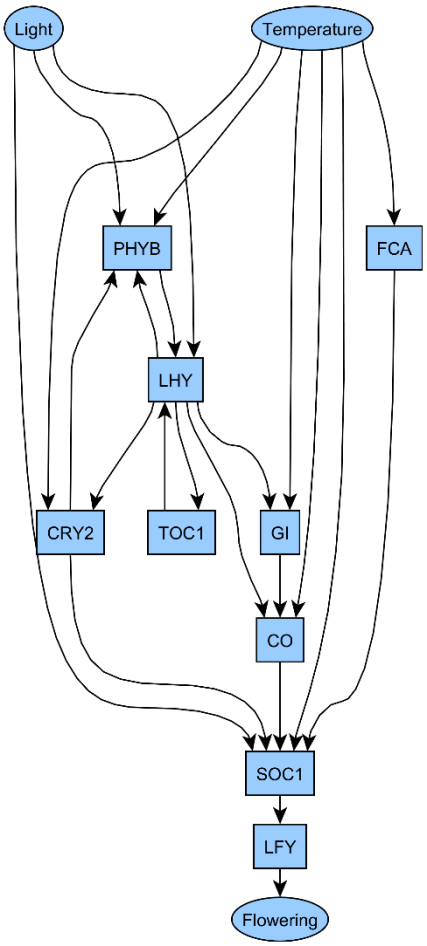


Figure 1.3. Topology of Dong's model (Dong, 2003). Rectangular and elliptic nodes denote genes and abstract concepts, respectively. Arrows represent regulatory interactions.

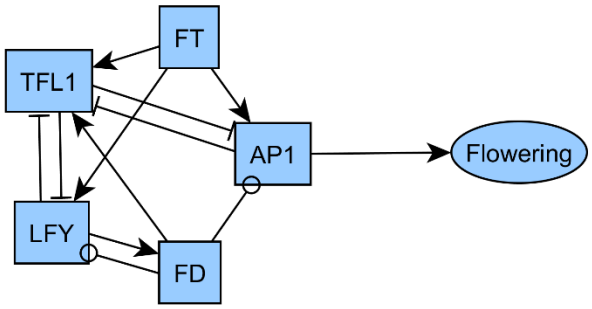


Figure 1.4. Topology of Jaeger and colleague's model. Rectangular and elliptic nodes denote genes and abstract concepts, respectively. V, T and O-shaped arrowheads represent activations, inhibitions and context-dependent regulatory interactions, respectively.

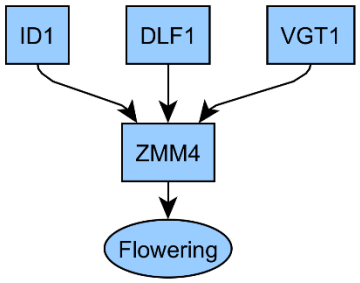


Figure 1.5. Topology of Dong and colleagues' model (Dong et al., 2012). Rectangular and elliptic nodes denote genes and abstract concepts, respectively. Arrows represent regulatory interactions.

1.4. THESIS OUTLINE

The core of this thesis is divided into 4 chapters, numbered 2 to 5. Chapter 2 is an attempt at modelling the regulation of flowering in rice, a crop species with direct real-world applications, but also a model for cereals due to its relatively small genome. However, despite the benefits of studying crop species, there

are still little data available about rice. This is why Chapters 3 to 5 focus on *A. thaliana* instead, as it is the *de facto* main model organism of plant biology.

Chapter 3 was carried out during a secondment at the University of Wageningen (Netherlands). It consists in an analysis of a previously developed model of the floral transition developed by partners from Wageningen, in addition to lab experiments and modelling work to integrate the effect of vernalization into a model of the floral transition. Chapter 3 raises questions concerning the relevance of models of the floral transition that completely ignore the spatial organisation of gene expression.

This is why Chapter 4, whose content was submitted as an article to PLoS Computational Biology, is about modelling the floral transition separately in various domains of the SAM, as well as the transitions of cells between these domains during development. Due to the qualitative nature of the data used, Chapter 4 departs from the ODE formalism and uses Boolean models instead. It shows that commonly used networks of the floral transition lack negative feedback loops that are crucial to the proper spatial organisation of the SAM. In Chapter 4, space is however not represented in a continuous coordinate system, but only as a set of compartments.

Chapter 5 attempts to lift that limitation, by translating a Boolean model of Chapter 4 into the ODE formalism, which is more suited to the modelling of spatial, gradient-generating phenomena, such as intercellular transport.

Finally, a sixth chapter discusses the results of the main four chapters.

2. MIXED-EFFECTS MODELS OF FLORIGEN REGULATION IN ITALIAN RICE CULTIVARS

2.1. INTRODUCTION

The floral integrator FT, which triggers flowering in *A. thaliana*, is conserved in all sequenced species of angiosperms (Klintenäs et al., 2012). Rice (*Oryza sativa*), a model organism of cereal species, has two known *FT* orthologs: *Heading date 3a* (*Hd3a*) and *RICE FLOWERING LOCUS T 1* (*RFT1*). *Hd3a* promotes flowering under short day conditions, while *RFT1* does so under long day conditions (Komiya et al., 2009). The regulation of these genes has however never been described quantitatively. Therefore, the possibility of developing mathematical models of the floral transition network in rice was investigated.

Other genes are conserved between the floral transition networks of *A. thaliana* and rice, including regulators of *FT*: *EARLY FLOWERING 3* (*ELF3*), *GI* and *CO* (*OsELF3*, *OsGI* and *Hd1* in rice, respectively). The roles of *OsELF3* and *OsGI* are similar to those of their *A. thaliana* homologs, however *CO* and *Hd1* have diverged functionally. *Hd1* is indeed not only able to activate *Hd3a* and *RFT1* in SD, but also to repress them in LD.

Although part of the floral transition network is conserved between *A. thaliana* and rice, some genes of the rice network have no homolog in *A. thaliana*, including *Early heading date 1*, *2* and *3* (*Ehd1*, *Ehd2*, *Ehd3*), *Grain yield, plant height and heading date 7* and *8* (*Ghd7* and *Ghd8*), *OsMADS50* and *OsMADS56.*

Ehd1 is an activator of *Hd3a* and *RFT1*, *Ehd2* is an activator of *Hd1* and *Ehd1*, *Ehd3* is an activator of *Ehd1*, and *Ghd7* is a repressor of *Ehd1*. Some of these rice-specific genes also have different behaviours depending on the photoperiodic conditions. Under LD conditions, *Hd1* is a repressor of *Hd3a* and *RFT1, Ghd8* is a repressor of *Ehd1*, *OsMADS50* is an activator of *Ehd1*, *OsMADS56* is a repressor of *Ehd1*, and *Ehd3* is additionally a repressor of *Ghd7*. Under SD, *Hd1* switches to being an activator of *Hd3a* and *RFT1*, *Ghd8* to being an activator of *Ehd1, OsMADS50* and *OSMADS56* no longer affect *Ehd1*, and *Ehd3* has no effect on *Ghd7*.

The network of the floral transition in rice under LD conditions (which corresponds to Fabio Fornara's data set) is summarized in Figure 2.1 (Brambilla and Fornara, 2013; Koo et al., 2013).

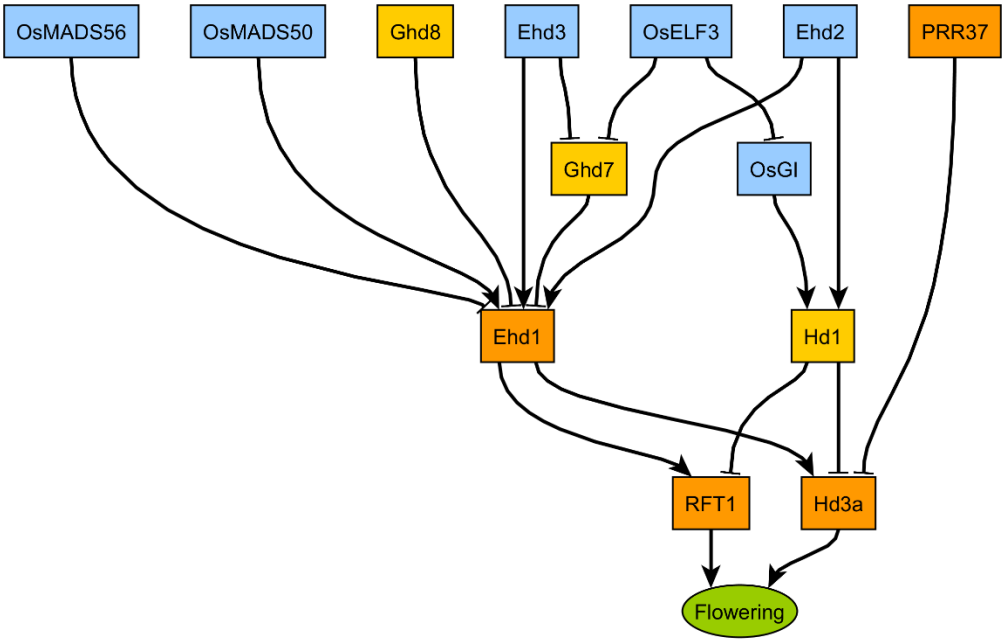


Figure 2.1. Floral transition pathway of rice (*O. sativa*) in LD. Rectangular nodes of the graph represent genes. Edges represent known regulatory interactions (V-shaped arrowheads: activations; T-shaped arrowheads: repressions). Genes for which qPCR measurements or genotypic information were available in this chapter are depicted in orange and yellow, respectively. Other genes are depicted in blue. The green ellipse represents flowering.

Based on data provided by Fabio Fornara and his team (University of Milan, Italy), two modelling opportunities were identified. The first was to predict florigen expression levels, based on those of their regulators, using dynamic or regression models. The second was to predict flowering time for several varieties of rice, based on their gene expression profiles.

2.2. MATERIAL AND METHODS

2.2.1. Gene expression data

Gene-expression data have been provided by Fabio Fornara's team on 17 varieties of rice grown in a field located near Milan in 2012 (Gómez-Ariza et al., 2015). The varieties are: Augusto, Balilla, Carnaroli, Eolo, Fragrance, Gladio, Lido, Loto, Nembo, Nipponbare, Panda, Roma, Sant'Andrea, Selenio, Thaibonet, Vialone Nano and Volano. The gene expression measurements come from leaf tissue samples collected from March to June, which means the plants were grown under LD conditions. For all varieties except Nipponbare, leaf tissue was sampled from 40 to 110 days after germination, at 14-day intervals. Nipponbare is a late-flowering cultivar, so additional samples were taken at time points 126, 140 and 154 days. The mRNA levels of genes *Ehd1* (Figure 2.2), *PRR37* (Figure 2.3), *Hd3a* (Figure 2.4) and *RFT1* (Figure 2.5) were measured by quantitative reverse transcription polymerase chain reaction (qRT-PCR) and normalized using *Ubiquitin* (*Ubq*) as the reference gene. For each variety, the functionality of the alleles of some regulators of *Hd3a*, *RFT1* and *Hd1* (*Hd1*, *Ghd8* and *Ghd7*) was also assessed through sequencing: the alleles were considered functional if their sequences did not exhibit any early STOP codons or reading-frame shifts. Given that many of these alleles are not functional by these standards, this genotypic profiling was meant to be a cheaper – albeit coarser – way to measure the inter-varietal variability of

genetic expression in the floral transition pathway and complement the qRT-PCR data.

As the plants were grown under LD conditions and flowering is mostly controlled by *RFT1* under LD, the florigen modelling work focused on *RFT1*.

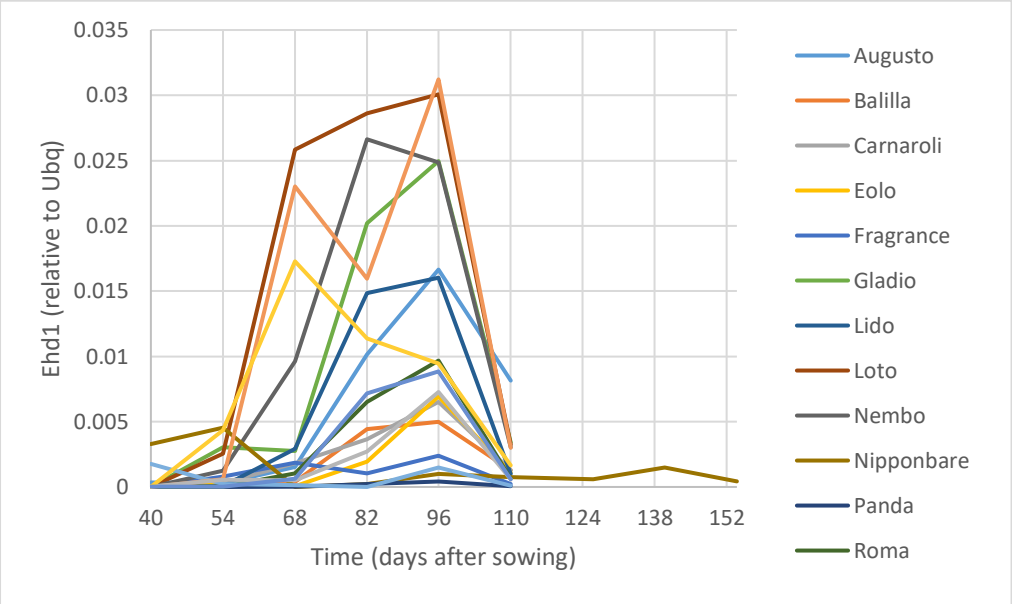


Figure 2.2. Time series of *Ehd1* expression in the 17 rice varieties. Each curve colour represents a different variety.

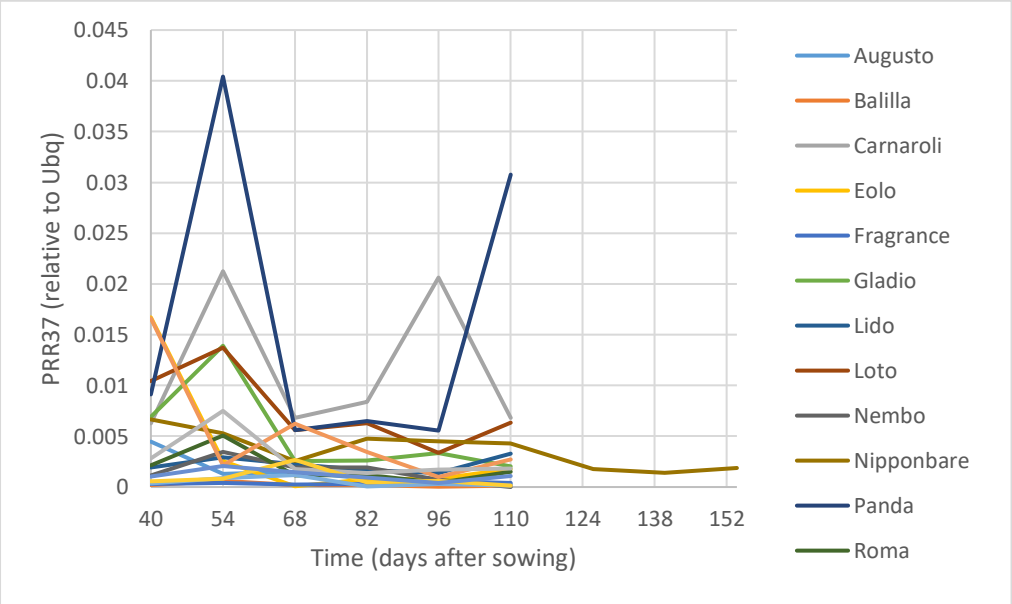


Figure 2.3. Time series of *PRR37* expression in the 17 rice varieties. Each curve colour represents a different variety.

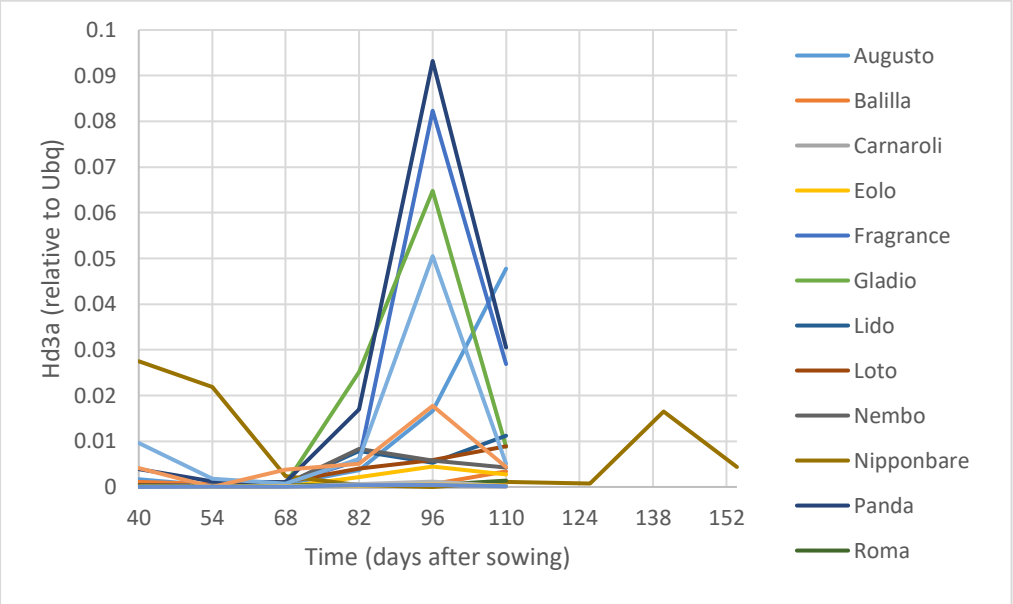


Figure 2.4. Time series of *Hd3a* expression in 17 rice varieties. Each curve colour represents a different variety.

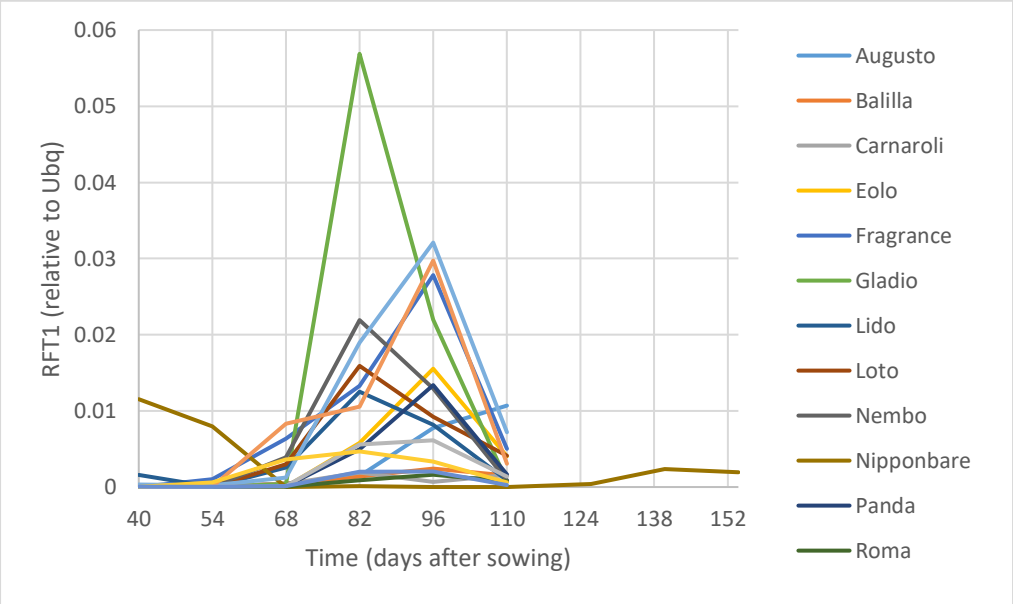


Figure 2.5. Time series of *RFT1* expression in the 17 rice varieties. Each curve colour represents a different variety.

2.2.2. Flowering time data

The plants were also scored for flowering time, i.e. the day when they started to flower was determined visually and recorded for each variety (Figure 2.6).

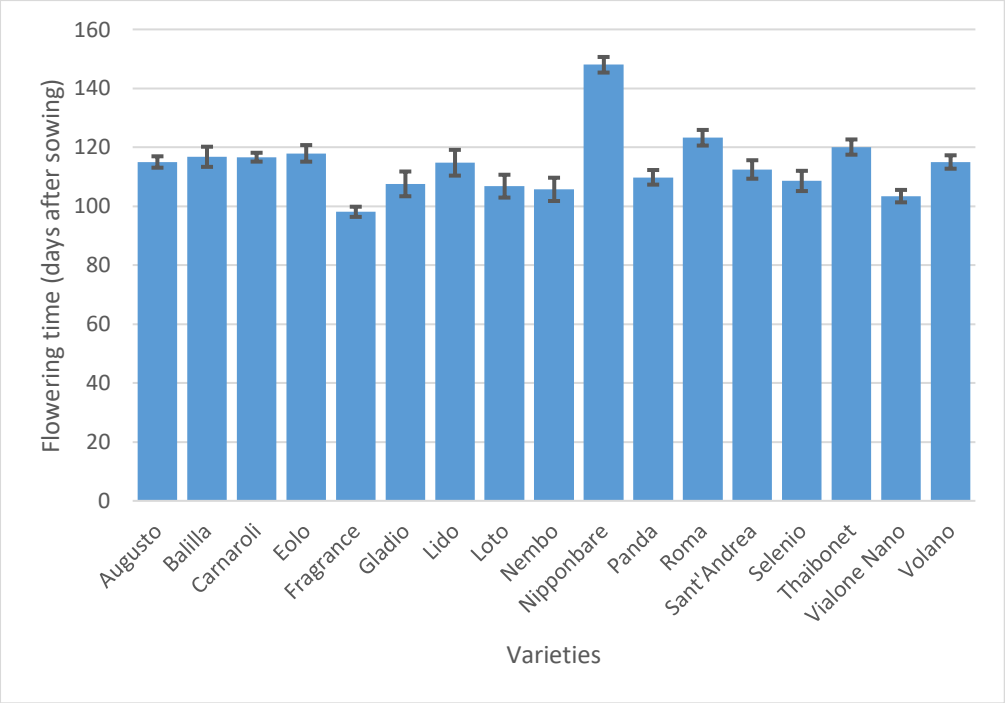


Figure 2.6. Flowering times of the 17 rice varieties. Error bars represent standard deviation.

2.2.3. Choosing a modelling formalism for the regulation of *RFT1*

Among the genes whose expression or functionality has been measured in Fabio Fornara's data set, there were two known regulators of *RFT1*: *Ehd1* and *Hd1*. Among these two, only *Ehd1* was measured quantitatively, and there seemed to be a strong linear dependency between *Ehd1* and *RFT1*. To assess whether a linear regression model would be sufficient to model *RFT1*, a linear regression model (Equation 2.1) and an ODE model (Equation 2.2) with variety-

specific parameters were fitted to the data, and their corrected Akaike information criteria (AICc) were compared.

The AICc (Cavanaugh, 1997) is a penalized likelihood criterion like the Akaike information criterion (AIC) (Akaike, 1973), used to select models that offer a good fit to the data (i.e. models with a high likelihood), while avoiding overfitting by penalizing overly complex models (i.e. models with many parameters; Equation 2.3). Unlike the AIC, it does not only apply to cases where the number of observations is much higher than the number of parameters. Concretely, the AICc has an additional penalty for extra parameters, which helps with preventing overfitting for smaller numbers of observations. There exist other model selection criteria with heavier penalties for extra parameters, such as the Bayesian information criterion (BIC) (Schwarz, 1978), but the AICc was selected because it converges towards the AIC for large numbers of observations, and the AIC is asymptotically optimal for selecting the model with the least mean squared error, under the assumption that the true model is not included in the candidates (Yang, 2005). As the general aim of this study was make the best predictions possible from a very limited set of input variables, this was a desirable property.

The regression model was fitted with the built-in lm function of R. The ODE model starts at t=40, which corresponds to the earliest measurements. It was simulated with the deSolve package for R (Soetaert et al., 2016), and was fitted using the Nelder-Mead algorithm (Nelder and Mead, 1965). deSolve relies on solvers from the ODEPACK collection (Hindmarsh, 1982). deSolve was used

with its default solver LSODA, which solves systems of the form $\frac{dy}{dt} = f$, switching automatically between a method for non-stiff systems (Adams) and a method for stiff systems (BDF). The Nelder-Mead algorithm is an algorithm for multidimensional unconstrained optimization without derivatives. Its main principle is that it starts from a simplex (a convex hull delimited by k+1 vertices in the k-dimensional parameter space) and minimizes the target function at its vertices by replacing the worst vertex at each iteration, through expansion or contraction of the simplex.

The coefficient of determination R^2 (Equation 2.4) was also computed for both models to provide insight into the percentage of the variability in the data accounted for by the model, but it was not used to determine which formalism to retain.

Equation 2.1. Simple regression model of *RFT1* with variety-specific coefficients.

$$RFT1_{it} = \alpha_i . Ehd1_{it} + \epsilon_{it}$$

$$\epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- $RFT1_{it}$: measured *RFT1* expression for variety i at time t
- α_i: sensitivity of *RFT1* to *Ehd1* in variety i
- $Ehd1_{it}$: measured *Ehd1* expression for variety i at time t
- ϵ_{it}: residual error in the expression of *RFT1* for variety i at time t
- σ^2: variance of the residual errors

Equation 2.2. Simple ODE model of *RFT1* with variety-specific parameters.

$$\widehat{RFT1}_\iota(40) = 0$$

$$\frac{d\widehat{RFT1}_\iota}{dt}(t) = \alpha_i . Ehd1_i(t) - d_i . \widehat{RFT1}_\iota(t)$$

$$RFT1_{it} = \widehat{RFT1}_\iota(t) + \epsilon_{it}$$

$$\epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- $\widehat{RFT1}_\iota(40)$: initial predicted value for all varieties at the first measurement
- $\frac{d\widehat{RFT1}_\iota}{dt}(t)$: predicted derivative of *RFT1* expression for variety i at time t
- α_i: sensitivity of *RFT1* to *Ehd1* in variety i
- $Ehd1_i(t)$: linear interpolation of the *Ehd1* measurements for variety i at time t
- d_i: degradation rate of *RFT1* in variety i
- $\widehat{RFT1}_\iota(t)$: predicted *RFT1* expression for variety i at time t
- ϵ_{it}: residual error for the expression of *RFT1,* for variety i at time t
- σ^2: variance of the residual errors

Equation 2.3. AICc and log-likelihood for the models of this study.

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

$$AIC = 2k - 2\ln L$$

$$\ln L = \sum_{i=1}^{I} \sum_{t\in\{40,\ldots,T\}} \left(\ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2}\frac{\epsilon_{it}^2}{\sigma^2} \right)$$

- k: number of parameters
- n: number of independent data series
- L: likelihood
- T: date of the last measurement
- ϵ_{it}: residual error for species i at time t
- σ^2: variance of the ϵ_{it}

Equation 2.4. Coefficient of determination R^2.

$$R^2 = 1 - \frac{\sum_{i=1}^{I}\sum_{t\in\{40,\ldots,T\}}\left(RFT1_{it} - \widehat{RFT1}_{it}\right)^2}{\sum_{i=1}^{I}\sum_{t\in\{40,\ldots,T\}}(RFT1_{it} - \overline{RFT1})^2}$$

- I: number of varieties
- T: date of the last measurement
- $RFT1_{it}$: measured value of *RFT1* for variety i at time t
- $\widehat{RFT1}_{it}$: predicted value of *RFT1* for variety i at time t
- $\overline{RFT1}$: average value of the *RFT1* measurements

2.2.4. Regression models of the regulation of RFT1

It was suspected that other measured genes than *Ehd1* might also affect the expression of *RFT1*. Therefore, a global linear regression model including all possible effects was designed (Equation 2.5). As *Ehd1* and *PRR37* were the only potential regulators for which an expression time series was available and *RFT1* exhibits clear temporal patterns for each variety, it was assumed that these temporal patterns should be controlled by at least one of *Ehd1* or *PRR37*. The effects of *Ehd1* and *PRR37* were therefore represented by linear functions in the equation of *RFT1*. The time-independent, binary variables (*Hd1, Ghd8 and Ghd7*) were assumed to modulate the effects of *Ehd1* and *PRR37* through the coefficients of the linear functions.

Equation 2.5. Global linear regression model.

$$RFT1_{it} = \left(\alpha + \alpha_{hd1} + \alpha_{ghd8} + \alpha_{ghd7}\right)Ehd1_{it} + \left(\beta + \beta_{hd1} + \beta_{ghd8} + \beta_{ghd7}\right)PRR37_{it} + \epsilon_{it}$$

$$\epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- $RFT1_{it}$: *RFT1* measurement for variety i at time t
- $Ehd1_{it}$: *Ehd1* measurement for variety i at time t
- $PRR37_{it}$: *PRR37* measurement for variety i at time t
- α: default coefficient of the effect of *Ehd1* on *RFT1*
- α_{hd1}: contribution of a non-functional *Hd1* allele to the effect of *Ehd1* on *RFT1*
- α_{ghd8}: contribution of a non-functional *Ghd8* allele to the effect of *Ehd1* on *RFT1*
- α_{ghd7}: contribution of a non-functional *Ghd7* allele to the effect of *Ehd1* on *RFT1*
- β: default coefficient of the effect of *Ehd1* on *RFT1*
- β_{hd1}: contribution of a non-functional *Hd1* allele to the effect of *PRR37* on *RFT1*
- β_{ghd8}: contribution of a non-functional *Ghd8* allele to the effect of *PRR37* on *RFT1*
- β_{ghd7}: contribution of a non-functional *Ghd7* allele to the effect of *PRR37* on *RFT1*
- ϵ_{it}: residual error for the expression of *RFT1,* for variety i at time t

All submodels (models including only a subset of these effects) were fitted using the lm function of R, scored according to AICc, and the best one was retained. The exhaustive assessment of submodels was done with the MuMIn package for R (Bartoń, 2016).

2.2.5. Characterization of the gene expression profile required to trigger the floral transition

The hypothesis tested in the second part of the study is that flowering happens a fixed time $\delta \geq 0$ after a critical value of *Hd3a* or *RFT1* expression is reached. Let F_i be the flowering time of variety i, X_i the function mapping time to Hd3a or RFT1 expression for variety i (interpolated linearly from the measurements), X_i^{-1} its inverse function defined as $x \mapsto \min_t\{t \,/\, X(t) = x\}$, and x^* the critical value of *Hd3a* or *RFT1* triggering flowering. Assuming the hypothesis is true, this would mean:

$$\forall i, \quad F_i = X_i^{-1}(x^*) + \delta$$

$$\forall i, \quad F_i - \delta = X_i^{-1}(x^*)$$

$$\forall i, \quad X_i(F_i - \delta) = x^*$$

This implies:

$$Var\big(X_i(F_i - \delta)\big) = 0$$

Therefore, there would be a δ for which $Var\big(X_i(F_i - \delta)\big) = 0$ if the hypothesis were true. In practice, the variance is unlikely to be perfectly 0, but it might be small. In this context, "small" should be relative to the values of X_i

measured at that time, because X_i varies manifold during the experiment. However, variance is not homogeneous with values of X_i, therefore, the standard deviation was compared to the average value of X_i instead, and the value minimizing this quantity is the estimate of the delay between reaching the critical expression level and flowering (Equation 2.6).

Equation 2.6. Estimator of δ (delay between crossing a threshold of gene expression and flowering).

$$\hat{\delta} = \underset{\delta}{\operatorname{argmin}} \frac{s\big(X_i(F_i - \delta)\big)}{\overline{X_i(F_i - \delta)}}$$

- $s\big(X_i(F_i - \delta)\big)$: empirical standard deviation of the X_i at δ before flowering
- $\overline{X_i(F_i - \delta)}$: empirical average of the X_i at δ before flowering

2.3. RESULTS

The first modelling subproject was about the control of florigen expression. As the expression data came from plants grown in LD conditions, this work focuses on the regulation of *RFT1*, which controls flowering in LD.

2.3.1. *Ehd1* and *RFT1* are linearly dependent

Ehd1 is a known regulator of *RFT1*, and multiple time series seem to indicate a strong linear relationship between *Ehd1* and *RFT1* (Figure 2.7), although this might be exaggerated by the scarcity of data points corresponding to medium levels of *Ehd1* and *RFT1*. This still suggested that *Ehd1* could explain most of

the variability of *RFT1*. However, the slopes of the regressions vary depending on the varieties. This seemed to indicate that the sensitivity of *RFT1* to *Ehd1* (the slope of the regression) was influenced by other factors. These other factors could include genomic sequence differences in the *RFT1* promoter, Ehd1 DNA-binding site amino acids, or both. Other differences in anatomy, physiology and alleles of other genes could also play a part.

To find out whether linear regression models are suitable or another formalism such as ODE models would be a better formalism to capture the relationship between *Ehd1* and *RFT1*, a simple ODE model was fitted to the data. Parameters values are given in Table 2.1. Both models have R^2 coefficients over 65% (Table 2.2), confirming that *Ehd1* has the potential to be a key regulator of *RFT1*. Their AICc values were also compared (Table 2.2), and indicated that the gain in goodness of fit resulting from the addition of degradation parameters (d_i) required by the ODE formalism was not worth the extra complexity (a two-fold increase in the number of parameters). Therefore, the rest of this study on the regulation of *RFT1* focuses on regression models. The key issue that remained to be addressed was why the apparent sensitivity of *RFT1* to *Ehd1* (the α_i coefficients) varied across varieties. To answer this question, potential effects of the other genes included in the data set were investigated.

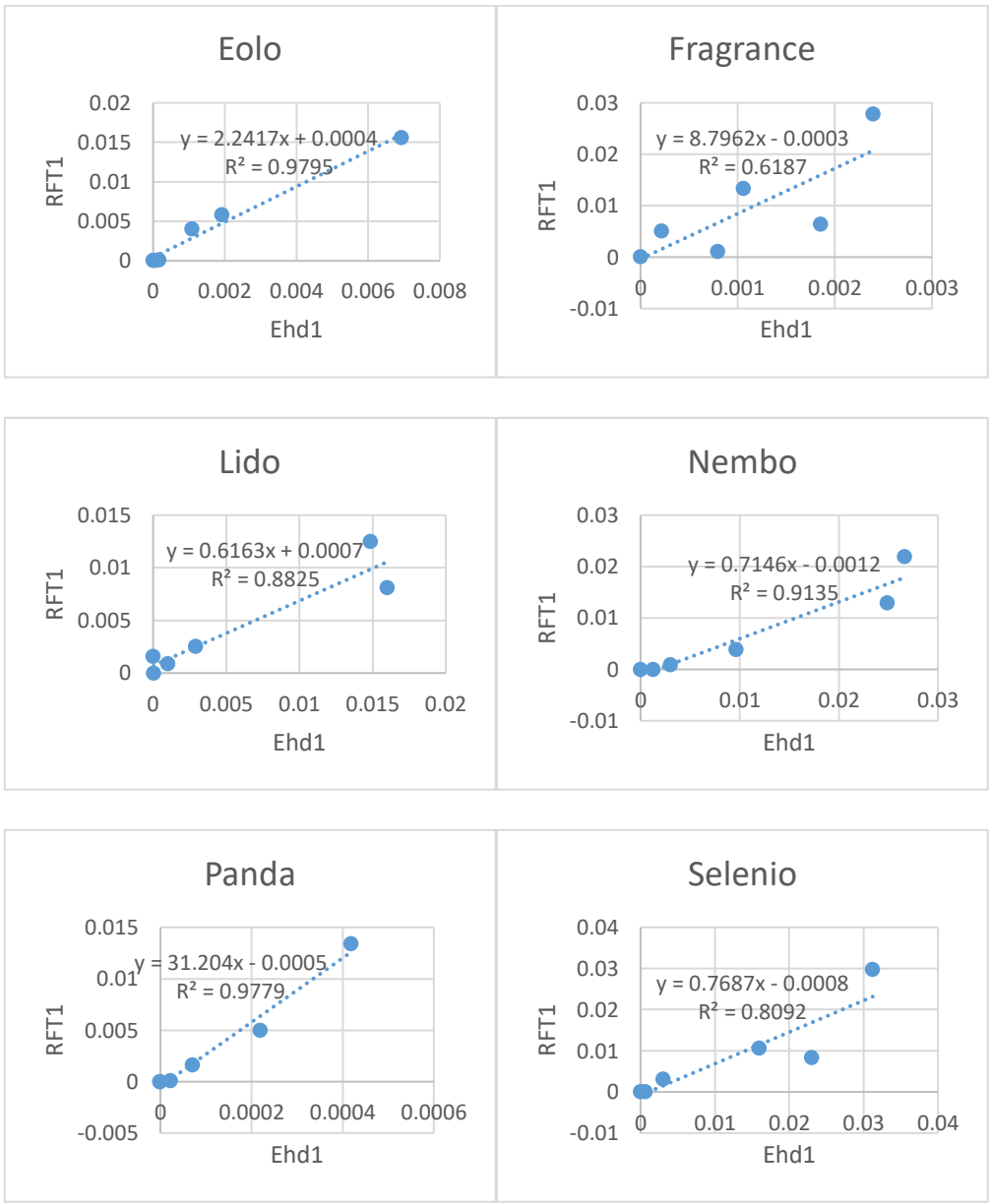


Figure 2.7: Linear regressions of *RFT1* expression against *Ehd1* expression in 6 varieties of rice. Some varieties show a clear linear relationship between *Ehd1* and *RFT1* levels (e.g. Panda, Selenio). Others seem compatible with such a relationship, but lack intermediate values to fully support this conclusion (e.g. Eolo, Lido).

Table 2.1. Parameter values for the simple regression and ODE models.

Variety	Regression	ODE	
	α_i	α_i	d_i
Augusto	0.515	0.478	0.855
Balilla	0.439	0.122	0.212
Carnaroli	0.232	0.014	0.026
Eolo	2.325	1.793	0.751
Fragrance	8.603	6.614	0.719
Gladio	1.605	2.330	1.471
Lido	0.667	1.579	2.624
Loto	0.342	0.406	1.127
Nembo	0.660	0.891	1.297
Nipponbare	2.200	2.542	1.413
Panda	29.731	21.712	0.669
Roma	0.159	0.002	-0.032
Sant Andrea	9.095	27.998	1.196
Selenio	0.734	1.040	1.452
Thaibonnet	0.992	0.481	0.487
Vialone Nano	0.280	0.231	0.733
Volano	0.249	0.190	0.586

Table 2.2. Goodness-of-fit and complexity statistics for the simple regression and ODE models.

Model	AICc	Log-likelihood	Number of parameters	R2
Regression	-772.94	408.45	17	65.12%
ODE	-752.52	418.11	34	70.98%

2.3.2. Exhaustive analysis of models of *RFT1* regulation hint at effects of *Hd1*, *Ghd8* and *Ghd7* on the sensitivity of *RFT1* to *Ehd1*

Exhaustive analysis of the submodels of the global model revealed the best model (according to the AICc) is the model including the effect of *Ehd1*, and its interactions with *Hd1*, *Ghd8* and *Ghd7*. Its fit is presented in Figure 2.8, and its parameter values, AICc, log-likelihood and R^2, as well as those of its own submodels, are reported in Table 2.3. Including the three interaction terms substantially improves R^2 from 26.34% to 42.35%, meaning the sensitivity of *RFT1* to *Ehd1* may be affected by the functionality of *Hd1*, *Ghd8* and *Ghd7*.

The maximum R^2 achievable by a linear regression model is 65.12% (achieved by the model with variety-specific coefficients), meaning there is still room for improvement. The variability unexplained by the selected model is not explained by the global model (the most complex one) either, which only achieves an R^2 of 43.44% - barely better than the selected model. This suggests that there may be factors affecting the expression of *RFT1* that have not been measured in this data set.

Looking at the fits of the selected model (Figure 2.8) and at the distribution of the α_i in the simple regression model with variety-specific coefficients yields some insight into the issue (Figure 2.9). It shows three outlying varieties, with much higher α_i that could not be predicted accurately by the selected model (Figure 2.8). This indicates that something in the regulation of *RFT1* is probably different in these varieties. However, there were no additional data to

determine what this or these differences might be. Excluding these varieties yields an R^2 of 49.90%.

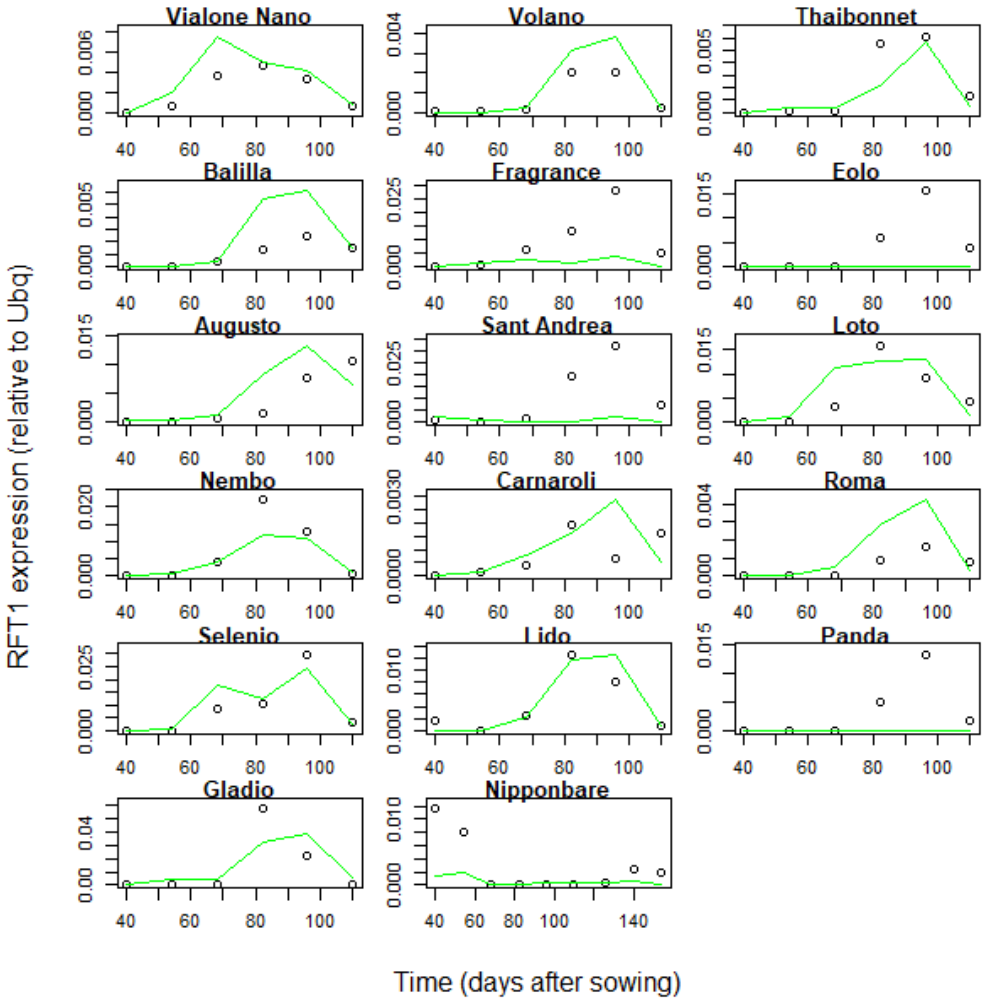


Figure 2.8. Fit of the selected model for the 17 studied varieties. Each frame represents a variety. Measurements are represented by circles, and the model predictions are the green lines. The expression level of RFT1 is predicted reasonably well for most varieties, but it is severely underestimated for some of them.

Table 2.3. Parameters and goodness of fit statistics for the best model and its submodels. An empty cell denotes a parameter not included in the model.

Parameters				Statistics		
Ehd1	Ehd1 x ghd7	Ehd1 x ghd8	Ehd1 x hd1	AICc	Log-likelihood	R^2
0.437	-0.7597	0.7818	0.3428	-753.5	382.1	42.35%
0.5509	-0.7136	0.9862		-751.1	379.8	39.76%
0.4391		0.5851	0.3193	-750.6	379.5	39.43%
0.5455		0.7874		-748.9	377.6	37.18%
0.4455			0.5416	-745.1	375.7	34.89%
0.4455	-0.3079		0.5814	-743.9	376.1	35.45%
0.688				-734.3	369.2	26.34%
0.6861	0.0328			-732.2	369.2	26.35%
				-677.1	339.6	-29.48%

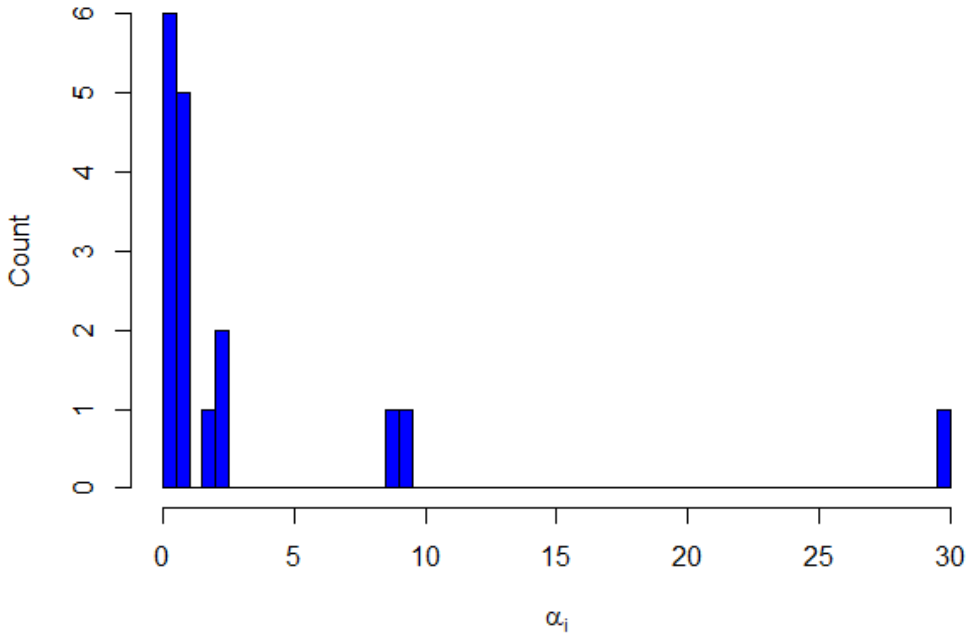


Figure 2.9. ***Histogram of the α_i for the selected model.*** The α_i are the apparent sensitivities of *RFT1* to Ehd1. The most outlying varieties are Fragrance, Sant'Andrea and Panda.

2.3.3. No pattern consistent across all varieties can be found between flowering time and *RFT1* or *Hd3a* levels

The hypothesis was that flowering occurred once a critical value of *Hd3a* or *RFT1* expression was reached, with an optional delay $\delta \geq 0$. Using the estimator detailed in Material and methods, δ was estimated to be 38.67 and 45.70 days before flowering for *RFT1* and *Hd3a*, respectively (Figure 2.10). However, even for these optimal values, the inter-varietal standard deviations of expression levels corresponding to these delays are still very high: 76% and 89% of the average values, respectively. Histograms of the distributions of *RFT1* (Figure 2.11) and *Hd3a* values (Figure 2.12) at their respective $\hat{\delta}$ before flowering revealed large disparities in their expression levels. For *Hd3a*, the inter-varietal fold-change ($\frac{\max_i X_i(F_i - \hat{\delta})}{\min_i X_i(F_i - \hat{\delta})}$) is over 150, and it is undefined for *RFT1* because of a 0 value. The minimum relative standard deviation of *Hd3a* looks like it might be inflated by an outlier (Figure 2.12). Removing the outlying variety (Selenio) makes the inter-varietal fold-change drop to 4.36. As for *RFT1*, excluding the 0 value yields an inter-varietal fold change of 8.68. While removing those extreme values results in substantial improvements, the fold-changes are still too high to infer the existence of a common *RFT1* or *Hd3a* threshold triggering flowering in all varieties. The search for a flowering trigger shared by all rice varieties was therefore unfruitful. Moreover, it was also impossible to propose a trigger for flowering on a per variety basis, as data

were only collected during a single growing season. Doing so would require inferring a pattern for each variety from a single time series.

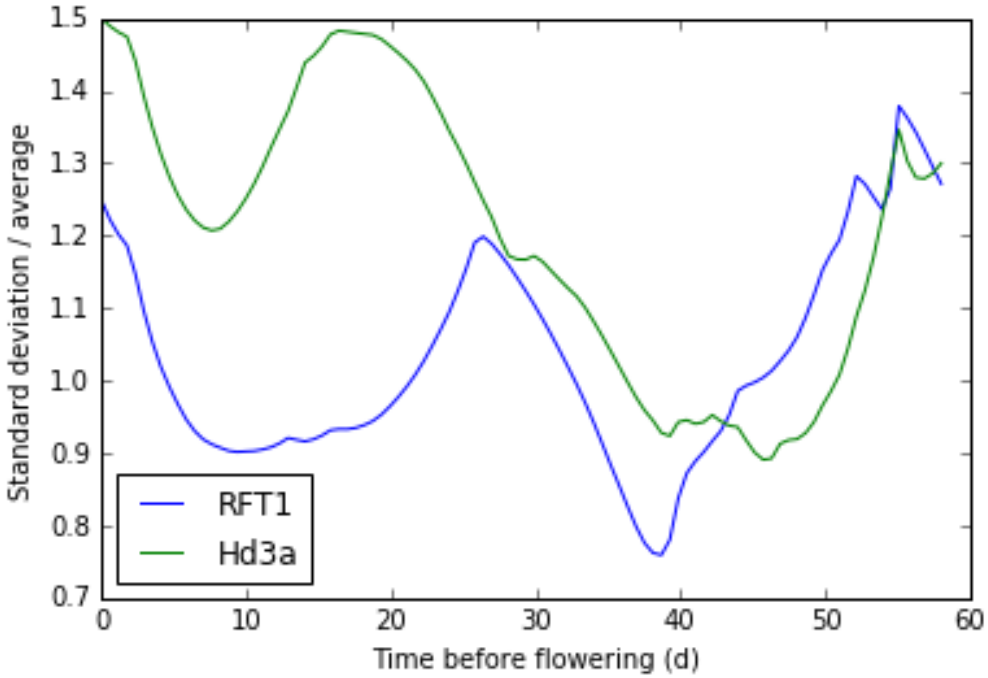


Figure 2.10. Relative standard deviation of RFT1 and Hd3a expression levels across the 17 rice varieties, as functions of time before flowering. Minima are reached at 38.67 and 45.70 days before flowering for *RFT1* and *Hd3a*, respectively.

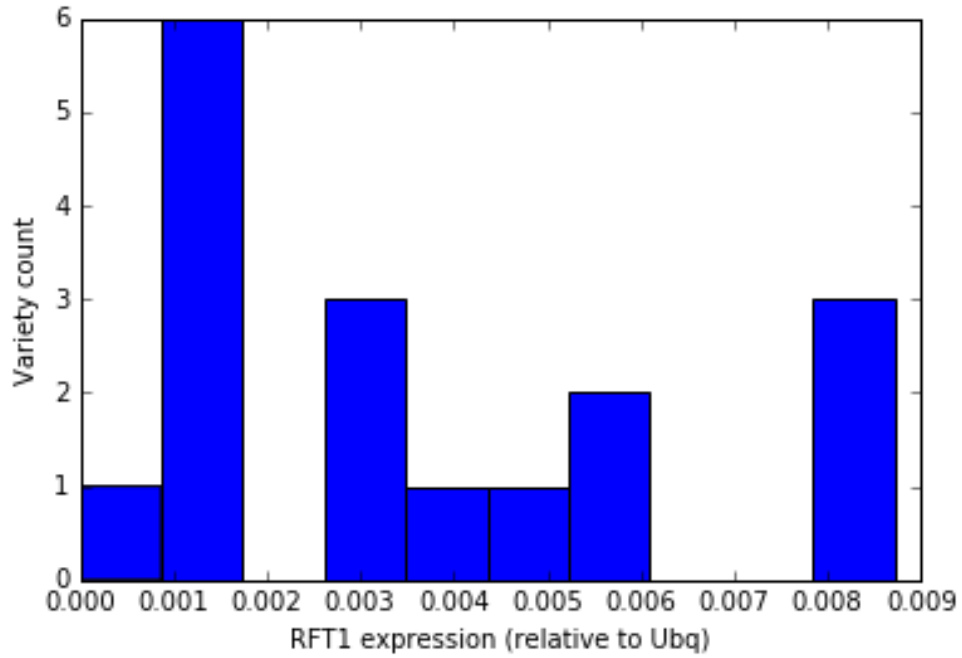


Figure 2.11. Distribution of RFT1 expression levels across the 17 rice varieties 38.67 days before flowering.

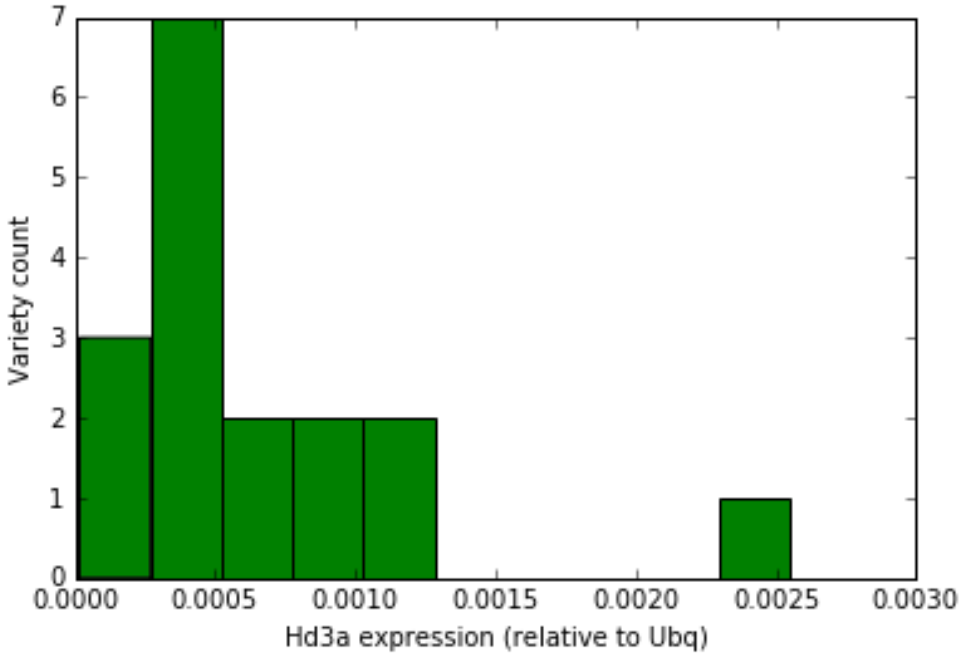


Figure 2.12. Distribution of Hd3a expression levels across the 17 rice varieties 45.70 days before flowering.

2.4. DISCUSSION

This study has shown a strong linear dependency of *RFT1* on *Ehd1*. This is consistent with the anterior finding that *Ehd1* is a regulator of *RFT1* (Doi et al., 2004). However, it also suggests that other genes may modulate the effect of *Ehd1*.

2.4.1. *Hd1*, *Ghd7* and *Ghd8* seem to modulate the control of *RFT1* by *Ehd1*

This study suggests that *Hd1*, *Ghd7* and *Ghd8* might be able to modulate the regulation of *RFT1* by *Ehd1*, however there is no biological evidence supporting this so far. Therefore, it might be worth investigating the potential involvement of *Hd1*, *Ghd8* and *Ghd7* in the regulation of *RFT1* by Ehd1 through biological experiments.

Hd1, Ghd8 or Ghd7 might be able to form a protein complex with Ehd1. To check for protein-protein interactions, one might use yeast two-hybrid assays (Y2H) (Fields and Song, 1989) or bimolecular fluorescence complementation (BiFC) (Kerppola, 2008), however the methods might not be able to pick up indirect binding. Should that be the case, Förster resonance energy transfer (FRET) (Sekar and Periasamy, 2003) might be able to provide an alternative, as it does not necessarily require the close proximity of the two proteins.

There could also be interactions between Hd1, Ghd8 or Ghd7, and the *RFT1* locus. This could be investigated via chromatin-immunoprecipitation-related methods (ChIP) (Gade and Kalvakolanu, 2012).

Interestingly, it was recently shown that Ghd8 and Hd1 are involved in a protein complex binding the promoter of *Hd3a* (Goretti et al., 2017). Given the homology between *Hd3a* and *RFT1*, it is possible that they might also be involved in complexes binding *RFT1*.

2.4.2. Three outlying varieties point to the control of *RFT1* by other factors

There are not enough data to suggest whether there is a common cause to the unusually high sensitivities of the three varieties, or multiple causes are involved. A possible explanation might be that the real sensitivities of *RFT1* to *Ehd1* in the three outlying varieties are not so different from those of the main group, but *RFT1* is also responding to another, unknown signal (e.g. an *Ehd1* homolog). That unknown signal may be present in all 17 varieties, meaning that its effects would be indiscernible from those of *Ehd1* within the main group of 14 varieties. However, varieties with lower *Ehd1* expression levels – like the three outliers – would appear to have higher sensitivities when the unknown factor is not accounted for. This hypothesis could be tested by screening for genes with temporal expression patterns correlated to those of *Ehd1* and *RFT1*, and mutating them.

Beyond this new insight into the regulation of *RFT1* in rice, this study also led to an interesting observation about the modelling of gene regulatory networks from time series data in general.

2.4.3. The low temporal resolution results in the relevance of linear regression models

The amount of experimental work required to measure gene-expression time-series has resulted in the temporal resolution being rather low. The interval between two consecutive measurements was usually 14 days. Compared to gene activation delays, which are usually on the scale of a few hours (Rosenfeld and Alon, 2003), these intervals are extremely large. This makes it unlikely that measurements would capture transient states of the floral transition pathway, such as the onset of a gene's expression. In most cases, measurements will be representative of quasi-steady states. Interestingly, this is what the regression models used in this study represent. Under the assumption that *RFT1* is controlled directly only by *Ehd1* and in a quasi-steady state, an ODE model can become equivalent to a regression model.

$$\left.\begin{array}{l} \dfrac{dRFT1}{dt}(t) = \alpha.Ehd1(t) - kd.RFT1(t) \\ \dfrac{dRFT1}{dt}(t) \to 0 \end{array}\right\} \Rightarrow RFT1(t) \to \frac{\alpha}{d} Ehd1(t)$$

A slow-evolving ODE model does therefore not differ much from a simple linear regression model in this case. A problem would have arisen if the genetic regulation graph had cycles, but in the very simple network studied here, this did not occur. This regression approach has the benefit of decreasing the number of parameters to be estimated, which is why the AICc chose the regression model over ODE model. It is particularly useful to do so, since, as

evidenced by the formulae above, parameters α and d would be difficult to separate based on quasi-steady state observations.

In those formulae and in the selected models, it was assumed that the synthesis rate of *RFT1* was linear with respect to its activator *Ehd1*, which is different from the convention that gene regulations follow Hill equations. However, for a Hill coefficient of 1, the Hill equation is equivalent to the Michaelis-Menten equation, which has a nearly linear domain, when the concentration of the activator is not saturating. The suitability of a linear response of *RFT1* to the expression of *Ehd1* might therefore indicate that the Ehd1 binding sites on the *RFT1* promoter are never saturated, and that the Hill coefficient of the activation of *RFT1* by Ehd1 is close to 1.

Finally, in addition to the modelling limitations resulting from the temporal resolution of the time series, there were also limitations caused by the high variability of gene expression across varieties.

2.4.4. High inter-varietal variability in gene expression levels prevents the prediction of flowering time based on genotypic data

The high inter-varietal variability in gene expression levels prevented the second modelling case from being solved, namely, the prediction of flowering time for each variety. The main roadblock was that the initiation of flowering could not be predicted satisfactorily from florigen expression, because florigen expression varies wildly across varieties (Figure 2.4 and Figure 2.5). It is unclear whether this variability is real and caused by genetic or physiological

variations, or results from issues with the measurements or their normalization. In particular, *Ubq* might not be a good reference gene, if its expression level varies between varieties.

In any case, the inter-varietal variability should be checked using additional reference genes and – if needed – new measurements. Should this inter-varietal variability be confirmed, separate models of the triggering of flowering would have to be established for each variety. This would in turn require multiple time series of gene expressions to be generated for each variety.

2.5. CONCLUSIONS

This work confirmed that *Ehd1* has the potential to be a key regulator of *RFT1*. It also suggested that *Hd1*, *Ghd8* and *Ghd7* might play a role in the control of *RFT1* by *Ehd1*, by modulating the effect of *Ehd1*. However, in the studied data set, the expression levels of all genes – including *RFT1* – vary greatly across varieties, which is suspicious. It is unclear whether those variations are real or come from a problem with the data. This point should be addressed first.

Assuming the variations are real, the sensitivity of *RFT1* to *Ehd1* is different for each variety. Part of this variability was attributed to the functionality – or not – of *Hd1*, *Ghd8* and *Ghd7*, but part of it remains unexplained. This could be addressed by quantitative expression measurements for these three genes or other genes involved in the floral transition. Such a data set would also be beneficial in identifying the molecular trigger of flowering, as there might be

better indicators of flowering than *RFT1* – assuming the inter-varietal variations of *RFT1* expression are real.

The above observations on the limitations of the data for rice suggest that a more widely studied species, for which more data are available, should be investigated. This is the subject of the next chapter.

3. QUANTIFICATION OF THE IMPACT OF VERNALIZATION ON THE FLORAL TRANSITION PATHWAY IN SF2 FRI COL-0 ARABIDOPSIS THALIANA

3.1. INTRODUCTION

This chapter was realized in collaboration with Aalt-Jan van Dijk and Gerco Angenent's laboratory, from Wageningen University.

In *A. thaliana*, the timing of the floral transition is controlled by multiple pathways. One of them is the vernalization pathway, whose function is to lift a block preventing plants from flowering when they are exposed to prolonged periods of cold (i.e. when they are vernalized).

The vernalization process has been described at the molecular level by Angel and colleagues (Angel et al., 2011), who also provided the only mathematical model of vernalization to date. Vernalization comes from the silencing of the *FLC* gene, which is a repressor of *SOC1* and *FT* (two crucial integrator genes of the floral transition). Silencing *FLC* therefore results in an increase in the expression of *SOC1* and *FT*. The vernalization-induced silencing of *FLC* relies on epigenetic modifications. At the beginning of a period of cold, the histones of a nucleation locus in the *FLC* gene start being methylated. As the period of cold continues, histone methylation spreads from that locus to the rest of the gene. When the temperature eventually rises to normal levels again, histone methylations prevent *FLC* from being expressed. Angel and colleagues' model focuses on demonstrating that the silencing of FLC occurs in an all-or-nothing

fashion (as can be observed by microscopy and GUS staining), where individual cells are only either on or off, and the quantitative aspect of *FLC* silencing only occurs at the tissue level, as a consequence of the proportions of on and off cells. Therefore, Angel and colleagues' model does not study the consequences of *FLC* regulation on the floral transition network. Mathematical models of the floral transition network of *A. thaliana* do exist, as reviewed in Chapter 1 (Dong, 2003; Jaeger et al., 2013; Welch et al., 2003). However, only Dong's model actually includes *FLC*. It unfortunately has limitations, as it was based on a now outdated network topology and primarily aimed at predicting flowering time. As a consequence, it did not aim at predicting gene expression accurately, so long as flowering time was predicted accurately.

The goal of this chapter is therefore to develop gene expression models of the floral transition network to quantify the effects of vernalization on the expression levels of floral transition genes, with a particular focus on the effect of *FLC* on *SOC1* and *FT*.

This chapter covers three sub-projects. The first is the modelling of the silencing of *FLC*, as a response to cold exposure. The second is comprised of experimental work aimed at gathering suitable data to fit an ODE model of the floral transition network including the vernalization pathway. The third is an analysis of a new model of the floral transition developed at Wageningen University, which was being considered as a candidate to be expanded to include the effects of vernalization.

3.2. MATERIAL AND METHODS

3.2.1. Implementation of the ODE model of vernalization

The ODE model of vernalization was implemented in R, and integration was performed using the deSolve package (Soetaert et al., 2016). deSolve is based on the ODEPACK solvers (Hindmarsh, 1982) and is a widely used R package for the integration of ODEs. More details are available in the Material and Methods section of Chapter 2.

3.2.2. Vernalization experiments

In this study, experiments were carried out to generate new gene expression data for vernalized *A. thaliana* plants. Not all varieties of *A. thaliana* are vernalization-sensitive: the Col-0 line, which is the most commonly used in experiments, is not. This is because Col-0 barely expresses *FLC* in the first place, due to its non-functional allele of the *FRI* gene (a key activator of *FLC*).

To investigate the vernalization pathway in a background comparable to that used in other flowering-related experiments, an introgression line created by introducing a functional *FRI* allele from the Sf2 ecotype into a Col-0 background (Lee et al., 1994) was used. Seeds were provided by Caroline Dean's lab (John Innes Centre, Norwich).

Sf2 FRI^{+} *A. thaliana* plants were grown according to the following protocol. Seeds were sown on cubes of rock wool (five seeds per cube) soaked in a 1g/L solution of Hyponex NPK = 7-6-19 fertilizer (HYPONeX JAPAN) and covered in

cling film to prevent desiccation. The seeds were then stratified by being exposed for 48h to a temperature of 4˚C, in darkness to prevent early germination. They were then transferred to a short-day growth chamber (23˚C, 8h light) for a duration of 2 weeks. The cling film was removed after 4 days, when the seedlings were visible. The plants were then sent to a cold room (4˚C, 12h light) for the cold treatment, except for one batch that went straight to the next phase, without cold treatment. The other batches were removed from the cold treatment after 1, 2 or 3 weeks, respectively. The final phase was a long-day growth chamber (23˚C, 16h light).

For the whole duration of the experiments, plants were watered two or three times a week with a solution of Hyponex (1g/L). Plants were sampled regularly at a frequency depending on the stage of the experiment, until they had spent 2 weeks in the final growth chamber. This duration was chosen based on a previous experiment on WT Col-0 plants, where the plants flowered in 12.6 days, and on the assumption that cold-treated FRI+ Col-0 plants would behave similarly to WT Col-0 plants. The sampled plants were dissected, with leaves and enriched meristem material being collected separately, except for the two earliest time points. As the seedlings were too small, they were not dissected but were collected as whole seedlings instead. For each time point, each kind of tissue was harvested in triplicates. Five to ten plants were pooled together for each sample, depending on the growth stage of the plants. Samples were first flash-frozen in liquid nitrogen, before being stored in a -90˚C freezer until they were processed for RNA extraction. Each sample was then homogenised

by shaking with glass beads, and subjected to an RNA extraction procedure using Invitrap® Spin Plant RNA Mini Kits (STRATEC Biomedical). The extracted RNA was subjected to a DNase treatment to avoid any contamination by genomic DNA, using Ambion® TURBO DNA-free™ DNase Treatment and Removal Reagents (Thermo Fisher Scientific), then RNA was quantified using a nano-drop. The cDNA was synthesized by Suraj Jamge and Froukje van der Waal (Wageningen University) and quantified using a BioMark qRT-PCR machine (Fluidigm) at Enza, a partner company of Wageningen University. The BioMark system is able to run qRT-PCRs on all combinations of 96 samples and 96 primers concurrently, while using very little cDNA. This system was selected for its efficiency, because the earlier time points of our experiment yielded very little RNA, as there was little biological material. The 96 primers used for the qRT-PCR were selected from published articles or designed by Suraj Jamge (Table 3.1).

Table 3.1. 96 primers used for the BioMark qRT-PCR. Compiled by Suraj Jamge.

Genes	ATG no.	Forward primer	Reverse primer
FT	AT1G65480.1	CTGGAACAACCTTTGGCAAT	AGCCACTCTCCCTCTGACAA
FD	At4G35900.1	CACCTCCTGCAACTGTTCTG	AGCCTCGAAAGAGGTGTTGA
LFY	At5G61850.1	ATTGGTTCAAGCACCACCTC	ACGGACCGAATAGTCCCTCT
YLS8	At5G08290.1	TTACTGTTTCGGTTGTTCTCCATTT	CACTGAATCATGTTCGAAGCAAGT
FLC	At5G10140.1	CGAACTCATGTTGAAGCTTGTT	GGAGAGTCACCGGAAGATTG
FLC_CD	At5G10140.1	GGCTAGCCAGATGGAGAATAA	TCAACCGCCGATTTAAGGT
SOC1	At2G45660.1	AGCTGCAGAAAACGAGAAGC	TGAAGAACAAGGTAACCCAATG
SVP.1	At2G22540.1	GAAGAGAACGAGCGACTTGG	GAGCTCTCGGAGTCAACAGG
SVP.3	At2G22540.3	ACCGGAAAACTGTTCGACATGA	TTCTTTACTCATTCGGGCGTGAT
MAF 1.2		CCTCAATGTTTTGAACTCGATC	TCGACATTTGGTTCTTCAAGCTTGC
MAF 1.3		GTCCCTTAAAGAAAAGGTTAGTG	CAAGAATCATCATAGCCTAGA
MAF 1.4		GATCGTTATGAAATACAACATGC	GTATTCTTTCCCATCTGGCTAGC
MAF 1.5		CAGTCCAAAGCAAGCTTGAAG	CAGTCCAAAGCAAGCTTGAAG
MAF 2.1		AAAACGGTGGGGAAGAAGAC	AAAAACTTCTGAATCAGGCTGT
MAF 2.2		AGCTCGAGACTGCTCTGTCC	TCAACTGATGAATTAGCTTCAAGA
MAF 2.3		CGGAGAACTTGCTGAGAGAAG	AGCCGTTGATGATGGTGATT
MAF 3.1		GCTTGAAGAATCAAATGTCGATAATG	TGAGCAGCGAAAGAGTCTCC
MAF 5.1		CAGGATAAGGAGAAGTTGCTGAA	ACTTGAGAAGCGGGAGAGTC
FUL	At5g60910	AAGGACAATTAGTCCAATGCTCCAA	CAACTCTCTCCACAAAGCCATCTCT
CAL		GATCGCTCATCAGACTTCTCCTTTC	GCCAAGGTAATTGTAAATGGGTTCA
AGL24	At4G24540.1	CGGAATTGGTGGATGAGAATAAGAG	GTTCCACTGTCGTAGCTTGACACAT

AGL15	AT5G13790	GTCAAGCGATTCAGTGACAACAAAC	CAGAGAACCTTTGTCTTTTGGCTTC
AGL16	At3g57230	ACATGAAAAGGTTTCAGAGGTCGAG	AGATGGACATGTTCGTTCGAGGTAT
AP1	AT1G69120	AAATCCAGCATCCTTACATGCTCTC	CAGTTCGAGATCATTCCTCCTCATT
AP2	AT4G36920	TGCCGAGTCATCAGGGAATCCTAC	TCCCAAGCTCAAATCGAGGTTGTG
STM	At1g62360	TCTCCGGTTATGGAGAGACAGCAA	TCGACTTCTTCCTCGGATGACCCA
FDP (bzip27)	AT2G17770.1	AATCAACCACCACCACCACCAC	AAGAGGCAGAGAGCCATAGAGAGC
FLD	AT3G10390.1	GGAAAGCAAGTCTTTGAGCACAGG	CACCAACATGTAAGGAACCACCAG
FRI	AT4G00650.1	AGTCACCGCTGGCATTTAAAGAAG	TGCCATCCTGGTAGTTCTTTCGC
SPL4	AT1G53160.1	TTTCTCTCAGGACTTAACCAACGC	CTTGGAGGTCATGAAACCTACTGC
SPL9	AT2G42200.1	TGTGGCTGGTATCGAACAGAGG	TTCCGGAAGCTGATGAAACCTG
SPL15	AT3G57920.1	TCGCTCCATCTCTTTACGGAAACC	TGCATCACTGATCTTGCGGTTG
AGL12	AT1G71692.1	CTCAGATTCGCTCTGCTAAGATGG	TGAGGACTCCTTCCTTGTTCCTC
AGL23	AT1G65360.1	TGACCACTTTCGAGGGTGTGTTG	TTCTACTTCCGCCTTCACCTCAG
AGL71	AT5G51870.3	TCGTATTGTCAGGTCAAGAAAGGC	TCTCGTTCAAGAGCTCCCTCTC
AGL72	AT5G51860.1	ACACGATAAAGCGATACGCTGAG	TTCCGGTTATGGACTTCAAGCAC
MRG1	At4g37280	CTTACCATGGTCCTCGCGTCTAC	CGGTATGTTTCAACAATCTATCCGC
MRG2	At1g02740	CTTCTGCTACCTGCTCCTCC	TTCGTCCCAACTTTTGTTCC
EFS/SDG8	AT1G77300	GTAAGCAAAAGGCGTGCTTC	TTCTTCTCCACAACCCAACC
SDG26	AT1G76710	CGGGTTCACGGTAACATTTC	CATGCTTCTGAGCGACATTC
TFL2/LHP1	AT5G17690.1	AGACAATGTCCAGGAAGTGTTGG	TGCTTCCTTCCCATCAGACCTC
CLF (SDG1)	AT2G23380.1	TTGTTTGCTAAACGGGACTTGCTG	TTCTTGCAGCTCTTTGGGCAACC
SNZ	AT2G39250.1	TGGGTGCCCATAGTAAAGGAAATG	CAACGGCTTCCCATGCAAACTC
SMZ	AT3G54990.1	AGCAAGTTTATTTGGGCGGGTTTG	TGATAGCAGCTCGGTCGTAAGC

TOE1	AT2G28550.1	AATAATCCCGCCGAGGGAAGAG	AACAATGGTGGTGGTTGTGGTC
TOE2	AT5G60120.1	TGGAGCAGCTTCATGGAAACATGG	GCTTCCCTTCCCTTCCATTGTACG
VRN1	AT3G18990.1	GTACCAGCCAACAAAGGGTATGC	GGCGTTGGCTCTTCAGCTTTAAC
VRN2	AT4G16845.1	TCGGGATAGCGAGGATGAAGTC	TCCACAAAGTCATCAAGCATCTGG
EMF1	AT5G11530.1	GTGGGAGGGATTTGTGCAGTTC	CATCTGTTAATCCCTCTGCCTCAG
ELF6	AT5G04240.1	TGGCATTCCCTGCTGTAGGTTG	TCCTTTGCTACGTTGAGCCACTG
SDG2 (putative)	AT4G15180.1	TGCTTGGTGGGTTGCCAGATTG	CTCGAAATTGATGAACCGGACCAG
GCN5 (putative)	AT3G54610.1	AATCTCAGGGCTCGTGCCAAAG	TTTGAGTCGTCCTGCTTGCTCCTC
VIN3	AT5G57380.1	GTATGGGATTGGGAGTGATGAT	CAAAACAACCTGAAACCTGTGA
COOLAIR (FLC antisense)		ACCTTATTCGTGTGAGAATTGC	TTGACAGAAGTGAAGAACACATACA
UBC	AT5G25760	CTGCGACTCAGGGAATCTTCTAA	TTGTGCCATTGAATTGAACCC
ACT	AT3G18780	TCCGCTCTTTCTTTCCAAGCTCA	CGAAGCGATGATAAAGAAGAAGTTCG
SAND	AT2G28390	CAGACAAGGCGATGGCGATA	GCTTTCTCTCAAGGGTTTCTGGGT
UBQ10	AT4G05320	GGTTTGTGTTTTGGGGCCTTG	CGAAGCGATGATAAAGAAGAAGTTCG
TIP41-like	AT4G34270	CATTTCAGTCTCTATCTGCGAAAGGGTATCC	CACCACAATAAGTCAGTGGAGTAACTCCTTAC
PP2AA3	At1G13320	GCGGTTGTGGAGAACATGATACG	GAACCAAACACAATTCGTTGCTG
bZIP29	at4g38900	CCAGAGACTTCATTCATCTTTCGGC	GCTGATGAGCGGATGAAATTAGGG
bZIP30	At2g21230	TCACTTGAATCCTGCTCTTATCCGC	AGTAAGGAGAAATGGGTGGAATCGG
bZIP59	At2g31370	GTCTTCCTCCTCCATCTCCATCAGG	CCGATGTCCAATCTTCTTAGGTGGG
bZIP70	At5g60830	AGTGTCCATCGCTCTCATTGTTTCG	AAGTCAGTGTTTGGTAGGAATGCCG
bZIP75	At5g08141	GAAGACGACGTCCATGTTCAAGACC	CGCGTTCTTCTTGCTGATTCTCG
SHL	AT1G62360	ATGCCCAAGCAAAAAGCTC	CGGTAGTGGAATTGTACTCG
EBS	AT4G22140	TGGTATCATCCTGCGTGTGT	CGCTTCGTTTCCACCTTAAC

SPL3	At2g33810	TTCAAACCGGGATCTCACAC	CAACGTTTCTGCCAACAATG
SPL5	At3g15270	CAGGACAGCATAGAGGGGACT	CATCATTCAAGCGACCACAG
TEM1	AT1G25560	GTCCGGTTCAGACTGTGGTT	GATAATCGCCTGCTTCTTGG
TEM2	AT1G68840	AGAGAAAACCCGGTTCAGGT	TATCGCCTGCTTCTTGGAAC
GI	AT1G22770	CTGTTCAGACGTTCAAAGGC	TGGTTTCCTCTTGGATTCAT
GA2ox7	AT1G50960	AAACCCTAGCGCCACTTCTC	CGTTCACTTGTTTCCCCAGT
GNL	AT4G26150	TTTGGAGACCCAGAGCAACT	AACCATTCCGTGCGATAGAG
GNC	AT5G56860	TGAGGGGTTGAGAAAGATGG	TCTTCCTCGCTTCATCATCA
TFL1	AT5G03840	GCTCTTTCCTTCTTCTGTTTCCTCC	CAGCGGTTTCTCTTTGTGCGT
BFT	AT5G62040	ATGTCAAGAGAAATAGAGCCACTAATA	TTAATAAGAAGGACGTCGTCG
TSF	AT4G20370	CACCACTGGAAATGCCTTTGGC	AACCGTTTGTCTTCCGAGTTGCC
MFT	AT1G18100	ACAATCCAGTGGACCCATTC	CCATTCCGATGAGCTTTACA
MAF 4	AT5G65070	TCGCACAAGGAGTTGCTAGA	GGGTCTTCACAAGCTCCATC
CO	At5g15840	AGCTGTGATGCTCAAGTTCACTCT	GCAGACCCGGACACGTTTAT
PIF4	AT2G43010	CCCATCACAGAACGATCTCGAT	AGGAGCCACCTGATGAGGAACT
PIF5	AT3G59060	AATTCCCGGTTATGAACCGGT	TACCTAGCGAGCTGCTCCGATA
FCA	AT4G16280	TGTTCGAACGAGAGCAACAG	AACGGCTGTAATTGGGTCTG
FVE (MSI4)	AT2G19520	ACTGGGCACCAAGATAATGC	GTCCCAATCGTTGTGATGTG
AGL14	At4g11880	TGCTGATGGAGAAGTGTGAGATGC	TGTCGAGTCTCAGGAGGTCCAATG
AGL18	At3g57390	GCCACTTGACTCCCAGAGTTATCG	ACTTCCTTGCAGTTGGGGTTGTC
AGL19	At4g22950	TGCATCAATGCCTTCTCCAAGCAA	TCAGCAAGCGAGAGACGAAACATC
AGL21	At4g37940	CTTCATGCTGGAGCTTGCAAAGTC	AGCTATTCTCTGTGATGCCGAGGT
AGL42	AT5G62165	AATTGTTCAAGGAGCAGTTGGAG	GGCAAACCGATGAATAAGTCAG

AGL17	At2g22630	TGCCAGCTCCAGTGTGAAATC	TTGCTCCTCCATCTTAGCCGT
miR-156a	At2g25095	CTCTCCCTCCCTCTCTTTGATTC	AGGCCAAAGAGATCAGCACCGG
miR-172b	AT5G04275	TTTCTCAAGCTTTAGGTATTTGTAG	TCGGCGGATCCATGGAAGAAAGCTC
MIR172a-2	AT2G28056	TTTCTCAAGCTTTAGGTATTTGTAG	TCGGCGGATCCATGGAAGAAAGCTC
CBF1	AT4G25490.1	CCGCCGTCTGTTCAATGGAATCAT	TCCAAAGCGACACGTCACCATCTC
CEN (ATC)	AT2G27550	TCCTGATGTTCCTGGACCTAGTG	TCGGTAACAATCCAGTGCAAGTG

In addition to the acquisition of these expression time series, flowering time was also measured using the leftover plants. Flowering time was measured in three ways: as the raw number of days from sowing to the apparition of visible flower buds, as the number of warm days (i.e. excluding days in the cold chamber), and as the number of rosette leaves at the beginning of flowering.

3.2.3. Analysis of qRT-PCR results

Normalized expression values were computed using the Eleven module (Smith, 2014) for Python, based on the GeNorm algorithm (Vandesompele et al., 2002). GeNorm enables the ranking of multiple reference genes and the normalization of expression data with a set of the best reference genes. However, as *YELLOW-LEAF-SPECIFIC GENE 8* (*YLS8*) – the reference gene used by Valentim and colleagues in a similar experiment presented below (Valentim et al., 2015) – was consistently more stable than any other, it was the only reference gene retained for normalization.

3.2.4. Data from Valentim and colleagues

Valentim and colleagues measured time series of the expression of genes involved in the floral transition (Figure 3.1 and Figure 3.2) (Valentim et al., 2015). The biological material comes from *A. thaliana* Col-0 plants grown in LD conditions at 21˚C. Two kinds of tissues were harvested every day, between 5 and 17 days after germination: plant apices (including the SAM) and leaves. In the apices, the expression of *SVP*, *FLC*, *FD*, *SOC1*, *AGL24*, *LFY* and *AP1* were measured by qRT-PCR and normalized by Y*LS8*. In the leaves, the expression levels of *SVP*, *FLC* and *FT* were measured and normalized in the same way.

They also measured the flowering times of the WT, and various single and double mutants grown under LD at 23˚C: *soc1-2*, *soc1-6*, *agl24-2*, *ft-10*, *fd-3*, *flc-3*, *svp31*, *svp32*, *svp41*, *soc1-2/agl24-2*, *soc1-2/svp32* and *svp41/agl24-2* (Figure 3.3) (Valentim et al., 2015). These data provide additional insight into the functioning of the floral transition network, as they can be used to get an idea of what happens when a node of the network is removed (detailed in the following section). As this data set did not include a mutant for *LFY*, a key gene of the floral transition, an extra data point was added using published data for the *lfy-12* mutant (Jaeger et al., 2013). That experiment showed the lfy-12 mutant flowered 9.24% slower than the WT, which equates to a flowering time of 13.80 days in this data set.

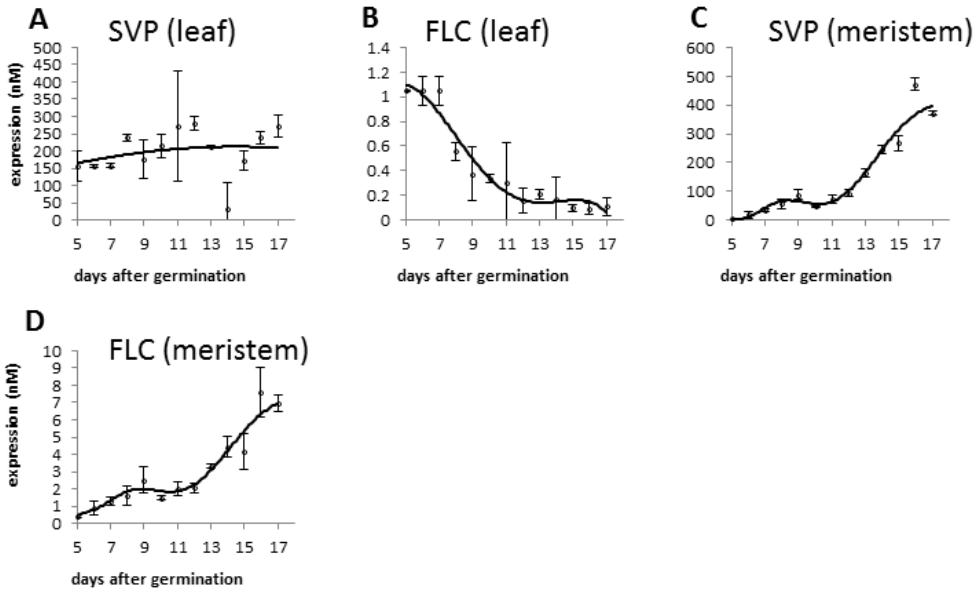


Figure 3.1. Gene expression measured by Valentim and colleagues for the input genes of their model. Data points represent the average measurements with their standard deviations. The lines are interpolations used by Valentim and colleagues' model. Reprinted from Valentim *et al.*, 2015 (CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/legalcode).

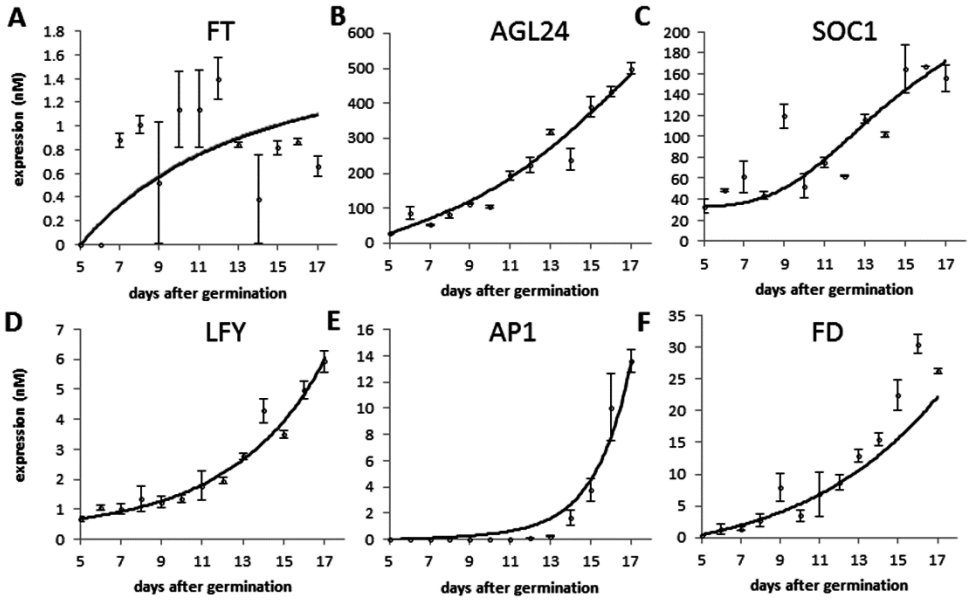


Figure 3.2. Gene expression measured by Valentim and colleagues for their modelled genes. Data points represent the average measurements with their standard deviations. The lines are the predictions of Valentim and colleagues' model. Reprinted from Valentim *et al.*, 2015 (CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/legalcode).

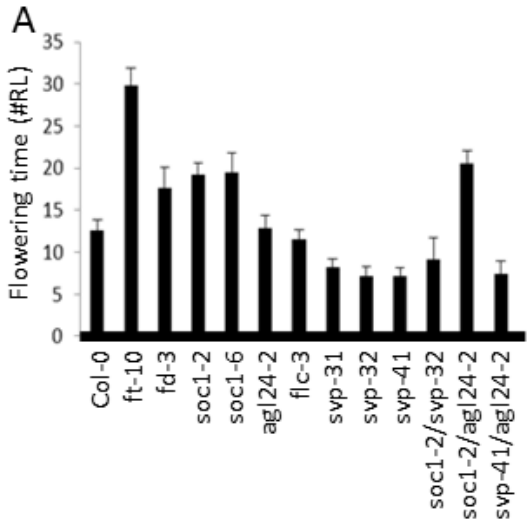


Figure 3.3. Flowering times measured by Valentim and colleagues for various mutants. Bars represent the average measurements with their standard deviations. Reprinted from Valentim *et al.*, 2015 (CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/legalcode).

3.2.5. Implementation of the ODE model of the floral transition

The ODE models of the floral transition were implemented in R, but dynamically translate the equations of the model into C++ and compile them for extra performance, using the inline package (Sklyar et al., 2015). Integration was performed with the deSolve package (Soetaert et al., 2016), using the LSODA algorithm.

Gene expression levels are predicted directly by the integration of the ODEs of the models for the Col-0 (WT) genotype. However, the models are also used to predict the gene expression levels of mutants. In that case, the initial values of

the mutated genes are set to 0 and the ODEs associated with those genes are replaced by the zero function.

The models can also yield another type of output: the flowering time of any mutant (or the WT). Flowering times are extrapolated from gene expression levels under the assumption used by Valentim and colleagues, i.e. flowering is marked by the expression of *AP1* above a certain threshold. Therefore, flowering was predicted to occur when *AP1* expression reached the value observed in the Col-0 background at flowering time (0.180 nM).

The parameters of the ODE models of the floral transition were optimized using the Robust Adaptive Metropolis (RAM) algorithm (Vihola, 2012), implemented by the adaptMCMC package for R (Scheidegger, 2012).

RAM is based on the Metropolis algorithm (Metropolis et al., 1953). The original Metropolis algorithm enables sampling from a probability distribution $\pi: x \mapsto \pi(x)$ (e.g. that of a vector of parameters) for which a quantity proportional to its function is known, without evaluating it at all nodes of a grid covering its definition domain, which would be extremely costly in high-dimensional spaces. The Metropolis algorithm works as follows:

1. An initial state X_1 (i.e. a vector of initial parameter in this case) values is chosen.
2. At iteration n, a proposal state Y_n is picked randomly from a symmetric distribution with density function $g: x \mapsto g(x|X_{n-1})$. This is often a normal distribution.

3. The proposal is accepted randomly, with an acceptance probability $\alpha_n = \min\left(1, \frac{\pi(Y_n)}{\pi(X_{n-1})}\right)$. If it is accepted, then $X_n = Y_n$, otherwise $X_n = X_{n-1}$.
4. Repeat from step 2.

The effectiveness of the sampling depends on the choice of g. The variance of g should not be too big, or the algorithm will tend to overshoot and propose states likely to result in a large loss of likelihood. Smaller steps have a lower risk of resulting in a large loss of likelihood, but if the variance is too small, the algorithm will take many steps – and therefore a lot of computational time – to explore the parameter space. RAM is an extension of the Metropolis algorithm that tackles the issue of choosing an adequate proposal distribution. In RAM, the distribution yielding x' is not fixed beforehand, but adapts based on the acceptance rate of previous proposals. Concretely, the algorithm changes as follows:

1. Initial values are chosen: a state $X_1 \in \mathbb{R}^d$ (where d is the number of parameters) and a lower triangular matrix $S_1 \in \mathbb{R}^{d \times d}$.
2. At iteration n, a proposal state Y_n is picked randomly as follows: $Y_n = X_{n-1} + S_{n-1}U_n$, where U_n follows a spherically symmetric distribution (e.g. a normal distribution where the covariance matrix is the identity matrix).
3. The proposal is accepted with probability $\alpha_n = \min\left(1, \frac{\pi(Y_n)}{\pi(X_{n-1})}\right)$. If it is accepted, $X_n = Y_n$, otherwise, $X_n = X_{n-1}$.

4. Using the Cholesky algorithm, the triangular matrix is updated as the lower triangular matrix with positive diagonal elements S_n satisfying: $S_n S_n^T = S_{n-1}\left(I + \eta_n(\alpha_n - \alpha^*)\frac{U_n U_n^T}{\|U_n\|^2}\right)S_{n-1}^T$, where $I \in \mathbb{R}^{d\times d}$ is the identity matrix, $\{\eta_n\}_{n\geq 1} \in (0, 1]$ is a step size sequence decaying to 0, and α^* is the target acceptance rate. In adaptMCMC, $\eta_n$is of the form n^{γ}, where $\gamma \in (0.5, 1]$. In this chapter, γ was set to 0.5001 and α^* to 0.234, the asymptotically optimal value for large number of parameters (Gelman et al., 1996).
5. Repeat from step 2.

Concretely, this means the proposal increments ($S_{n-1}U_n$) are not sampled from a spherically symmetric distribution like U_n, but from an elliptically symmetric distribution, to account for the correlation of parameters. The shape of that ellipsoid is updated at each iteration, by shrinking or expanding in the direction of $S_{n-1}U_n$, depending on the value of the acceptance rate relative to the target rate.

RAM can be used in a Bayesian way to sample the posterior distribution of the parameters of a model. In that case, the π function is a function of the joint prior distribution of the parameters, of the predictions, and of the data:

$$P(X|D) = \frac{P(X \cap D)}{P(D)} = \frac{P(X).P(D|X)}{P(D)}$$

$$\pi(X) = P(X).P(D|X)$$

$$\pi(X) \propto P(X|D)$$

Where X is the parameter vector, D is the data set, $P(X|D)$ is the joint posterior distribution of the parameters, $P(X)$ is the joint prior distribution of the parameters, $P(D|X)$ is the likelihood of the errors observed between the data and the model, with that combination of parameters, and $P(D)$ is an unknown constant representing the probability of observing this data set. The fact that $P(D)$ is unknown is not an issue, because π simply needs to be proportional to the posterior distribution, which is the target.

In this chapter, the choice for $P(X)$ is a product of independent marginal distributions. Most marginal distributions are uniform distributions on $(0, M_i]$, where the M_i are chosen depending on the type of parameters they apply to. Model equations are built around Hill equations and exponential degradation terms and typically look like $\frac{dX}{dt} = \beta \frac{Y^n}{K^n + Y^n} - d.X$, where X and Y are modelled species, and β, K, n and d are parameters. The upper bounds for β, K and n parameters were chosen as 10^9, 100 times the maximum value of Y, and 10, respectively. For d-type parameters, the upper bound is 10^9 when the parameters are optimized without prior information. In the later models though, prior information from a study of mRNA half-lives was taken into account (Narsai et al., 2007), leading to the replacement of the uniform prior for degradation parameters by a normal distribution $\mathcal{N}(1.000, 0.392)$ matching the reported average value and confidence intervals.

$P(D|X)$ is a function of the model predictions and the observed data, for two types of measurements: gene expression and flowering time.

$$P(D|X) = P(D_g|X).P(D_f|X)$$

$$= \left(\prod_{i=1}^{n_g}\prod_{t=1}^{n_t} P(D_{g_{it}}|X)\right).\prod_{j=1}^{n_f} P\left(D_{f_j}\middle|X\right)$$

$$= \left(\prod_{i=1}^{n_g}\prod_{t=1}^{n_t} \phi_{0,\sigma_{g_{it}}^2}\left(D_{g_{it}} - m_{g_i}(A_{it}, X)\right)\right)\cdot\prod_{j=1}^{n_f} \phi_{0,\sigma_f^2}\left(D_{f_j} - m_f(B_j, X)\right)$$

Where D_g and D_f represent the gene expression and flowering time measurements, respectively, n_g is the number of genes, n_t is the number of observations per gene, n_f is the number of flowering time measurements, ϕ is the normal probability density function, $\sigma_{g_{it}}^2$ and σ_f^2 are the variances of the $D_{g_{it}}$ and D_{f_j}, respectively, m_{g_i} and m_f are functions representing how the model computes the expression level of gene i and flowering time, respectively, and A_{it} and B_j are input data used to compute the expression level of gene i at time t and flowering time of plant j, respectively.

The $\sigma_{g_{it}}$ are derived from the data.

$$\sigma_{g_{it}} = s_{g_{it}} + 0.01\left(\max_{t\in[\![1,n_t]\!]} D_{g_{it}} - \min_{t\in[\![1,n_t]\!]} D_{g_{it}}\right)$$

Where $s_{g_{it}}$ is the empirical standard deviation of the data. The added 1% of the range of the measurements is to account for measurements where the empirical standard deviation is 0 (typically, when the measurements themselves are all 0).

For flowering times, simply using the empirical standard deviations resulted in the flowering time data being neglected in favour of the expression data.

Therefore, they were replaced by a common σ_f proportional to the range of measurements:

$$\sigma_f = 0.001\left(\max_{i\in[\![1,n_i]\!]} D_{f_i} - \min_{i\in[\![1,n_i]\!]} D_{f_i}\right)$$

The 0.001 value was chosen by trial and error. Higher values (e.g. 0.01) also resulted in the flowering time data being mostly ignored in favour of minor fit improvements on the expression data. This was possibly because low-variance gene expression data points had too much weight.

Using the formula of ϕ (where π is not a distribution, but the usual $\pi = 3.14\ldots$):

$$\phi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

It comes that

$$P(D|X)$$

$$\propto \left(\prod_{i=1}^{n_g}\prod_{t=1}^{n_t}\exp\left(\frac{\left(D_{g_{it}} - m_{g_i}(A_{it},X)\right)^2}{2\sigma_{g_{it}}^2}\right)\right)\prod_{j=1}^{n_f}\exp\left(\frac{\left(D_{f_j} - m_f(B_j,X)\right)^2}{2\sigma_f^2}\right)$$

Let

$$U(X)$$

$$= \left(\prod_{i=1}^{n_g}\prod_{t=1}^{n_t}\exp\left(\frac{\left(D_{g_{it}} - m_{g_i}(A_{it},X)\right)^2}{2\sigma_{g_{it}}^2}\right)\right)\prod_{j=1}^{n_f}\exp\left(\frac{\left(D_{f_j} - m_f(B_j,X)\right)^2}{2\sigma_f^2}\right)$$

Based on the formula of $P(X|D)$, $P(X|D) \propto P(X).U(X)$.

Instead of working with $U(X)$ as a product of exponentials, it is computationally cheaper to work with its logarithm.

$$\ln U(X) = \sum_{i=1}^{n_g} \sum_{t=1}^{n_t} \frac{\left(D_{g_{it}} - m_{g_i}(A_{it}, X)\right)^2}{2\sigma_{g_{it}}^2} + \sum_{j=1}^{n_f} \frac{\left(D_{f_j} - m_f(B_j, X)\right)^2}{2\sigma_f^2}$$

Let

$$\pi'(X) = P(X).U(X)$$

$$\ln \pi'(X) = \ln P(X) + \ln U(X)$$

$\ln \pi'$ is the argument expected by the adaptMCMC package.

3.3. RESULTS

The first subproject was the possibility of integrating the vernalization pathway into an ODE model of the floral transition.

3.3.1. An ODE model of vernalization was able to reproduce the quantitative regulation of *FLC* expression at the tissue scale

Angel and colleagues developed a stochastic model of the silencing of *FLC* during vernalization (Angel et al., 2011). This model focused on demonstrating that the silencing of *FLC* at the plant scale resulted from a stochastic process, whereby individual cells decided to repress *FLC* in an independent, all-or-nothing way. Quantitative variations in the expression of *FLC* at the plant scale therefore depend on the proportions of “on” cells and “off” cells, as opposed to synchronous, quantitative variations in the expression levels of all cells.

Angel and colleagues modelled the histones of the *FLC* locus individually. During the simulation, histones have probabilities of switching between three states: unmodified, methylated and activating. Methylated histones silence the expression of *FLC*, while activating ones enhance it. The methylated and activating states can spread, i.e. their presence increases the probabilities of other histones to switch to their respective states and away from their opposite states, which creates regulatory loops conducive to a bistable system at the cell level. The transition probabilities are also affected by whether they belong to a special region of the locus or not (the nucleation region) and temperature. Initially, most histones are in the activating state, and remain so, due to a slight bias of transition probabilities toward activation. When the plant is exposed to cold, the probability of histone methylation increases at the nucleation region, thereby making the histones of the nucleation region more likely to be methylated. When the plant returns to warm conditions, there may be enough methylated histones in the nucleation region to shift the bias toward methylation in the rest of the locus, resulting in the propagation of histone methylation to the whole locus. Therefore, individual cells either silence *FLC* completely or not at all. This stochastic model can however still explain the fact that the overall *FLC* expression of a plant decreases gradually with the duration of the exposure to cold, because this duration does affect the methylation of the nucleation region and therefore the probability that a cell will silence *FLC*. At the plant or tissue level, the expression of all cells are

averaged, so the overall expression level is determined by the proportion of silenced cells.

In order to avoid running stochastic simulations of cell populations to feed their results into an ODE model, the possibility of directly modelling the silencing of *FLC* at the plant or tissue level using ODEs was studied.

An ODE model inspired by the mechanism proposed by Angel and colleagues was developed (Equation 3.1). It only considers two histone states: methylated and unmodified, as this was sufficient to create a system where methylation levels are stable before and after cold treatment. It relies on the following principles:

1. Two regions are considered within the *FLC* locus: the nucleation site and the distal region. At the beginning of the simulation, they are both unmethylated ($m_{nucleation} = 0$ and $m_{distal} = 0$).
2. In warm weather, the histones of both regions remain unmethylated:
    a. $\frac{d(m_{nucleation})}{dt} = 0$, because $cold = 0$.
    b. $\frac{d(m_{distal})}{dt} = 0$, because $m_{nucleation}$=0.
3. During a period of cold ($cold = 1$), the histones of the nucleation site get methylated, both spontaneously (due to ϵ) and as the result of the spreading of the methylation marks (due to $m_{nucleation} \cdot (1 - m_{nucleation})$). The spreading term is a logistic function because it requires both the presence of methylated and unmethylated histones in the nucleation zone to occur. Methylation also starts spreading from

the nucleation site to the distal region, which is represented by a product of three terms: $m_{nucleation}.(1-m_{distal}).(1-\beta.cold)$. The first two factors are because the spreading of methylation marks from the nucleation zone to the distal zone requires both methylated histones in the nucleation zone and unmethylated histones in the distal zone. and warmth. The third factor is because that spreading is faster in warm weather than in cold weather ($0<\beta<1$).

4. After returning to warm weather ($cold=0$), methylation stops at the nucleation site, and the level of methylation sustains itself, therefore no degradation term was added to the equation of $\frac{d(m_{nucleation})}{dt}$. Meanwhile, the spreading of methylation from the nucleation zone to the distal zone picks up, as it is facilitated by warmth. However, the methylation of the distal zone reaches a steady state that is not always complete methylation, but is a non-decreasing function of the duration of the exposure to cold. Therefore, a degradation term was added to the equation of $\frac{d(m_{distal})}{dt}$.

The expression of *FLC* was modelled as inhibited by cold and the methylation of the distal region. A delay was introduced into the effect of temperature ($1-cold(t-\delta)$), to prevent a transient expression of *FLC* after returning to warm conditions, during the time it takes for methylation marks to spread from the nucleation region to the distal region. The effect of distal methylation was

raised to the power of n ($\left(1 - m_{distal}(t)\right)^{n}$) to sharpen the response of *FLC* expression to variations in m_{distal}.

Similarly to Angel and colleagues' model, the ODE model was able to capture that short cold exposures barely affect the expression level of *FLC*, but longer ones do lead to its repression. The intensity of the repression increases with the duration of the exposure, until saturation is reached (Figure 3.4).

Equation 3.1. ODE model of FLC silencing.

$$\frac{d(m_{nucleation})}{dt} = cold.(\alpha_{nucleation}.m_{nucleation}.(1 - m_{nucleation}) + \epsilon)$$

$$\frac{d(m_{distal})}{dt} = \alpha_{distal}.m_{nucleation}.(1 - m_{distal}).(1 - \beta.cold) - k.m_{distal}$$

$$FLC(t) = \big(1 - cold(t - \delta)\big).\big(1 - m_{distal}(t)\big)^{n}$$

- $m_{nucleation}$: proportion of methylated histones at the nucleation site of FLC
- m_{distal}: proportion of methylated histones in the distal region of FLC
- t: time
- $cold \in \{0, 1\}$: whether the current temperature is under a certain threshold or not
- $\alpha_{nucleation}$: propagation rate of histone methylation rate at the nucleation site in cold weather
- ϵ: rate of spontaneous histone methylation at the nucleation site in cold weather
- α_{distal}: propagation rate of histone methylation in the distal region
- $\beta \in [0, 1]$: inhibition coefficient for the propagation of histone methylation in cold weather
- k: histone demethylation rate in the distal region
- δ: delay in the effect of temperature on FLC expression (d)
- n: coefficient regulating the stiffness of the response of FLC expression to the methylation of the distal region

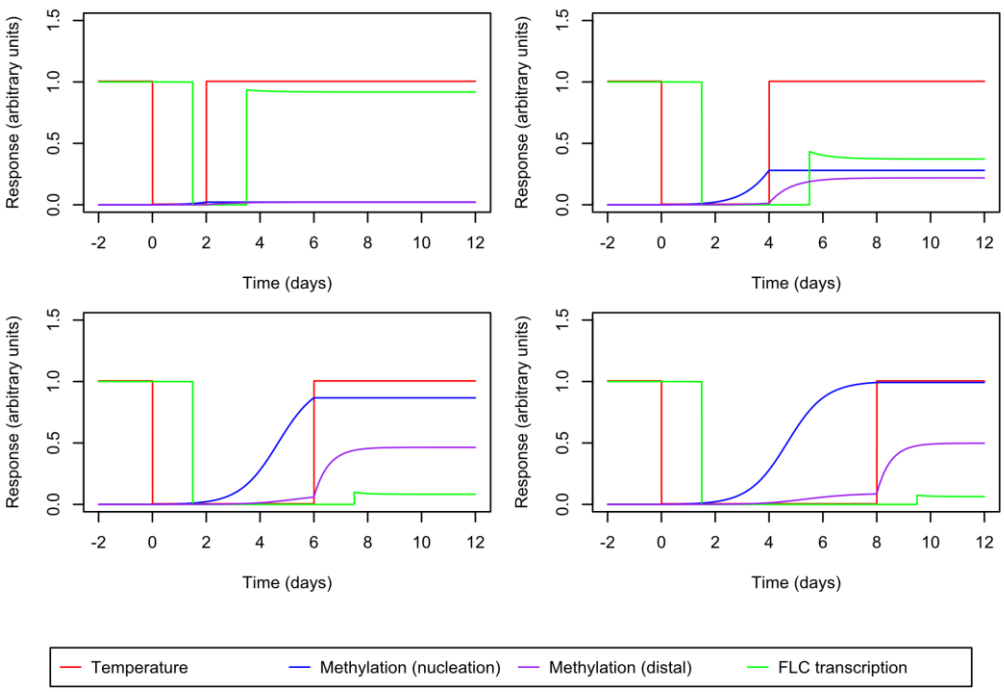


Figure 3.4. Gradual silencing of FLC expression in response to varying durations of exposure to cold. For short durations of exposure to cold, virtually no reduction in FLC expression occurs. As the duration of exposure increases, the expression level of FLC after returning to warm temperatures decreases, until that effect reaches saturation. Parameter values: $\epsilon = 10^{-4}$, $\alpha_{nucleation} = 0.07$, $\alpha_{distal} = 0.05$, β =0.9, $k = 0.05$, $\delta = 1.5\ d$, $n = 4$.

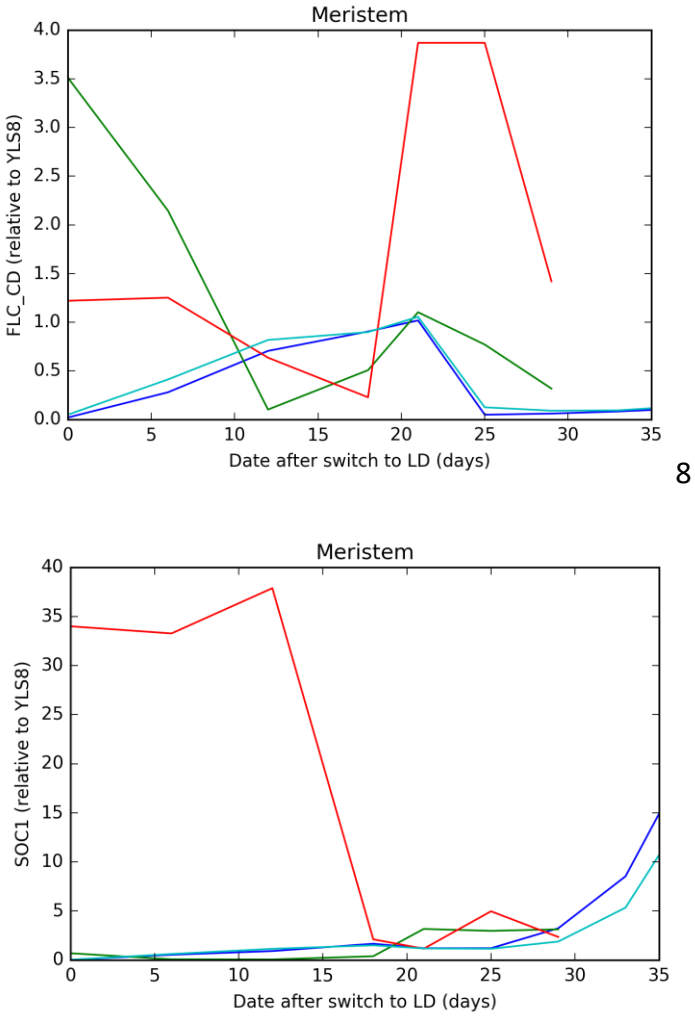
This demonstrates that, even though the methylation of histones at the *FLC* locus is a stochastic process in nature, a deterministic ODE model can represent it accurately at the tissue level. It is therefore not necessary to model vernalization at the cell level to study its effect on the floral transition pathway.

The original plan was to integrate this model of vernalization into an ODE model of the floral transition developed from scratch. However, it appeared

through a collaboration with Wageningen University that an ODE model of the floral transition in *A. thaliana* was already being developed by Valentim and colleagues, from Aalt-Jan van Dijk's team (Valentim et al., 2015). The plan therefore shifted to expanding Valentim and colleagues' model to include the effects of vernalization. An attempt to generate data suitable to fit such a model was part of this effort.

3.3.2. The measurement of the effects of vernalization on gene expression in the floral transition network was affected by repeatability issues

Samples harvested from plants that had undergone a 3-week cold treatment were subjected to qRT-PCR analyses on four separate occasions by Enza, a partner company of the University of Wageningen., however the results exhibited so much variability that the actual kinetics of genetic expression could not be ascertained. The measurements are shown for two genes, *FLC* and *SOC1*, as examples of the observed variability (Figure 3.5) These genes were selected as examples because their theoretical behaviour during and after a cold treatment is known: FLC should be silenced by the cold treatment and remain silent thereafter, while SOC1 is normally upregulated during the floral transition, which should happen after the cold treatment, if at all. As these expected behaviours cannot be seen consistently in the results, these experiments were considered unsuccessful.



8

Figure 3.5. *FLC* and *SOC1* expression levels relative to *YLS8* in the meristems, as measured by four Fluidigm qRT-PCRs. Each line colour corresponds to a different experiment. All four qRT-PCRs were carried out by Enza on the same biological samples. The cold treatment occurred from day 0 to day 21. After day 21, the plants were grown in warm conditions.

3.3.3. Cold treatment decreases flowering time

Studying the effect of cold treatments on flowering time is not straight-forward because a 4˚C cold treatment all but halts the development of

seedlings, at least from a morphological point of view. It is unclear whether development at the molecular level (gene expression profiles) is affected as much as at the macroscopic level (size of the plant and number of leaves), but it seems plausible and is commonly accepted, hence the wide-spread practice of counting flowering time in number of rosette leaves instead of actual time. In this study, the notion of adjusted bolting time was introduced to respond to the same concern, while keeping measurements of flowering time in a standard time unit. It is defined as the bolting time (from sowing to bolting), minus the duration of the cold treatment.

Results show that increasing durations of cold treatment lead to reductions in bolting time and adjusted bolting time (Table 3.2) The cold treatment required to saturate the vernalization response does not seem to have been reached, and this might be why flowering occurred later than expected. This unfortunately resulted in the pool of plants allocated to dissection being depleted before flowering could be observed macroscopically, as the number of plants to grow for the experiment had been calculated under the assumption that the plants undergoing a 3-week cold treatment would flower in 49 days at most (i.e. 14 days in the last chamber). Another argument in favour of stopping the harvesting was that the floral transition might have already occurred at the molecular (gene expression) level.

While the experiments were being carried out, an analysis of Valentim and colleagues' model was also made. Those results are presented in the next section.

Table 3.2. Influence of vernalization on the flowering time of FRI+ Col-0 plants. Bolting time values are given as average ± standard error.

Cold-treatment (weeks)	Bolting time (days after sowing)	Bolting time (days after induction)	Adjusted bolting time (days)	Number of individuals
0	90.64 ± 0.70	76.64 ± 0.70	90.64 ± 0.70	45
1	76.05 ± 1.68	62.05 ± 1.68	69.05 ± 1.68	19
2	75.25 ± 0.78	61.25 ± 0.78	61.25 ± 0.78	40
3	68.11 ± 0.22	54.11 ± 0.22	47.11 ± 0.22	90

3.3.4. Valentim and colleagues' model of the floral transition can be simplified

At first glance, Valentim and colleagues' model (Valentim et al., 2015) seemed very complex with respect to its predictions, therefore a complexity reduction was attempted. This attempt was supported by the finding in Chapter 2 that gene regulation equations can involve very few parameters, while the original model (Valentim et al.'s) featured 35 parameters.

The original model was fitted to two types of data: time series of gene expression and flowering time measurements. However, the parameters retained by Valentim and colleagues fitted the flowering time data rather poorly (Figure 3.6), because *AP1* is overestimated (Figure 3.2). Reoptimizing the parameters with RAM resulted in better fits (Figure 3.7 and Figure 3.8), but also revealed other issues.

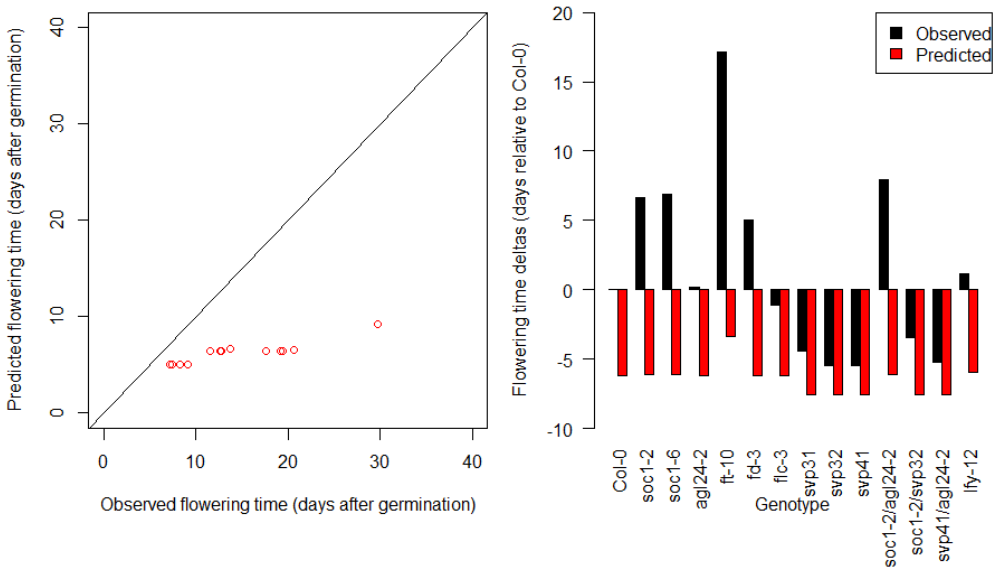


Figure 3.6. Predicted flowering times using the original model with the reported parameters. Flowering time is underestimated for all genotypes.

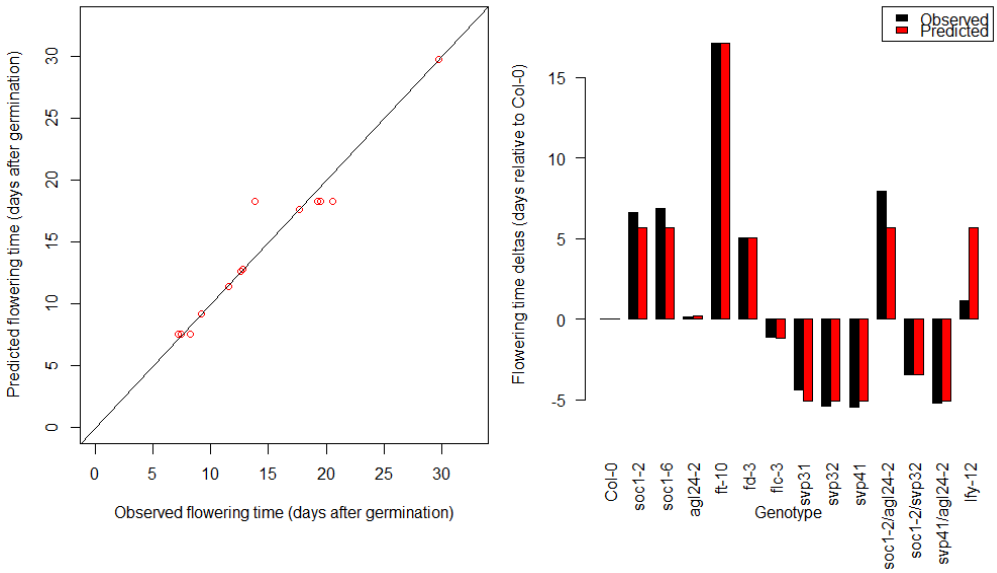


Figure 3.7. Predicted flowering times using the original model reoptimized with RAM. Flowering times are well predicted, except for *lfy-12*.

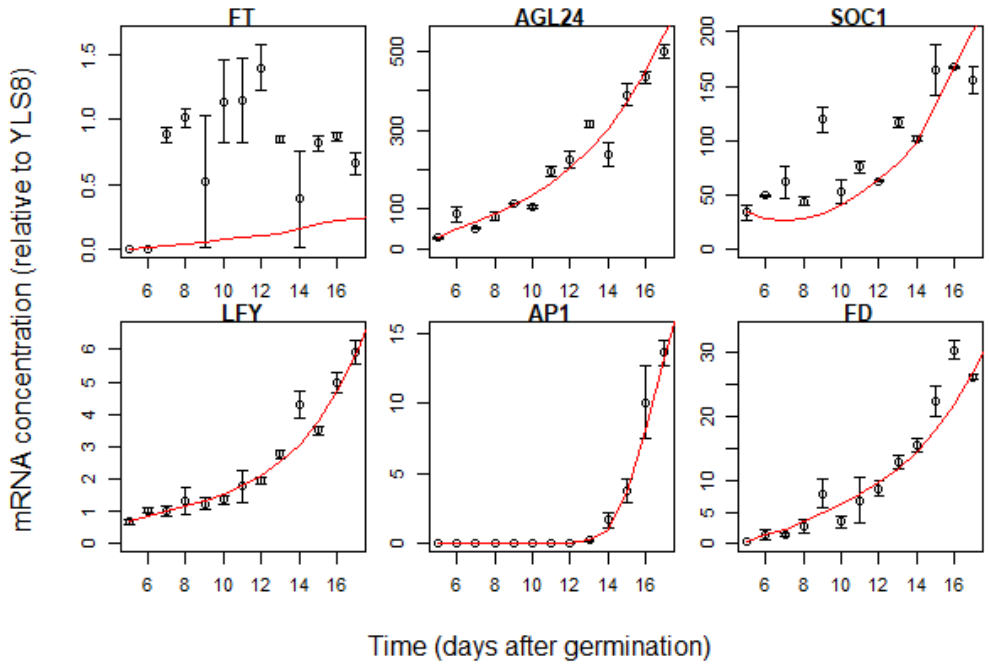


Figure 3.8. Predicted gene expression using the original model reoptimized with RAM. Gene expression is well predicted, except for *FT*.

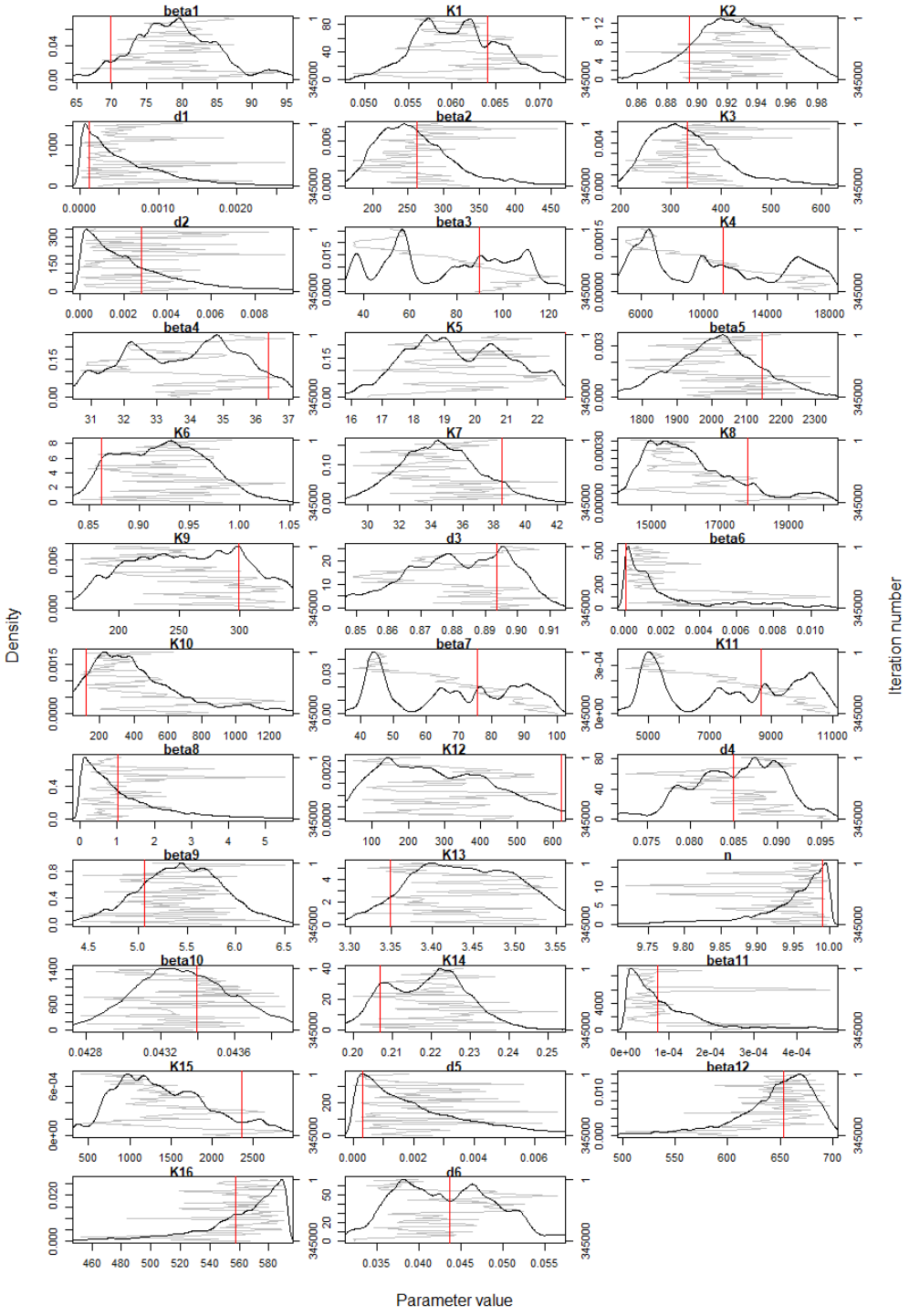


Figure 3.9. Distribution of the parameters of the original model sampled by RAM. The grey line is the trace of the values sampled by RAM. The black line is the kernel estimation of the density. The red line is the optimal value.

Table 3.3. Optimal values of the parameters of the original model.

Parameter	Value
beta1	69.88392
beta2	262.571
beta3	90.02974
beta4	36.38466
beta5	2144.1
beta6	7.70E-05
beta7	75.60066
beta8	1.024824
beta9	5.06396
beta10	0.043398
beta11	7.49E-05
beta12	652.836
d1	0.000122
d2	0.002795
d3	0.893741
d4	0.084881
d5	0.000333
d6	0.0436
K1	0.064031
K2	0.895105
K3	333.3299
K4	11211.01
K5	2.29E+01
K6	0.862488
K7	38.46168
K8	17808.81
K9	298.9328
K10	117.813
K11	8681.21
K12	620.6245
K13	3.35E+00
K14	0.206951
K15	2369.714
K16	557.6477
n	9.989972

First, out of the six degradation coefficients, three have near 0 values (Figure 3.9, Table 3.3). This seems at odds with a study that measured the half-lives of several mRNAs including *SOC1*, which was found to have a degradation

coefficient of 1.000 d^{-1} (95% confidence interval: [0.24, 1.78]) (Narsai et al., 2007). As none of the other modelled mRNAs were investigated in that study, the gene-specific degradation parameters d_i were replaced by a single parameter d applying to all genes. A prior distribution $\mathcal{N}(1.000, 0.392)$ (corresponding to the reported confidence interval) was used for d.

Second, many K_i coefficients (K_3,K_{12}, K_{14}, K_{15} and K_{16}) of Hill equations for activations ($x \mapsto \frac{x^n}{K_i^n + x^n}$) had values over 8 times the maximum observed value of x. This means those Hill equations actually became quasi-polynomial (or quasi-linear if n=1) in the domain corresponding to the simulation. They were therefore replaced by polynomial or linear equations accordingly.

Third, the K_i coefficients of some Hill equations for inhibitions ($x \mapsto \frac{K_i^n}{K_i^n + x^n}$) had values resulting in fold changes of less than 1.08 in the range of x observed. This was the case of K_8 and K_9. The corresponding Hill equations were therefore replaced by 1.

Fourth, the distributions of some β_i parameters are "leaning" on 0, showing their associated effects are negligible. This is the case of β_6 (AGL24 → LFY) and β_{11} (FD → AP1). Those effects were therefore removed from the equations.

Finally, some mutations (*flc*, *agl24* and *lfy*) have little or no effect on flowering time, according to the data. Therefore, *LFY* was removed from the equation of *AP1*. However, this would make *AP1 SOC1*-independent, and the *soc1* mutation is known to have a late-flowering effect, so a term similar to the *LFY*

one was reintroduced using *SOC1* instead. *LFY* was also removed from the equation of *FD*, because since *LFY* was the only regulator of *FD*, a *lfy* mutation should have resulted in the complete silencing of *FD*, but the *fd* mutation has a stronger late-flowering effect than the *lfy* one. As no regulator of *FD* remained, *FD* was no longer modelled, and interpolated measurements were used as inputs for the other equations of the model. As *agl24* has nearly no effect on flowering time, the effects of *AGL24* were removed from the equations of *SOC1*, which is upstream of *AP1* (it had already been removed from the equation of *LFY*). Finally, the effect of *FLC* was removed from the equations of *FT* (it had already been removed from the equation of *SOC1*). However, *FT* only had one incoming regulation left – a repression by *SVP* – and the SVP measurements in leaves (Figure 3.1) cannot explain the upregulation of *FT* observed in the data. Therefore, a time-dependent term representing regulations by unknown species was added to the equation of *FT*.

These changes are summarized in Table 3.4. The predictions of Model 1 are shown in Figure 3.10 and Figure 3.11. The optimal parameter values (likelihood-wise) are in Table 3.5.

Table 3.4. Changes between the original model and Model 1. Changes are highlighted in red.

Derivative	Original model	Model 1
$\frac{dFT}{dt}$	$\beta_1 \frac{K_1}{K_1 + SVP_{leaf}} \cdot \frac{K_2}{K_2 + FLC_{leaf}} - d_1.FT$	$\beta_1 \frac{K_1}{K_1 + SVP_{leaf}} \cdot \frac{t^{n_t}}{K_t^{n_t} + t^{n_t}} - d.FT$
$\frac{dAGL24}{dt}$	$\beta_2 \frac{SOC1}{K_3 + SOC1} - d_2.AGL24$	$\beta_2.SOC1 - d.AGL24$
$\frac{dSOC1}{dt}$	$\left(\beta_3 \frac{AGL24}{K_4 + AGL24} + \beta_4 \frac{SOC1}{K_5 + SOC1} + \beta_5 \frac{FT}{K_6 + FT} \frac{FD}{K_7 + FD}\right) \cdot \frac{K_8}{K_8 + SVP} \frac{K_9}{K_9 + FLC} - d_3.SOC1$	$\beta_4 \frac{SOC1}{K_5 + SOC1} + \beta_5 \frac{FT}{K_6 + FT} \frac{FD}{K_7 + FD} - d.SOC1$
$\frac{dLFY}{dt}$	$\beta_6 \frac{AGL24}{K_{10} + AGL24} + \beta_7 \frac{SOC1}{K_{11} + SOC1} + \beta_8 \frac{AP1}{K_{12} + AP1} - d_4.LFY$	$\beta_7 \frac{SOC1}{K_{11} + SOC1} + \beta_8.AP1 - d.LFY$
$\frac{dAP1}{AP1}$	$\beta_9 \frac{LFY^n}{K_{13}^n + LFY^n} + \beta_{10} \frac{FT}{K_{14} + FT} + \beta_{11} \frac{FD}{K_{15} + FD} - d_5.AP1$	$\beta_9 \frac{SOC1^n}{K_{13}^n + SOC1^n} + \beta_{10}.FT - d.AP1$
$\frac{dFD}{dt}$	$\beta_{12} \frac{LFY}{K_{16} + LFY} - d_6.FD$	

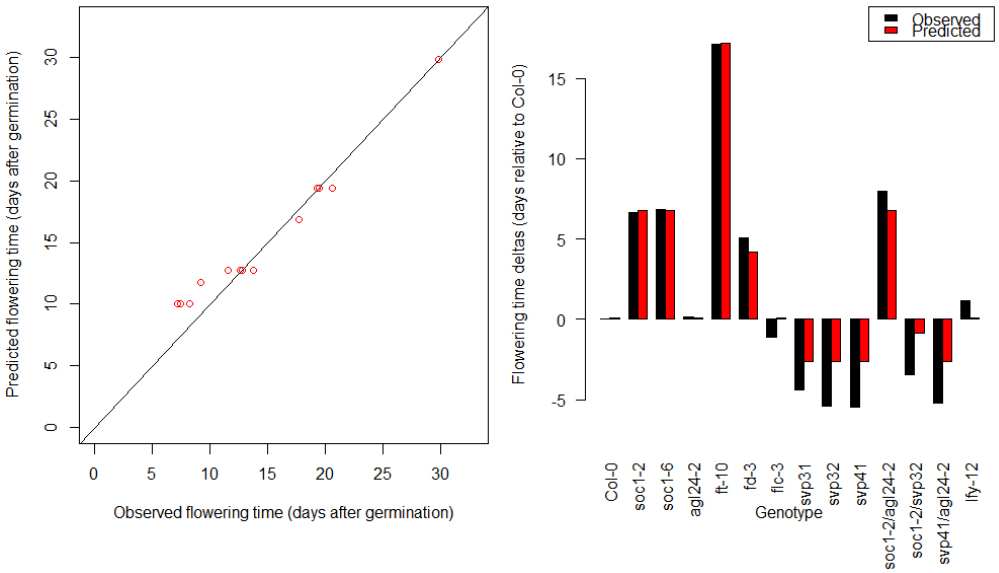


Figure 3.10. Flowering times predicted by Model 1. Flowering times are overall well predicted, but the effect of the *svp* mutation are underestimated.

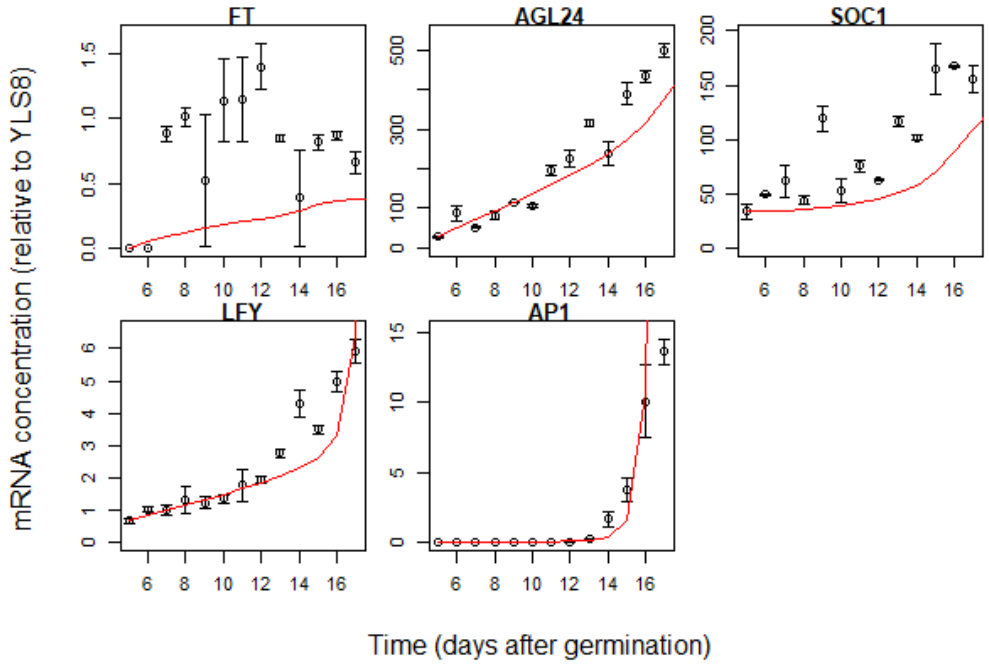


Figure 3.11. Gene expression predicted by Model 1. The predictions are not too far from the data points, but the curvatures of the predictions are off.

Table 3.5. Optimal parameter values for Model 1.

Parameter	Value
beta1	0.142997
beta2	0.748411
beta4	515.1157
beta5	200220.1
beta7	37.20443
beta8	0.074734
beta9	1E+09
beta10	0.067994
d	0.050156
K1	73.02524
K5	8567.636
K6	81.58542
K7	1256.276
K11	6001.799
K13	597.4586
Kt	3.95E-10
n	9.213486
nt	0.348851

The simplifications introduced in Model 1 fixed the issue with the flowering time of the *lfy* mutant. However, they also introduced other issues. Gene expression fits are overall poorer (Table 3.8), especially for *AP1*, whose final measurement is extremely overestimated due to the steepness of the predicted curve, resulting in an NMRSE of 140%. The flowering times of the various *svp* mutants were also overestimated. As the *svp* mutation results in the upregulation of *FT*, this suggested that the effect of *FT* was underestimated in that context.

To improve the fit of the *AP1* and flowering time predictions, the FT term was raised to the power of n_2, to provide more flexibility in the regulation of *AP1*. The effect of *AP1* on *LFY* was removed, as it seemed to negatively affect the curvature of the *LFY* prediction and did not seem required, based on the gene

expression and flowering time data. Finally, all remaining K_i coefficients were removed, as they were all over 10 times the maximum expression value observed for their respective regulators. All these changes were implemented into Model 2.

The predictions of Model 2 are shown in Figure 3.12 and Figure 3.13. The optimal parameter values are in Table 3.7.

Model 2 only contains 13 parameters, yet rivals the original model in terms of goodness of fit (Table 3.8). It therefore appears to be a good replacement.

Interestingly, it could even be simplified further if flowering time data are ignored, as is shown in the following section.

Table 3.6. Changes between Model 1 and Model 2. Changes are highlighted in red.

Derivative	Model 1	Model 2
$\frac{dFT}{dt}$	$\beta_1 \frac{K_1}{K_1 + SVP_{leaf}} \cdot \frac{t^{n_t}}{K_t^{n_t} + t^{n_t}} - d.FT$	$\beta_1 \frac{K_1}{K_1 + SVP_{leaf}} \cdot \frac{t^{n_t}}{K_t^{n_t} + t^{n_t}} - d.FT$
$\frac{dAGL24}{dt}$	$\beta_2.SOC1 - d.AGL24$	$\beta_2.SOC1 - d.AGL24$
$\frac{dSOC1}{dt}$	$\beta_4 \frac{SOC1}{K_5 + SOC1} + \beta_5 \frac{FT}{K_6 + FT} \frac{FD}{K_7 + FD} - d.SOC1$	$\beta_4.SOC1 + \beta_5.FT.FD - d.SOC1$
$\frac{dLFY}{dt}$	$\beta_7 \frac{SOC1}{K_{11} + SOC1} + \beta_8.AP1 - d.LFY$	$\beta_7.SOC1 - d.LFY$
$\frac{dAP1}{AP1}$	$\beta_9 \frac{SOC1^n}{K_{13}^n + SOC1^n} + \beta_{10}.FT - d.AP1$	$\beta_9.SOC1^{n_1} + \beta_{10}.FT^{n_2} - d.AP1$

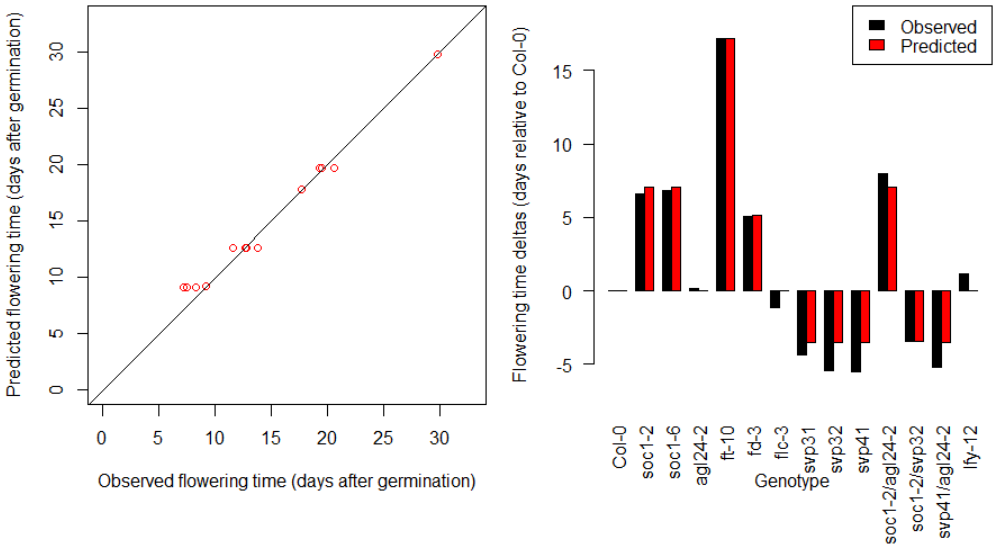


Figure 3.12. Flowering time predictions of Model 2. Flowering times are overall well predicted, although the effects of the *svp* mutation is still slightly underestimated.

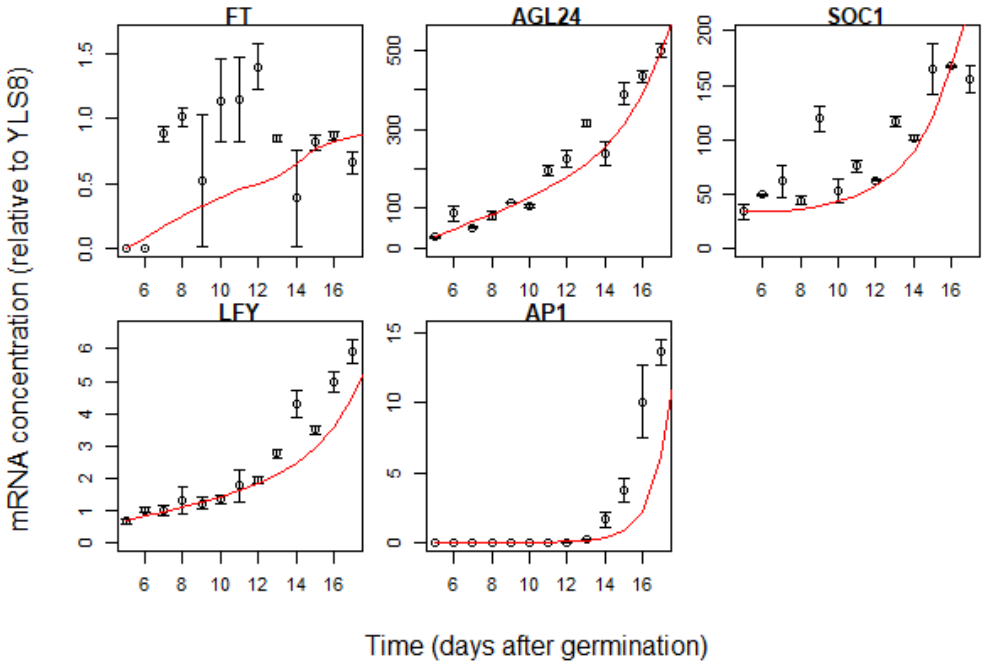


Figure 3.13. Gene expression predictions of Model 2. Gene expression is overall well predicted, though AP1 is slightly underestimated past 13 days (after flowering has occurred).

Table 3.7. Optimal parameter values for Model 2.

Parameter	Value
beta1	0.293154
beta2	0.6502
beta4	0.043441
beta5	2.251851
beta7	0.005528
beta9	7.30E-08
beta10	1.46E-01
d	0.040861
K1	83.66373
Kt	4.849048
n1	3.377892
n2	9.980918
nt	9.967795

Table 3.8. Normalised root-mean square error (NRMSE) values of modelled genes and flowering time, for all fitted models.

Variable	Original model with reported parameters	Original models reoptimized with RAM	Model 1	Model 2
FT	27%	54%	47%	35%
SOC1	19%	25%	37%	27%
AGL24	7%	7%	14%	10%
LFY	7%	7%	15%	15%
AP1	14%	4%	140%	23%
Flowering time	42%	6%	7%	5%

3.3.5. The gene expression time series alone do not contain enough information to determine the topology of the regulatory network

Valentim and colleagues' model is sensible from a biological point of view. The interactions it involves are supported by biological evidence and are modelled by Michaelis-Menten or Hill equations to account for the existence of upper limits on synthesis rates. However, the fact that it could be simplified drastically (from 35 to 13 parameters) while retaining nearly the same goodness of fit as the original shows there are not enough data to fit such a complex model. The time series are particularly uninformative, since they do not exhibit any deceleration of gene expressions and therefore provide no way of estimating the parameters associated with the plateaux of the Hill and Michaelis-Menten functions. To illustrate this, a deliberately uninformative

ODE model, where *SOC1*, *AGL24*, *LFY* and *AP1* are each exclusively self-activated ($\frac{dx_i}{dt} = a_i x_i, a_i \in \mathbb{R}^+$), was fitted to their time series. The resulting fits were once again good (Figure 3.14), showing that the time series of the key floral transition genes are essentially exponential curves and therefore contain very little information, as it would be possible to build models with completely erroneous topologies if it were not for the prior biological evidence available in the literature. One could for instance build a model by picking a random activator for each of *SOC1*, *AGL24*, *LFY* and *AP1*. *FT* is the only gene that does not suffer from this issue.

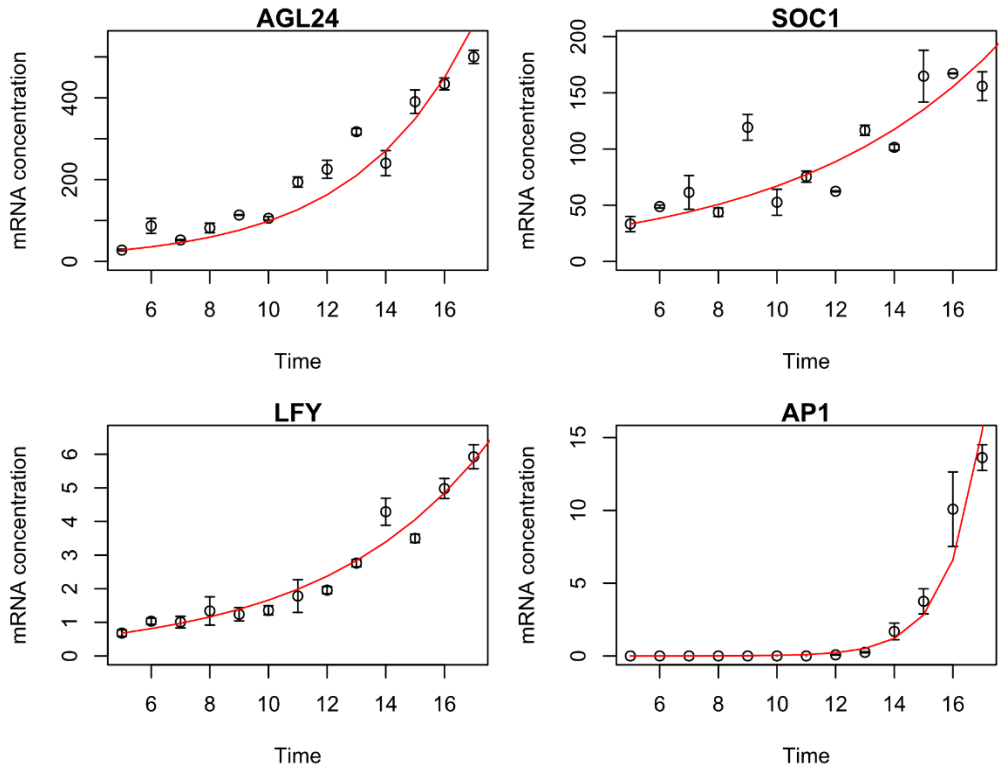


Figure 3.14: Fit of the deliberately uninformative self-activation model. *In this model, AGL24, SOC1, LFY and AP1 are only up-regulated by themselves, resulting in exponential growths of their expression levels. This overly simplistic, biology-unrelated model still provides adequate fits, which indicates that the time series are not very informative.*

It was apparent in Figure 3.14 that some time series look very similar to each other. To assess the matter in a more quantitative way, the correlations for each pair of genes were computed (Figure 3.15).

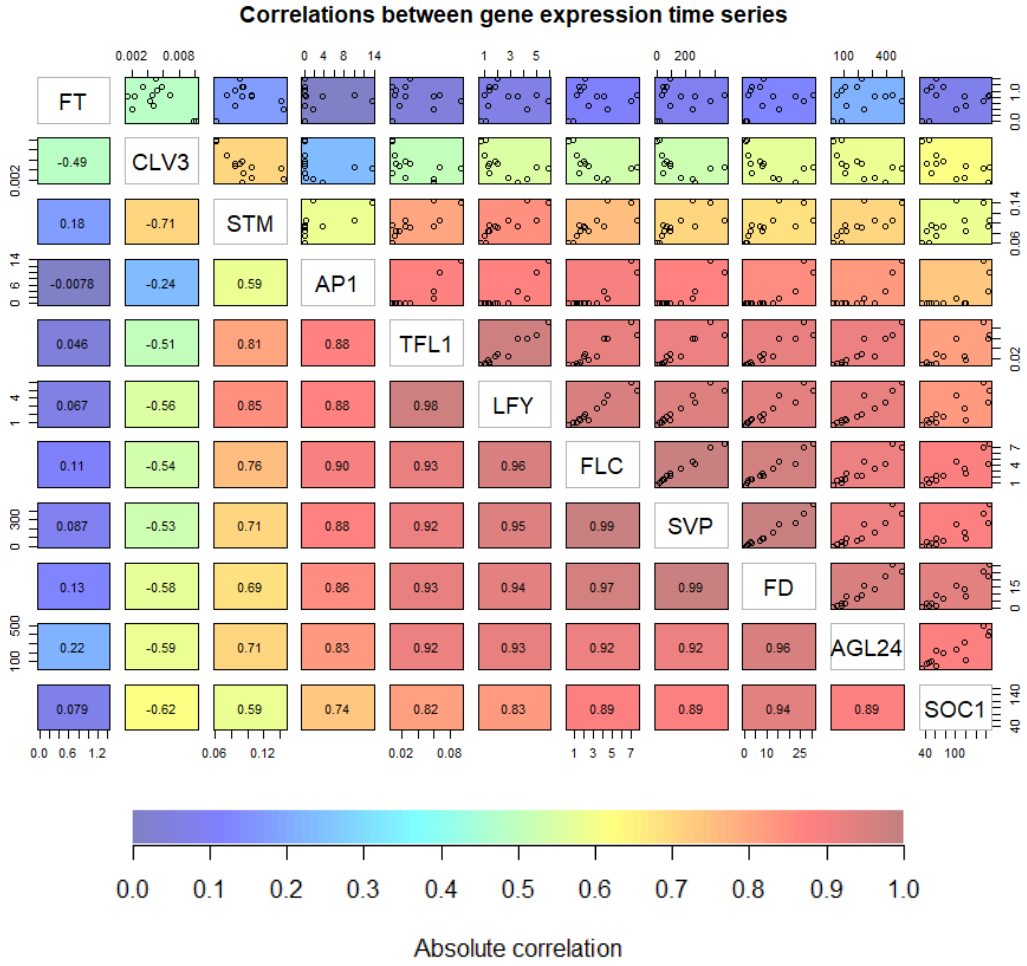


Figure 3.15. Correlations of time series in the WT. *Background colours indicate absolute values of correlation.*

Genes *CLV3*, *STM* and *TFL1* are not part of any of the models studied, but were added to the set for two reasons. *STM* and *CLV3* were included because they are meristem markers, and *TFL1* is a candidate gene to be included into an expanded model, as it is a key repressor of the floral transition. The expression of these three genes was measured on the biological samples used by Valentim and colleagues.

AP1, *TFL1*, *LFY*, *FLC*, *SVP*, *FD*, *AGL24* and *SOC1* form a cluster of mutually highly correlated genes. The cause of these correlations is still unclear, however, two hypotheses have been proposed:

1. These genes are genuinely all up-regulated by a common factor. This makes sense for meristem genes (*SOC1*, *AGL24*, *LFY*, *AP1*, *TFL1*, *FD*), which are all under the control – direct or indirect – of *SOC1*. However, *FLC* and *SVP* expression are also correlated to those of the previously mentioned genes, which is unexpected, as they are supposed to inhibit them.
2. The correlated time series are artefacts resulting from the method of data acquisition. Three reasons have been considered.
    a. *YLS8* is unsuitable as a reference gene, because its expression level or the size of its expression domain decreases during development, causing an apparent increase in the expression levels of the genes of interest.
    b. The proportions of the expression domains of the genes of interest with respect to the size of the biological tissue sampled increases over time. This could be due to the fact that the subset of the tissue actually expressing the genes of interest is small compared to the total size of the sample, but as the plant and its apex grow, harvesting the meristem becomes more accurate, causing the fraction of meristematic tissue in the sample to increase. However, in truth, the meristem is a highly

heterogeneous tissue, and it does not really make sense to talk about a ratio of “meristematic tissue” in the sample, as not all meristematic genes are expressed in all cells of the “meristem”.

c. A more generic version of hypothesis 2b is that the composition of the sample – in terms of number of cells of each cell type – changes over time. This is highly likely for floral identity genes, such as *LFY* and *AP1*, as they are expressed in floral primordia, which are non-existent at the beginning of the time series, appear during the floral transition and become more and more numerous subsequently. For inflorescence identity genes, this is more debatable. The expression domains of *STM* and *AGL24* seem to spread down the shoot after the floral transition (Geier et al., 2008; Michaels et al., 2003), which could contribute to the kinetics observed for those genes. The measurements of *CLV3* – a gene expressed at the very tip of the SAM – do not show the same increase as the floral transition-related genes. Their values actually decrease over time (Figure 3.16) Assuming that the pool of CLV3-expressing cells remains the same size all the time due to homeostasis and that the expression level of these cells remains constant, this would support the hypothesis that the composition of the sample varies over time. However, in the case of *CLV3*, it would be a dilution rather than an enrichment. The main limit of this hypothesis is that there is no

obvious reason why the large majority of the correlated genes should be upregulated proportionally to each other when their expression domains have different shapes. A possible explanation would be that the SAM has some built-in regularity causing the sizes of all expression domains to respect universal ratios. This is somewhat plausible, as the SAM has a repetitive pattern stemming from the continuous generation of identical lateral organs.

These hypotheses are not mutually exclusive, but there is evidence that hypothesis 2c needs to be investigated further.

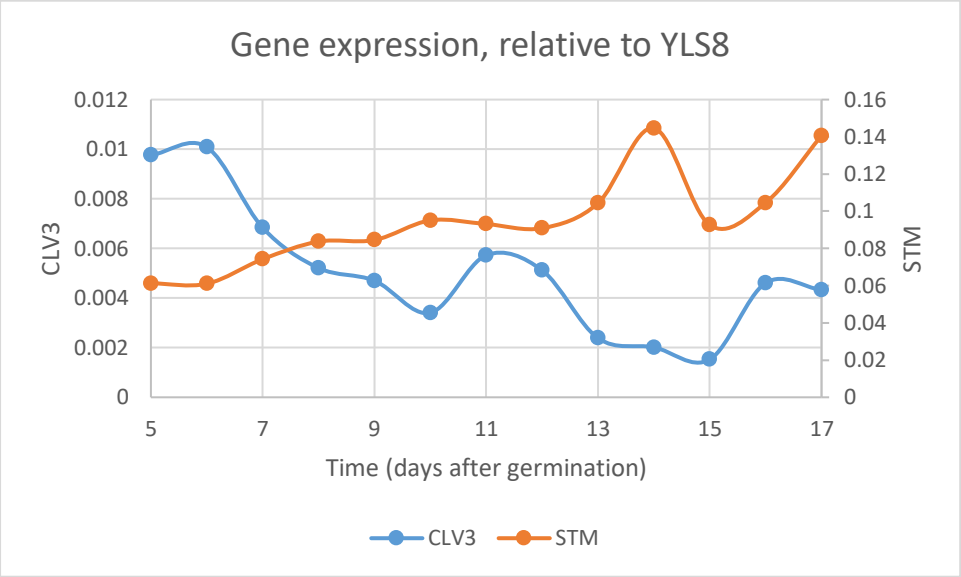


Figure 3.16. Evolution of CLV3 and STM relative concentrations during the growth of the WT plants. The measurements of CLV3 seem to decrease over time, however it could simply be that CLV3 gets more and more diluted in the samples, as the samples (and therefore the reference gene's expression domain) grow in absolute size but the expression domain of CLV3 does not. This would create a distortion between what the data show (an apparent decrease in CLV3 expression) and the actual variable of interest (the intensity of CLV3 expression, presumed to be constant).

3.4. DISCUSSION

The modelling work presented in this chapter has shown that, like in the previous chapter, genetic expression time series can be modelled with very simple models. It also raises important questions about the purposes of models.

3.4.1. Modelling scale matters

As shown with the ODE model of vernalization, modelling a process in a satisfactory manner at a given scale does not require a model to be accurate at smaller scales.

In this particular case, the behaviour of individual cells was ignored and only their aggregate behaviour at the scale of the whole tissue was considered. It was possible to ignore the spatial organization of the tissue because the location of *FLC* transcription does not change its effect, as the FLC protein diffuses throughout the tissue, spatially averaging its distribution. There is also no coordination of *FLC* expression between cells through intercellular exchanges, which might have resulted in hard to predict effects.

In other cases, however, the spatial distribution of genetic expression can have crucial roles.

3.4.2. The spatial organization of the meristem is important to describe how it works

The expressions of some genes are mutually exclusive in a cell. This is the case of *AP1* and *SOC1*, or *AP1* and *FD*. This is because AP1 represses these two activators in negative feedback loops (Kaufmann et al., 2010).

If gene expression is analysed at the tissue level (e.g. by qRT-PCR), *AP1* will appear to be co-expressed with *FD* and *SOC1*, because those three genes are expressed in the SAM after the floral transition. However, *FD* and *SOC1* are expressed in the inflorescence meristem, and *AP1* is expressed in the floral

meristems. This leads to a problem when trying to fit a single-compartment model (i.e. a model that assumes the meristem is homogeneous) to these data, as the single-compartment formalism makes it impossible to reconcile the expression levels measured in the whole meristem with the topology of the regulatory network (e.g. switches and negative feedback loops) and complementary observations, such as *in* situ hybridization (ISH) studies, which are more precise spatially but less precise quantitatively.

The following chapter of this thesis explores in more detail the benefit of modelling heterogeneous tissues like the SAM at a higher spatial resolution, taking advantage of sources of data like ISH experiments.

3.4.3. Ignoring spatial organization in development studies can negatively affect the experimental design itself

One of the premises of the modelling work initially planned for this chapter was that the expression levels of all the genes included in the model are uniform within the meristem. This is however not the case, as mentioned above and detailed in the next chapter. The experiments described in this chapter and in the works of other authors (Jaeger et al., 2013; Valentim et al., 2015) were built on that premise, since they model the SAM as a single compartment. This consequently compromises their accuracy and interpretability.

The incorrectness of this premise has consequences beyond the formalism used for modelling. It also impacts data acquisition methods. In the study

carried out by Valentim and colleagues, the model was fitted to qRT-PCR data. qRT-PCR requires raw measurements to be normalized to control for variations in the quantities of cDNA in the sample. This is usually done by comparing the expression levels of the genes of interest to those of housekeeping genes, and works fine when the number of cells expressing the gene of interest is proportional – or equal – to the number of cells in the sample, across all samples. It is however not the case in the SAM during the floral transition, as the composition of the sampled tissue varies over time, as a consequence of the meristem fulfilling its function: generating new lateral organs.

Another related issue is that the sampled tissue is larger than just the SAM and therefore includes non-meristematic tissue, such as the petioles of the latest leaves and the stem of the main shoot, whose ratio is unknown, non-negligible, and most likely varying over time.

qRT-PCR time series are not the only source of data that this work and that of Valentim and colleagues were based on. Flowering time data from various mutants were also used. However, measuring only flowering time presents some issues.

3.4.4. Flowering time is only one dimension of the floral transition

In floral transition studies, genes are often characterized as promoters or inhibitors of the floral transition, depending on whether the associated mutants are late or early flowering. However, flowering genes have more functions than simply accelerating or slowing down the timing of flowering, as

evidenced by the morphological alterations often accompanying mutations of those genes (Irish and Sussex, 1990; Schultz and Haughn, 1991). Fitting models of the floral transition primarily on flowering time data therefore seems too specific.

Following on that reasoning, one might question the use of *AP1* as a marker of the floral transition, as done by Valentim and colleagues. The onset of *AP1* expression empirically marks the completion of the floral transition in WT plants, however there is no guarantee that this will apply to mutants. The most striking argument, perhaps, is that *ap1* mutants still produce flower-like structures (although they are devoid of petals, hence the full name of the gene, *APETALA1*). This clearly shows that *AP1* is not actually required to produce flowers. It suggests that, conversely, mutants could have their *AP1* expression levels and timing affected, without the timing of their floral transition being affected.

3.4.5. Recommended experimental design for future experiments

Considering the observations made above, a set of recommendations for future experiments was proposed.

First, with respect to the gene expression measurement method, it appears crucial to address the ambiguity regarding whether variations in the measured values stem from changes in expression intensity or expression domains. The best method to ensure this would probably be to use single-cell measurement methods (e.g. single-cell transcriptomics or quantification of fluorescent

proteins using confocal microscopy; microscopy might have the added benefit of being non-destructive and therefore allow longitudinal studies of the same plants). Methods working on subpopulations of cells isolated by micro-dissection or isolation of nuclei tagged in specific cell types (INTACT) are also possible alternatives, although micro-dissection coupled to RNA sequencing can lead to highly variable measurements (Torti et al., 2012). However, if practical constraints preclude the use of any other method than qRT-PCR on crudely dissected samples, it would be very important to identify reference genes suitable to normalize measurements while preserving information about the intensity of gene expression in the cells of interest. Different reference genes could be used for different genes of interest, depending on their expression domains.

The other points address vernalization-specific concerns. First, the cold treatments used in this study did not seem to saturate the vernalization response of the *FRI*$^{+}$ Col-0 plants. According to other sources (Heo and Sung, 2011), it might require between 30 days of cold treatment, which could not be done for this study, due to growth chamber space constraints. Heo and Sung's work also indicates that the reduction in *FLC* expression levels is highly non-linear. Cold treatments of 10, 20 and 30 days result result in ~5%, ~50%, ~85% reductions in *FLC* expression, respectively, compared to the pre-treatment value. This means the 3-week cold-treated plants probably still had about half their normal *FLC* expression, causing them to flower later than expected. For future experiments, it is worth noting that even a saturating cold treatment

does not fully suppress *FLC* expression. There is a residual ~15% *FLC* expression, which may make fully-vernalized *FRI*$^{+}$ plants still flower later than non-vernalized Col-0 plants.

An indirect consequence of the unsaturated vernalization response is that the gene expression time series acquired after the cold treatment stop before the floral transition actually happens. This should be addressed, either by increasing the duration of the cold treatment, or extending the gene expression measurement window. Doing both would probably be optimal, as the former would sharpen the cold-treatment response and the latter would enable the capture of the deceleration of the expression of floral transition genes.

3.5. CONCLUSIONS

An ODE model of the silencing of *FLC* during the vernalization process has been developed, however it was not integrated into Valentim and colleagues' wider model of the floral transition. This is because fundamental flaws in the formalism adopted by the pre-existing model were identified. Most notably, its single-compartment formalism is not able to model the behaviour of the SAM in a biologically relevant way, as it is bound to simulate the joint effects of genes that would normally be expressed in spatially disjoint domains. This in turn brought to light a flaw in the experiment Valentim and colleagues' model is built on: gene expressions were measured and normalized under the assumption that gene expression was uniform in the sampled biological

material, however that material is actually comprised of several kinds of tissues with different expression profiles and whose ratios are expected to vary during development. As a consequence, it is hard to determine from the measurements how intensely the genes of interest are expressed, and therefore, how they regulate each other.

The importance of the spatial organization of genetic expression in the SAM was probably the main revelation from this study, which is why the next chapter is dedicated to its detailed analysis and its exploitation to elucidate the structure of regulatory networks.

4. THE LOGIC OF THE FLORAL TRANSITION: REVERSE-ENGINEERING THE SWITCH CONTROLLING THE IDENTITY OF LATERAL ORGANS

4.1. ABSTRACT

Much laboratory work has been carried out to determine the gene regulatory network (GRN) that results in plant cells becoming flowers instead of leaves. However, this also involves the spatial distribution of different cell types, and poses the question of whether alternative networks could produce the same set of observed results. This issue has been addressed through a survey of the published intercellular distribution of expressed regulatory genes and techniques both developed and applied to Boolean network models. This has uncovered a large number of models which are compatible with the currently available data. It shows that an exhaustive exploration would be unfeasible due to the massive number of alternative models, so genetic programming algorithms have also been employed. This approach allows exploration on the basis of both data fitting criteria and parsimony of the regulatory processes, ruling out biologically unrealistic mechanisms. One of the conclusions is that, despite the multiplicity of acceptable models, an overall structure dominates, with differences mostly in alternative fine-grained regulatory interactions. The overall structure confirms the known interactions, including some that were not present in the training set, showing that current data are sufficient to determine the overall structure of the GRN. The model stresses the importance of relative spatial location, through explicit references to this aspect. This approach also provides a quantitative indication of how likely

some regulatory interactions might be, and can be applied to the study of other developmental transitions.

4.2. INTRODUCTION

Computational approaches have become routinely used in the study of gene regulatory networks. One of the fundamental key outcomes of gene-network activity is specification of the differentiated cell types during development that lead to different tissues and organs. To address this particular question, computational models have to capture the unfolding, both in time and space, of the program embodied by interactions between genes, transcription factors and other molecular complexes. This necessity to describe spatio-temporal patterns of gene activity entails an important computational cost. In addition, the data available to build and assess computational models are typically incomplete or ambiguous, since precise spatio-temporal patterns of gene expression are seldom available for multiple genes in a single data set. This paper proposes tools designed to represent the specification of new cell identities during development, and to fit models against incomplete data. This work focuses on the floral transition, see below, but the methods aim to be applicable to other systems involving cell differentiation and the underlying spatial patterning of biological tissues.

Flowers are the reproductive organs of plants. Therefore, their formation is crucial for reproductive success. From a developmental perspective, flower formation starts with the triggering of specific pathways in the founder cells of

lateral organs (i.e. leaves initially), so that they develop into flowers instead. This developmental switch is called the floral transition. It is one of many aspects of cell-fate specification in the shoot apical meristem (SAM), which comprises multiple tissues, each with their own gene-expression profile but all produced from a single stem-cell population. This early specification of cell types, through the interactions between genes and hormones, enables newly formed tissues to later develop into all the aerial parts of a plant (Adrian et al., 2009; Simon et al., 1996). The transition goes through three well-characterized stages, starting with a vegetative meristem, which produces leaves. Upon the trigger by the appearance of the protein FT, this meristem becomes an inflorescence meristem, from which floral meristems appear that produce flowers.

While the pathways involved in the floral transitions have been reviewed (Fornara et al., 2010; Liu et al., 2009) and modelled using Ordinary Differential Equations (ODEs) (Dong, 2003; Jaeger et al., 2013; Valentim et al., 2015) and neural network (Welch et al., 2003) formalisms, these studies give little if any attention to the spatial organization of the SAM and do not include any representation of space. The side effects of this simplification obviously include the inability to explain how the spatial organization of the SAM is acquired, but also the prediction of unrepresentative gene-expression profiles, because the gene expression measurements have come from multiple cell types. This potentially leads to the consideration of combinations of regulatory

interactions that cannot actually occur *in vivo*, because the genes involved are not, in reality, expressed in the same cells.

The present study focuses on how the gene-regulatory network of the SAM is able to determine the transition of its daughter cells into stem, leaf, flower or other cell types, based on environmental and positional cues. To address the lack of spatial information found in previously published studies, a novel approach was required. We therefore propose a modelling framework which includes an explicit representation of space. Regulations known from the literature may be ambiguous, so the proposed methodology comprises a method for the inference of models, based on experimental data. This entailed generating a compendium of published *in situ* hybridization (ISH) experiments, to describe groups of jointly expressed genes. Models deemed plausible had to reproduce both the observed patterns of co-expression and the known developmental transitions. This offers the potential to explore alternatives to current thinking about the regulatory mechanisms and predict novel regulatory interactions for laboratory testing.

If ODE modelling is used, the number of possible alternative regulatory interactions, even among a small number of genes, would lead to unfeasibly long parameter-estimation times. However, a formalism particularly well suited to this task is Boolean modelling, which naturally handles binary (on or off) variables that accord with the resolution of the ISH data. For a brief introduction to Boolean models, please refer to S1 Text. Even though Boolean models are lightweight, the space of possible models for a given set of genes

remains computationally expensive to explore. In simple cases, this “model” space can be explored through exhaustive searches, but it quickly becomes intractable as the number of possible regulatory interactions increases. In more complex cases, heuristic techniques are required. In this work, a genetic programming algorithm has been employed to find suitable models that explain all observed data.

Boolean network models have been used successfully to study developmental processes, such as floral development (Espinosa-Soto et al., 2004), which directly follows the floral transition. By representing genes as binary variables influencing each other, they enable us to run simulations and find steady states of the system. These steady states can then be interpreted as cell identities or expression profiles. The idea of matching biological observations to steady states in not new: the logical rules built by Espinosa-Soto and colleagues resulted in steady-states matching biological observations. This work describes a related process: building up the logical rules from the biological observations. It is similar to what has been done by La Rota *et al.* for the regulatory network controlling sepal formation (La Rota et al., 2011).

Genetic algorithms have previously been used in conjunction with Boolean modelling (Kang et al., 2011; Roli et al., 2011). These methods operate on Boolean models at the level of truth tables, whereas genetic programming operates at the level of equations. While truth tables can always be generated from equations and equations can be factorized from truth tables, working on equations has several benefits: factorizing equations is more expensive than

deriving truth tables, equations are human-readable, and constraints of complexity can be enforced on them.

This work has shown, for the floral transition, that an exhaustive search of all possible regulatory interactions is prohibitive. Restricting the search to models supported by the published regulatory networks explains the steady states but, when attempting to explain the dynamic transitions between them, results in many ambiguous regulatory events. Using genetic programming to find models that correspond to the ISH data and known cell type transitions reduced the ambiguity almost entirely, identified other regulatory interactions that have been independently confirmed in other published work.

4.3. RESULTS

The most common representation of the core regulatory-network (Fornara et al., 2010) is shown in Figure 4.1, though other regulatory components have also been reviewed by Liu et al. (Liu et al., 2009). As a necessary first verification, one needs to assess whether this topology is sufficient to generate the observed patterns of gene expression, or if new regulators or interactions are required. As detailed below, a given topology, or regulatory graph, can be achieved by a large number of distinct models and one needs to determine whether at least one of them is able to generate the required expression patterns. In some cases, all the potential models can be listed exhaustively, but it will soon become clear that in the general case the space to explore is too large to allow for an exhaustive search.

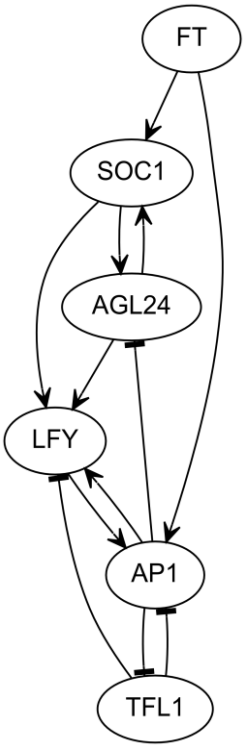


Figure 4.1. Common representation of the core regulatory network of the floral transition. Nodes represent genes and edges represent regulatory interactions. V-shaped and T-shaped arrow heads respectively denote activation and repression by the regulatory nodes.

4.3.1. The cost of running an exhaustive search on the whole space of possible models is prohibitive

Classically, three meristematic identities are distinguished: vegetative, inflorescence and floral (Adrian et al., 2009; Simon et al., 1996), and are normally defined by five main genes. *SOC1* and *AGL24* are markers of the inflorescence identity, and *LFY* and *AP1* of the floral identity (Mandel et al., 1992; Mandel and Yanofsky, 1995; Weigel et al., 1992; Weigel and Nilsson, 1995), while *TFL1* inhibits the floral identity (Gustafson-Brown et al., 1994;

Weigel et al., 1992) and is a marker of vegetative identity (the inflorescence also expresses *TFL1* though, which can be attributed to the inflorescence conserving some vegetative traits). A sixth gene, *FT*, encodes a mobile protein that is synthesized in leaves, moves to the SAM through the phloem (Jaeger and Wigge, 2007) and triggers the transition from the vegetative to the inflorescence and floral identities. However, owing to a memory effect, *FT* is not needed to maintain the inflorescence and floral identities after the floral transition (Adrian et al., 2009). Using this information, characteristic expression profiles can be established for each meristematic identity (Table 4.1 and Figure 4.2). The question arises of whether or not there are any other regulatory combinations of these genes than those reviewed in the literature that result in the same set of identities.

Table 4.1. The three classical meristematic identities.

Vegetative	*TFL1* (Adrian et al., 2009)
Inflorescence	*(FT)* *SOC1* (Adrian et al., 2009) *AGL24* (Adrian et al., 2009) *TFL1* (Adrian et al., 2009; Pidkowich et al., 1999)
Floral	*(FT)* *AP1* (Adrian et al., 2009; Pidkowich et al., 1999) *LFY* (Pidkowich et al., 1999)

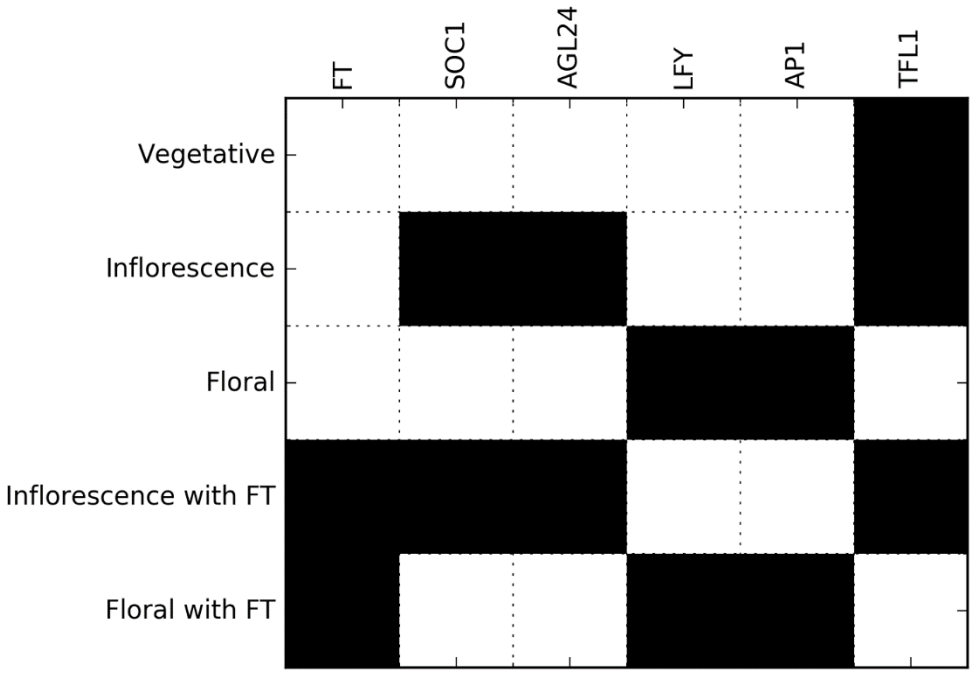


Figure 4.2. Expression profiles of the three classical meristematic identities. Each row corresponds to a desired steady state, and each column to a gene. Black and white cells indicate whether a gene is expressed or not, respectively.

The number of models to examine is a function of the numbers of input nodes (nodes with no inbound regulation) and internal nodes (nodes with inbound regulations). As discussed in more detail in S1 Text, a Boolean model is nothing

other than a map acting on the set of all possible states (combinations of "on"/"off" status of each node) of the system; each state is sent to a "successor" state by this map, hence describing dynamical evolution (steady states being their own successor). There are 6 nodes in total, so there are $2^6 = 64$ possible Boolean states. To define a model, a successor must be defined for each of these 64 states. The behavior of input nodes is fixed, so successors are uniquely characterized by the behaviors of the five internal nodes. Looking naively at the full set of all Boolean models, there are therefore $2^5 = 32$ possible choices of successor for each of the 64 states, i.e. $32^{64} = 2^{320} \simeq 2.10^{96}$ potential models. This is more than the estimated number of atoms in the observable universe, which is ~10^{80}. Even with a computer able to check 10 billion models per second, it would still take ~6.10^{78} years. This quick estimate shows that a brute force approach is impractical and that one needs to constrain the search space using prior biological knowledge.

4.3.2. The topology summarized by Fornara et al. can explain the steady states but not the dynamic behavior

The first, obvious, constraint on the search space is to exclude models containing regulatory interactions that are not backed by any biological evidence. As an added benefit, should solutions be found, this would demonstrate that the set of evidence-backed interactions is comprehensive enough to explain the behavior of the system. In an attempt to find a reasonably sized set of regulatory interactions that can explain the behavior of

the system, the Fornara et al. network (Fornara et al., 2010) has been used as the main source of prior knowledge, without any additions from Liu et al. (Liu et al., 2009) which would require additional genes. This set can be determined very cheaply, as it is comprised of all the models whose truth tables follow a pattern depending solely on the required steady states and the topology of the network (see S2 Text).

The outcome of this search was a set of 262,144 models compatible with Fornara et al.'s topology and exhibiting the required steady states. This topology is therefore sufficient to explain the steady states of the system. However, it cannot reproduce state transitions undergone by the real biological system during development, and, most crucially, does not include the activation of *SOC1* by *FT* (see S1 Fig).

In our modelling framework, we describe transitions as the given of an initial steady state I, a perturbation P to be applied to that steady state, and a final steady state F, resulting from the spontaneous evolution of the system following the perturbation. Both I and F correspond to one of the cell identities described in a matrix such as Figure 4.2, built using biological knowledge about gene expression domains, and P to the toggling of one or a few variables representing the appearance or disappearance of non-cell-autonomous factors. P is derived from knowledge about the motion of cells, relative to the domains of these factors, during development. The factors toggled by P are therefore effectively the triggers of the transitions from the modelling perspective. The associated biological interpretation is that non cell-

autonomous species form spatial patterns in the SAM that are constantly perturbed by growth and cell divisions. This causes cells to enter some patterns and exit others, as those patterns reorganize. The topology by Fornara et al. lacks a trigger with a pattern matching the position of floral primordia. Thus, for these reasons, the topology by Fornara et al. cannot explain the dynamic behavior of the SAM.

4.3.3. The addition of two interactions yields models that are able to mimic the changes in cell identities

The failure of this exhaustive search to explain dynamical behavior requires the model to be enlarged with two interactions from Liu et al.: *AP1* → *SOC1* and Auxin → *LFY*. These choices were guided by parsimony, the intuitive fact that they are likely to counteract the irresponsiveness of *SOC1* to *FT*, and the absence of difference between the unsteady states leading to the inflorescence and floral identities observed in our first exploration. However, this will increase even further the number of possible models.

Constraining the search space of Boolean models with a defined network topology greatly reduces the number of models to explore. The exact figures depend on the topology. The Boolean network model formalism dictates that the state of any internal node is only dependent on the states of its regulators. Therefore, if node i has r_i regulators, its truth table will have 2^{r_i}entries. As a consequence, there are $2^{2^{r_i}}$ ways of choosing the truth table of node i. Building the whole model is equivalent to picking a combination of truth tables

for all nodes, so the number of models in the search space is given in Equation 4.1.

$$\prod_{i=1}^{n} 2^{2^{r_i}} = 2^{\sum_{i=1}^{n} 2^{r_i}} \quad (4.1)$$

With the topology from Fornara et al. plus the two extra interactions, $\sum_{i=1}^{n} 2^{r_i}$ equals 54.

As a consequence, there are 2^{54} models in the search space after excluding models that do not conform to prior knowledge (down from 2^{320}). Details of the calculation are provided in Table 4.2. Furthermore, most of them can be ruled out because they are not compatible with the observed steady states (see S2 Text). In this case, only 2^{37} solutions presented the required steady states (Figure 4.2). As evidenced by the formulae, adding new interactions becomes more and more expensive. In particular, the latest two interactions added into the data set, *AP1*→*SOC1* and Auxin→*LFY*, increased the size of the search space 2^4-fold and 2^{16}-fold, respectively. This brought the problem close to the limit of what was computationally feasible. Performing the exhaustive search on this problem takes about 1.5 years with current CPUs, but was achieved using a 192-core High-Performance-Computing cluster running for 3 days. The search returned 1.6 billion suitable models. These solutions were used to build an aggregate topology graph of the GRN (Figure 4.3), using the methods described in S6 Text.

Table 4.2. Contributions of each gene to the number of models to explore.

i	r_i	2^{r_i}
SOC1	3	8
AGL24	2	4
LFY	5	32
AP1	3	8
TFL1	1	2

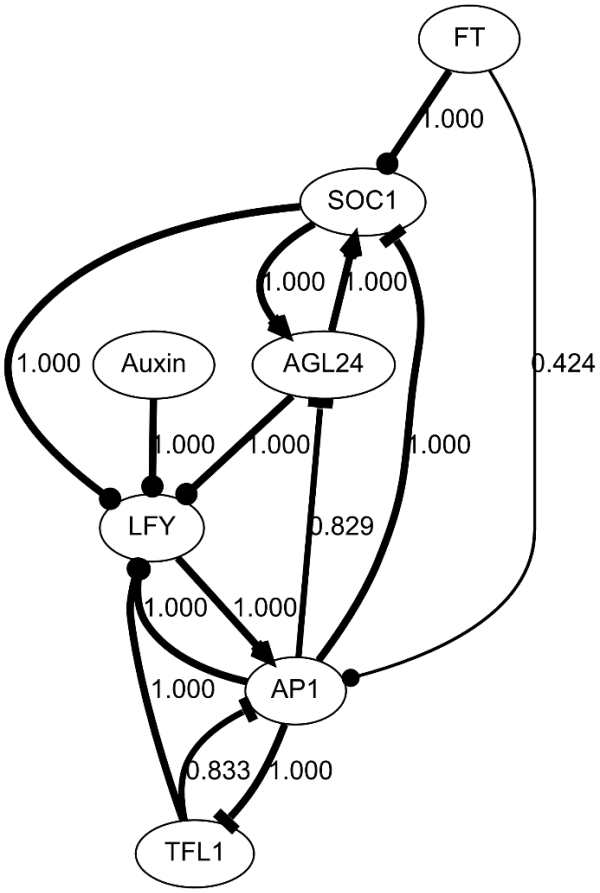


Figure 4.3. Aggregate graph of the models generated by exhaustive search on Fornara data set with 2 extra interactions. The nodes of the graph represent the species of the regulatory network, which are also nodes of the Boolean network models. Edges represent regulatory interactions between regulators and their targets. Arrowheads are placed on the side of the target species. V-, T- and O-shaped arrowheads respectively denote up-regulation, down-regulation, and interactions that can fall in either category, depending on the context and the model. Edge thicknesses and edge labels indicate the frequency of occurrence of the associated interactions, across all the models generated. Owing to the very large number of models obtained, a frequency displayed as 1.000 does not necessarily mean all models.

The 1.6 billion models represent networks with mostly similar topologies. Among the 14 interactions allowed in the search space, all appear in at least some models, and 11 appear in all models. 7 interactions can clearly be labelled as positive or negative, but the other 7 remain ambiguous. This happens because either an interaction is sometimes positive and negative in the same model, depending on which other regulators are present, or it is positive in some models and negative in others.

Figure 4.4 shows the proportions of models in which each interaction is positive, negative, ambivalent, and absent. In most models, the interactions controlling *LFY* are ambivalent, meaning that the regulators of *LFY* can be both activators and repressors, depending on the combination of other regulators. Such behaviors do not seem very plausible. Instead, it is likely that these models are simply artefacts resulting from the high number of regulators of *LFY* and the comparatively small amount of information about the behavior of *LFY*: many combinations of *LFY* regulators are possible, but the actual behavior of *LFY* is unknown in most of them.

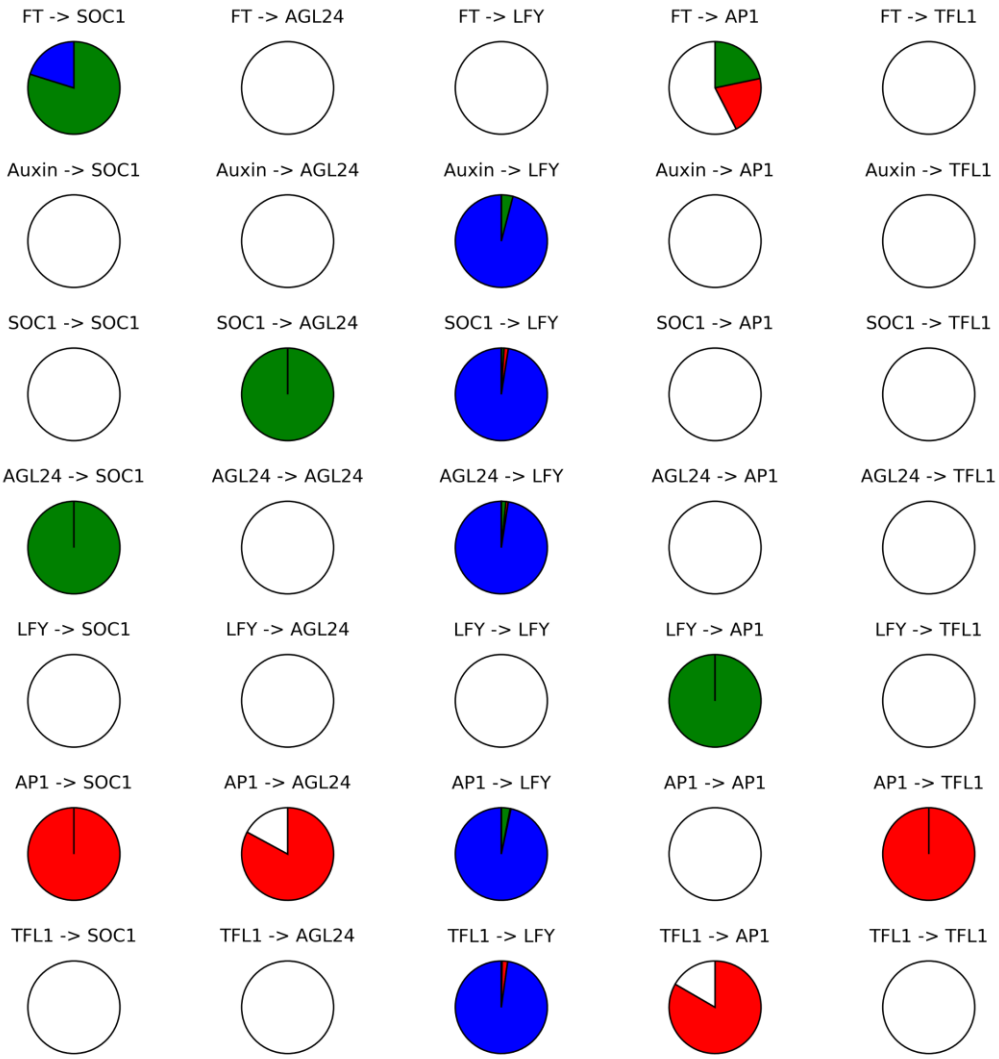


Figure 4.4. Distribution of interaction types per interaction across the set of models generated by exhaustive search. Each pie chart indicates the proportions of models in which the associated interaction is positive (green), negative (red), ambivalent (blue) or non-existent (white).

4.3.4. A higher resolution description of gene expression during the floral transition can be established from *in situ* hybridization (ISH) data

A survey of published ISH studies has been carried out for genes *AGL24*, *AP1*, *LFY*, *SOC1*, *TFL1* and *FD*, which interacts with FT (see S4.1 Table). The

expression domains of each of these genes were analyzed at various developmental stages to establish co-expression maps. For most genes, proteins were assumed to be distributed following the same pattern as their respective mRNAs, as nothing indicated otherwise. However, in the case of the *TFL1*, there was clear evidence that the TFL1 protein was mobile and had a distribution pattern different from that of its mRNA. As well as the three classical meristematic identities (Table 4.1), this survey has revealed additional identities, and most of them can be matched to zones already characterized in studies of SAM development (Clark, 1997), see Figure 4.5 and Figure 4.6.

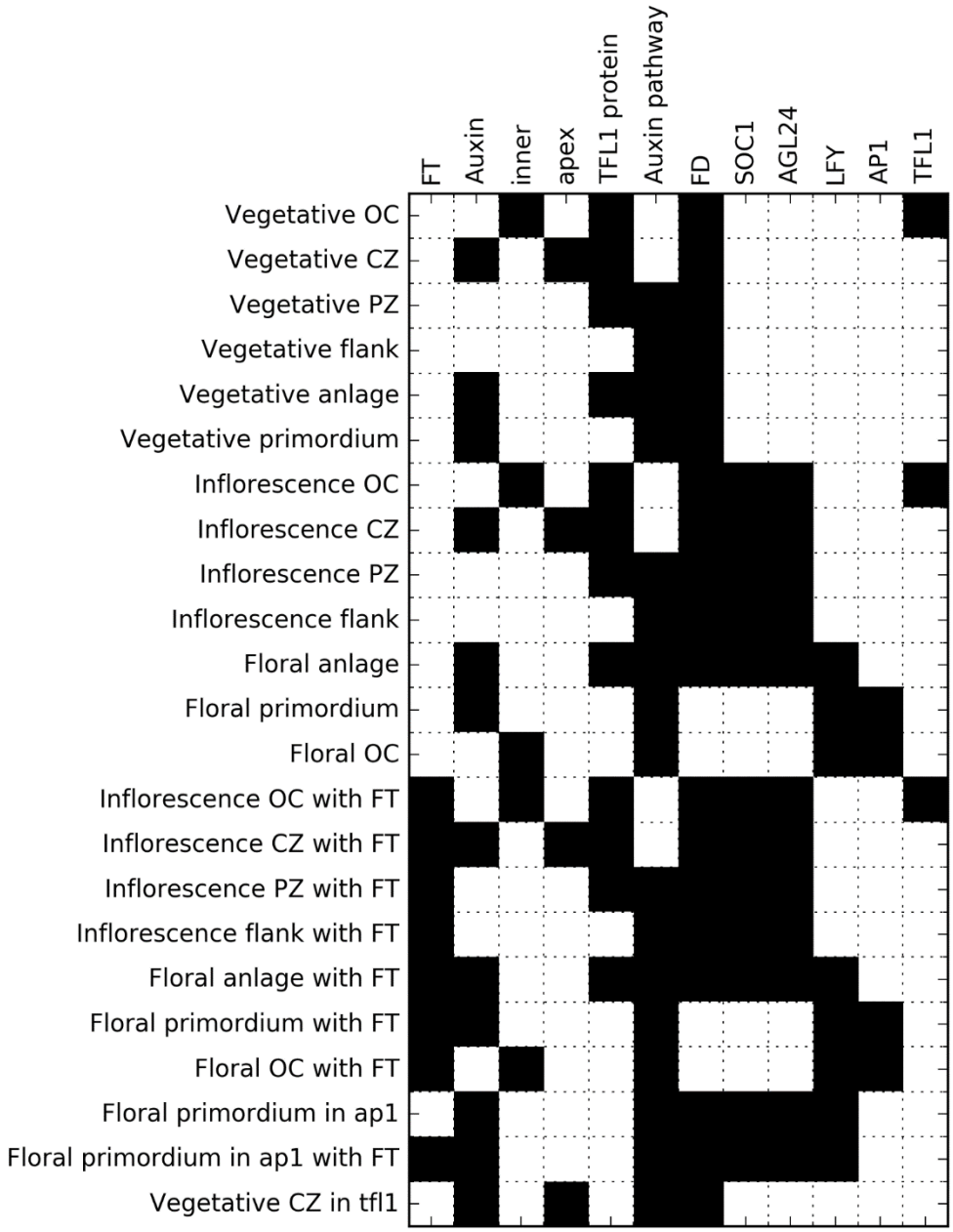


Figure 4.5. Matrix of gene expressions in cell populations identified from ISH pictures. Rows correspond to cell populations and columns to chemical species or other variables. A black square means a species is present or a variable is on in the associated tissue.

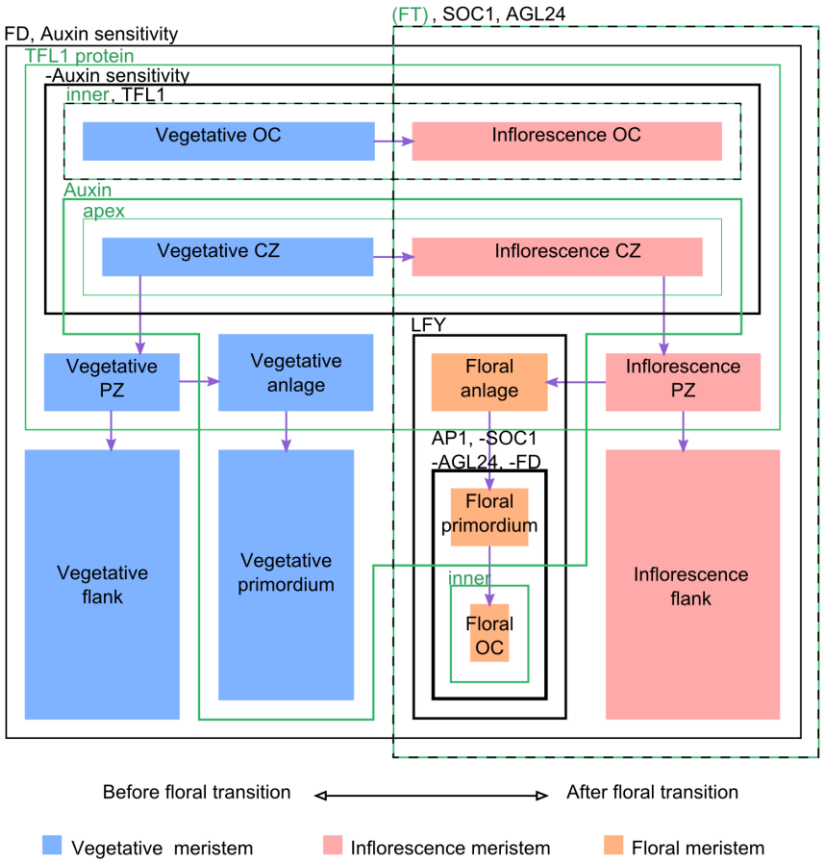


Figure 4.6. Diagram of gene expression domains in time and space. Green and black contours mark the expression domains of the species mentioned in the upper left corners of the boxes. A "-" sign before a gene name means the frame marks a hole in the expression domain of that gene. The green species are those used as triggers of the transitions between developmental stages. Transitions (symbolized by purple arrows) are triggered by toggling the variables associated with the green species (i.e. crossing green lines on the diagram), which pushes the system towards a new identity, often causing black species to also toggle their values (i.e. cross black lines on the diagram). Identities are represented as colored areas for clarity, the surface of these areas is not representative. The left-hand and the right-hand halves of the picture are temporally separate, all other separations are spatial.

These zones are described below, and unless stated otherwise, exist in both the vegetative and floral phases, although the genes they express change.

The first zone is the organizing center (OC). It is classically defined as the expression domain of *WUS*, but it also seems to express *TFL1* (Conti and Bradley, 2007; Liu et al., 2013), which encodes a mobile protein that is transported towards the apex. The second zone is the central zone (CZ), which contains stem cells and is located at the very apex of the meristem. These cells are unable to initiate the formation of a primordium in response to auxin (Reinhardt et al., 2000), possibly because their auxin sensitivity has been disrupted, as suggested by the expression patterns of some genes of the *ARF* family (Vernoux et al., 2011). The third is the peripheral zone (PZ), vegetative or inflorescence, which surrounds the CZ. We define its border as that of the diffusion domain of the TFL1 protein (Conti and Bradley, 2007). Within the PZ, some cells actually belong to another (fourth) identity: anlagen or founder cells of lateral organs. Their defining characteristic is a high concentration of auxin. Floral anlagen start expressing *LFY* (Blazquez et al., 1997). The fifth identity is the primordia for anlagen that have gone through the boundary of the TFL1 protein domain, which express *AP1* (Wang et al., 2009; Wigge et al., 2005), but not *FD* (Wigge et al., 2005), *SOC1* (Wang et al., 2009) or *AGL24* (Michaels et al., 2003). Finally, the sixth is the meristem flank, which surrounds primordia. Compared to the peripheral zone, its differences are that it does not have TFL1 proteins (Conti and Bradley, 2007) and it is insensitive to auxin treatment (Reinhardt et al., 2003).

In addition to these known steady states, knowledge of the processes involved in plant development has enabled us to generate a list of initial steady states, perturbations and resulting steady states (Table 4.3). These steady states and transitions were also complemented with information inferred from the phenotypes of the *tfl1* and the *ap1* mutants (see Table 4.4). Studying the *ap1* mutant led us to consider a seventh zone: the floral OC, which does not have any counterpart in the vegetative SAM. In WT plants, it is very similar to the floral primordium, except that it is located deeper within the meristem, and we assume it does not have a high concentration of auxin. In the *ap1* mutant, this territory is expected to turn into an inflorescence OC instead, paving the way for a recursive, cauliflower-like inflorescence architecture.

Table 4.3. Developmental transformations in WT

Initial steady state	Perturbation	Final steady state
Vegetative CZ	- apex - auxin	Vegetative PZ
Vegetative PZ	- TFL1 protein	Vegetative flank
Vegetative PZ	+ auxin	Vegetative anlagen
Vegetative anlagen	- TFL1 protein	Vegetative primordium
Inflorescence CZ	- apex - auxin	Inflorescence PZ
Inflorescence PZ	- TFL1 protein	Inflorescence flank
Inflorescence PZ	+ auxin	Floral anlagen
Floral anlagen	- TFL1 protein	Floral primordium
Floral primordium	+ inner - auxin	Floral OC
Vegetative OC	+ FT	Inflorescence OC
Vegetative CZ	+ FT	Inflorescence CZ
Inflorescence CZ with FT	- apex - auxin	Inflorescence PZ with FT
Inflorescence PZ with FT	- TFL1 protein	Inflorescence flank with FT
Inflorescence PZ with FT	+ auxin	Floral anlage with FT
Floral anlage with FT	- TFL1 protein	Floral primordium with FT
Floral primordium with FT	+ inner - auxin	Floral OC with FT

Table 4.4. Transitions in mutant plants.

Mutation	Initial steady state	Perturbation	Resulting steady state
tfl1	Vegetative CZ	+ *FT*	A state with *AP1*
ap1	Floral anlagen	- TFL1 protein	Floral primordium in *ap1*
ap1	Floral primordium in ap1	+ inner - auxin	Inflorescence OC (similar to WT)
ap1	Floral primordium in ap1	+ apex + TFL1 protein	Inflorescence CZ (similar to WT)

The additional data provided by ISH were unfortunately shown by exhaustive search to be incompatible with the supplemented Fornara topology, as some of the observed steady states (Figure 4.5) provide conflicting information about the regulation of some genes, implying that the topology is incomplete. As a consequence, in order to solve this problem, it is crucial to develop a method that can suggest new regulatory edges for the network. One approach involves the use of genetic programming. There are two motives for developing such an algorithm: the need for simpler over complex/implausible regulatory interactions, and a non-exhaustive strategy of exploration of the search space should be more cost-effective and allow the solving of complex cases that involve more species and interactions. This performance gain can also be used to explore models that do not perfectly match prior knowledge, and hence potentially identify previously unknown interactions.

4.3.5. A genetic programming algorithm proposes Boolean models that explain the meristem development during the floral transition

465 models fitting the observations were generated using the genetic programming algorithm. As these results included models that shared the same truth table, they could be filtered down to 103 distinct models (i.e. models with distinct truth tables). These models can be clearly classified according to their fitness values (Figure 4.7; lower is better). The presence of clearly separated peaks is due to the way the fitness function was constructed. Each peak represents a different number of novel interactions. The number of copies per distinct model from the first peak (fitness < -0.18) is plotted in Figure 4.8. It empirically shows that not all models of approximately equal fitness will be found with similar frequencies by the algorithm.

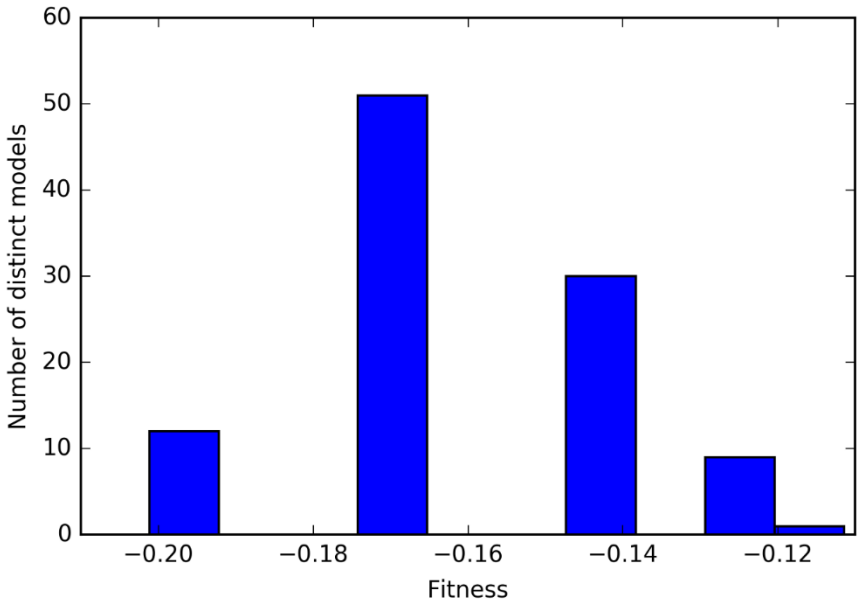


Figure 4.7. Distribution of the fitness values of the 103 distinct models generated by genetic programming. The formula for this can be found in S5 Text. Models found by genetic programming spontaneously segregate into clusters corresponding to their fitness values. Each cluster corresponds to a different number of novel interactions introduced into the regulation network. The algorithm attempts to find models with the fewest novel interactions possible, i.e. those with the lowest fitness values. It does however not always succeed in finding models with the actual lowest possible number of novel interactions, hence the presence of several clusters on the diagram.

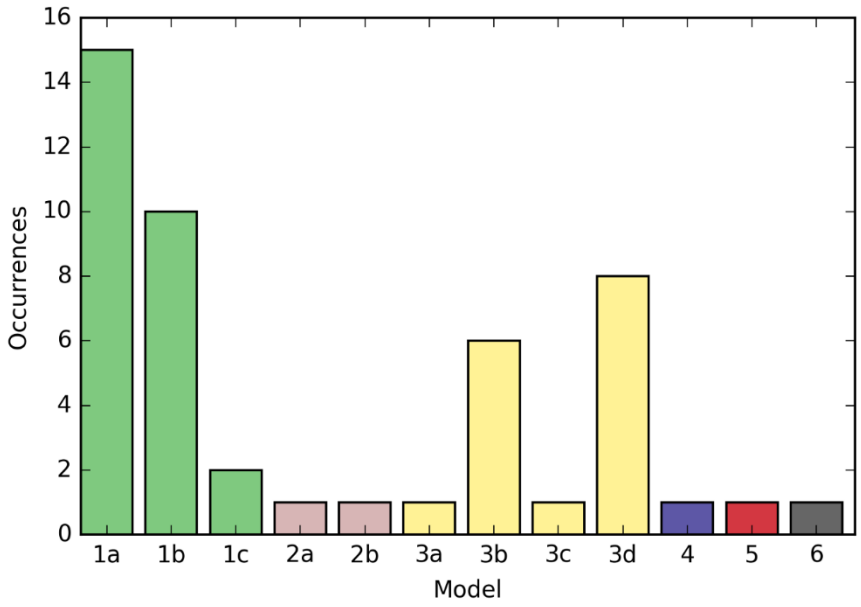


Figure 4.8. Counts of the distinct models generated by the genetic programming algorithm, with fitness < -0.18, in order of increasing fitness (lower is better). Models with the same number have the same fitness value. The algorithm favors models with lower fitness values, but even at a given fitness value (1a-1c, 3a-3d), not all models are found with the same frequency, suggesting that some may be easier to find than others.

As mentioned previously, the topology provided to the algorithm did not allow, as is, for any solutions to be found. As a consequence, all solutions proposed by the algorithm involve additional interactions that were not part of the prior knowledge. An aggregate graph of the topologies of the 103 models is presented in Figure 4.9. It reveals numerous novel interactions, many of which occur at low frequencies (< 10%). This is because the set includes sub-optimal models, as far as the parsimony of new interactions is concerned (i.e. they

include models that have more novel interactions than necessary). This can be addressed by retaining only the models with lower (i.e. better) fitness values.

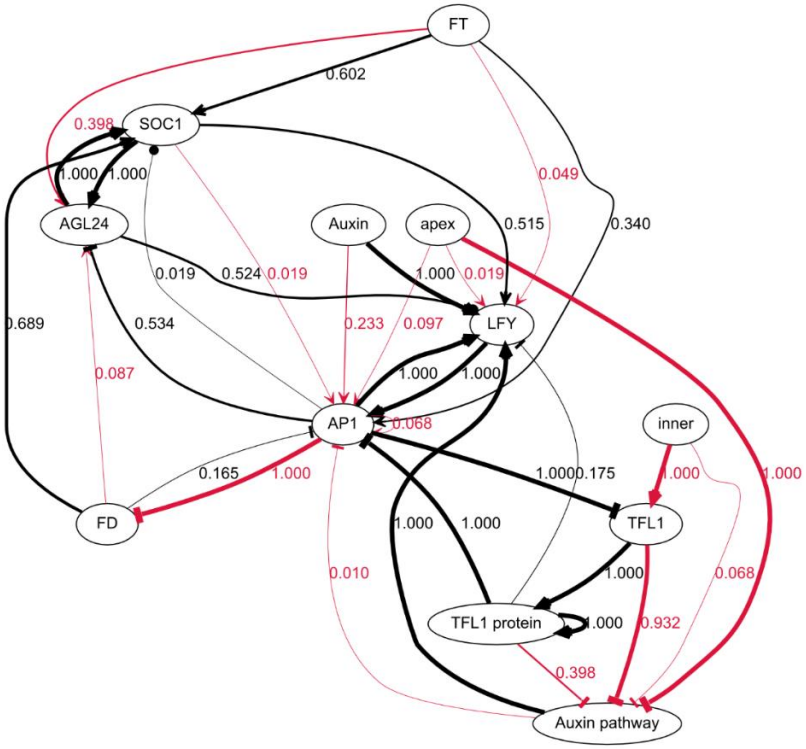


Figure 4.9. Interactions found in the 103 networks generated by genetic programming. Black edges are part of the prior knowledge, red edges are not. All edges were allowed in the search, however red edges incur penalties, and their inclusion is therefore minimized. Edge labels represent the frequencies of their respective edges. Many novel interactions appear in at least some of the 103 models, but most of them with low frequencies. The interactions involving apex and inner however both have frequencies of 1. This confirms that the variables apex and inner, as they were defined, would be able to explain the patterning of the auxin signaling pathway and *TFL1*, respectively, although additional work would be needed to explain how apex and inner can be defined molecularly. This also shows that no way to substitute apex or inner with other variables could be found, unless it would involve substantially more novel interactions.

In the following, only the best models (fitness values <-0.18) were retained, as they are – by construction – the models with the fewest novel interactions (4 in total). Some are more parsimonious than others in terms of known interactions (see section S4.2 Table), but we will consider them equally relevant here, as our main focus is the study of minimal sets of novel interactions able to complement published networks. The aggregate graph of this selection is presented in Figure 4.10. The 12 models selected this way suggest:

- *FD* is repressed by AP1; this would constitute a negative feedback loop, whereby FD activates floral identity genes before indirectly turning itself off;
- *SOC1* is not necessarily repressed directly by AP1; the results of the exhaustive search had shown that a negative feedback loop was necessary, but it might be the same as that of FD;
- *AP1* is not necessarily activated directly by FT; an indirect activation pathway through SOC1 and LFY is sufficient;
- *TFL1* is upregulated by a non-modelled factor present in the inner tissue of the meristem, or a modelled factor with unknown interactions occurring in the inner tissue of the meristem;
- The auxin pathway is disrupted by TFL1 and a non-modelled factor present in the CZ, or a modelled factor with unknown interactions occurring in the CZ of the meristem.

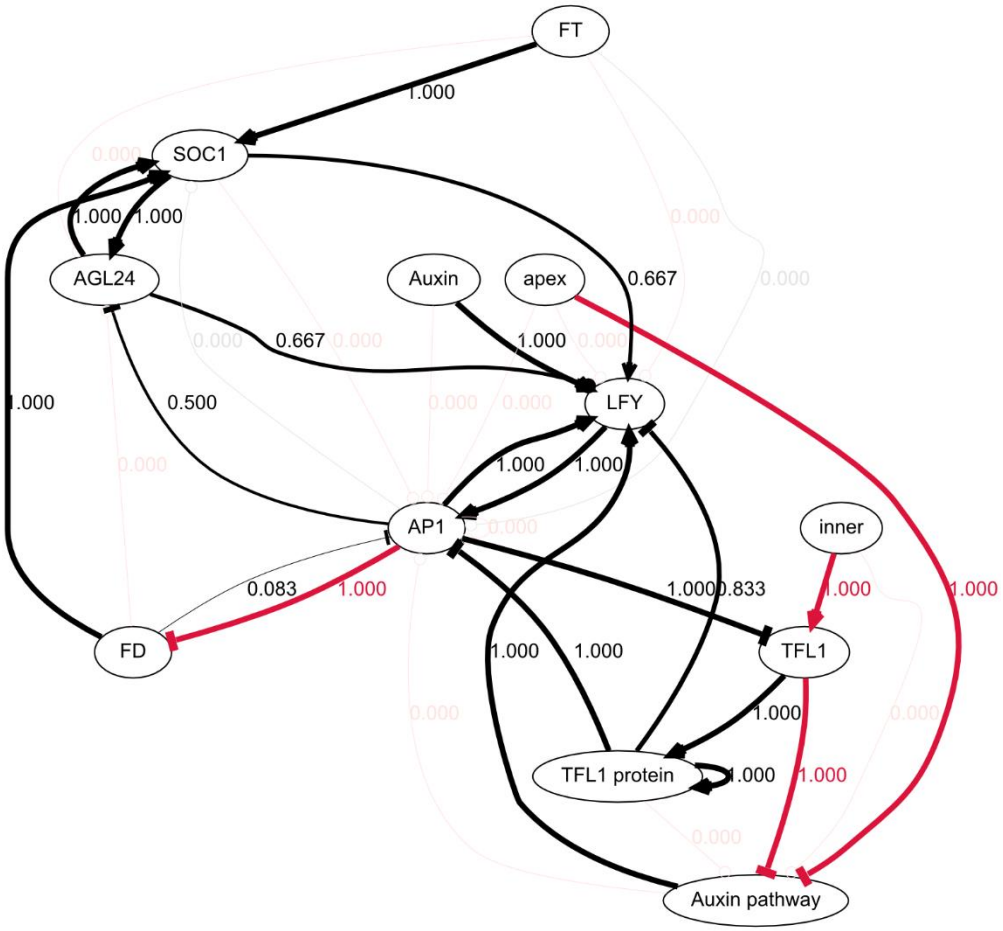


Figure 4.10. Interactions found in the 12 networks in the first peak of fitness. This shows the repressions of *FD* by AP1 and of the auxin pathway by TFL1 are the most straightforward additions required to make the network consistent with the data. This also shows that some interactions are not required to explain the data, namely *FT* → *AP1*, *FD* → *AP1* and *AP1* → *SOC1*.

In this subset of solutions, only one interaction (*AGL24* → *LFY*) is of undefined nature in the aggregate of the 12 models. This interaction is however never undefined within any given model (Figure 4.11), instead there are some models where it is positive, and some where it is negative. This shows that this

method is able to avoid complex models. The equations of the 12 models are given in S4.2 Table.

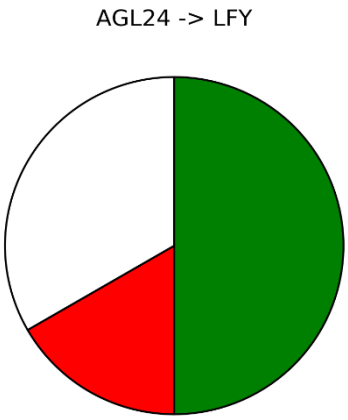


Figure 4.11: Breakdown of the type of the *AGL24* → *LFY* interaction across the subset of 12 models. The pie chart indicates the proportion of models in which the associated interaction is positive (green), negative (red), or non-existent (white).

Among these 12 distinct models, 5 interactions are not present in all models:

- TFL1 protein → *LFY*;
- *FD* → *AP1*;
- *SOC1* → *LFY*;
- *AGL24* → *LFY*;
- *AP1* → *AGL24*.

Principal component analysis (PCA) was carried out to determine the number of degrees of freedom in the set of 12 models (S6 Text). It showed this set was really 5-dimensional, but 91% of variance could be explained by the first three

components (Table 4.5). The first component only covers interactions *SOC1* → *LFY* and *AGL24* → *LFY*, with opposite coefficients, showing the *SOC1* and *AGL24* nodes can play similar roles in the regulation of *LFY* in the generated models. The second component is mostly composed of *AP1* → *AGL24*, probably because it is not necessary for a model to fit the observations: the most concise models generated do not include that interaction at all (see S4.2 Table).

Table 4.5. Principal components of the variability in the subset of 12 models. The three main components explain 91% of the variance. The first component indicates that *SOC1* and *AGL24* can play similar roles in the regulation of *LFY* in the generated models. The second component is strongly influenced by *AP1* → *AGL24*, an interaction that is highly optional in the set of 12 models.

Component	TFL1 protein → *LFY*	*FD* → *AP1*	*SOC1* → *LFY*	*AGL24* → *LFY*	*AP1* → *AGL24*	Percentage of variance explained
#1	-0.000	0.000	0.707	- 0.707	0.000	0.366
#2	-0.357	0.362	0.190	0.190	0.819	0.295
#3	0.622	-0.259	-0.350	-0.350	0.548	0.253

Looking at combinations of interactions model per model provides additional insight. Noticeably, LFY is always upregulated by *SOC1*, *AGL24* or both, in each of the proposed solutions (Table 4.6). It highlights the importance of an activation path from inflorescence genes (*SOC1* and *AGL24*) to the floral identity gene *LFY*, and confirms that one such path is theoretically sufficient. However, if only one of them activates *LFY*, the algorithm is not able to suggest which one from the available data.

Interestingly, none of the configurations reported in Table 4.6 involves all of the five interactions, even though they do all feature in the topology reviewed by Fornara et al. However, the missing interactions can be either of the five. This shows there is not only redundancy between *SOC1* and *AGL24*, but also at a higher level.

Table 4.6. Combinations of interaction types in the best cluster of models generated by genetic programming and their numbers of occurrence in the 12-model set. Empty cells denote the absence of the associated interactions.

TFL1 protein → *LFY*	*FD* → *AP1*	*SOC1* → *LFY*	*AGL24* → *LFY*	*AP1* → *AGL24*	Occurrences
-1		1			2
-1			1	-1	2
-1			1		2
-1		1		-1	2
	-1	1	-1	-1	1
-1		1	1		1
		1	-1		1
-1		1	1	-1	1

4.4. DISCUSSION

Even though Boolean models are simple and cheap to simulate, they are still very flexible. The downside of this flexibility is that, for most model reverse-engineering applications, it is impractical to test all possible models exhaustively to find those that fit observations. This work shows that this can be improved by constraining the search space to models that conform to a given topology, which is not helpful when the network topology is unknown.

The genetic programming method used here is able to handle incomplete topologies. Unlike exhaustive search approaches, it is also able to favor models with simple Boolean equations. These are more likely to represent biological regulatory mechanisms, because a given regulator rarely changes from being an activator to a repressor. However, like any method, the validity of its results depends on the quality of the input data.

A large part of the input data in this work has been extracted from *in situ* hybridization experiments. This shows the locations the mRNA of the studied genes, but not their proteins, which is an issue for mobile proteins, such as FT and TFL1. Although the greatest care was taken when interpreting ISH pictures, comparing plants of different ages at different times, and grown in different conditions may be a source of errors. Confocal imaging of multiple fluorescent fusion proteins could help with both matters, as it provides a way of tracing proteins and studying how they co-localize. Following the development of the same plant through time is also possible with this technique.

4.4.1. Lack of mutant data

The core of our approach is based on the use of ISH data to approach expression profiles at single-cell resolution to infer regulatory interactions. Unfortunately, this kind of data is usually not available for mutants. This has consequences for the models that can be generated. Indeed, real biological regulatory networks are usually robust to mutations, as regulators are often encoded by a family of related genes, providing redundancy. However, our

genetic programming algorithm aims at generating models as simple as possible, and as we have little data about expression profiles in mutants, the algorithm has no reason to try to replicate the robustness of the real network. This means the algorithm will build models featuring little – if any – redundancy.

4.4.2. Applicability of the method

This method, based on co-expression profiles and genetic programming, has been successfully applied to the case of the network controlling cell identity in the SAM. Although it has not been tested on other biological networks, it should be applicable to other networks providing appropriate data sets are available. It would be interesting to see how well the method performs on other cases, and, in particular, if the trade-off between computation time and quality of the output models is satisfactory across all cases. It is entirely possible that this trade-off could be improved using a different set of parameters for the genetic-programming algorithm, both as default values and as problem-specific values. This is because little optimization has been carried out in this area, due to the high computational cost associated with it.

4.4.3. Roles of AP1 and TFL1

This work suggests that *AP1* represses *FD*. While this was not reported by Liu et al. or Fornara et al., it has since been published (Kaufmann et al., 2010). The genetic programming output also suggested AP1 does not necessarily need to directly down-regulate *SOC1*, as this would be redundant with an indirect

repression via *FD*. This might be tested experimentally in an *FD*-overexpressing plant. If *SOC1* is not down-regulated in floral primordia, it would confirm that the repression of *SOC1* by *AP1* goes through *FD*. Alternatively, it is possible that both regulatory features occur and this is a case of feed-forward repression.

One of the aims of this work is to investigate the place of TFL1 in the regulation of cell identity in the SAM. To make this possible, variables inner and apex were introduced for the following reasons. First, very little is known about the regulation of *TFL1*, which makes it difficult to produce models where *TFL1* is expressed in the right conditions. The patterning of *TFL1* is, however, very similar to that of *WUS*, for which a patterning mechanism combining inhibition in outer tissues and sensitivity to activation in inner tissues has been proposed (Chickarmane et al., 2012). An "inner" node was added to the network to enable similar models for *TFL1*. Second, *TFL1* seems to affect the identity of CZs. Indeed, floral meristems, which are usually determinate, become indeterminate and generate recursive cauliflower-like patterns in the *ap1*/*cal* mutants, where *TFL1* is expressed ectopically. Conversely, the SAM becomes determinate in *tfl1* mutants, as the meristem turns into a flower after the floral transition. Since the apices of the SAM and floral meristems appear to have similar behaviors in some genetic backgrounds, we postulated that those apices share some unknown properties responsible for this shared behavior, and introduced a variable called "apex" accordingly.

It is not clear which molecular species correspond to the spatial information implied by variables inner and apex, but some genes exhibit the relevant

expression patterns. Inner seems to correlate with *AHK4* (Chickarmane et al., 2012) and apex to *CLV3* (Geier et al., 2008). Interestingly, these two genes are involved in the *WUSCHEL-CLAVATA* negative feedback loop. As *WUS* and *TFL1* share similar expression patterns and their expression levels are correlated, it seems likely that *TFL1* and genes of this loop are somehow connected. Should it not be the case, the patterns of *AHK4* and *CLV3* still prove that genes with patterns appropriate to explain those of inner and apex do exist.

4.4.4. Extension to quantitative modelling

Inferring a quantitative model of the floral transition – such as an ODE or PDE model - by genetic programming might be possible. The major challenges, however, are that it would add a parameter optimization problem for each system of equations to assess, and the simulations of ODE models are more expensive than those of Boolean models. However, instead of trying to infer a quantitative model directly, another approach could be to convert the Boolean models into ODE models using predefined methods (Mendoza and Xenarios, 2006; Wittmann et al., 2009). These quantitative models could then be simulated in a spatially explicit context, such as a 3D tissue mesh, which would enable the simulation of transitions in a more explicit way (growth, cell division, diffusion, transport). The main limitation of such developments is the lack of any nondestructive experimental method to measure quantitatively the gene expression patterns of cells *in situ* in organs, so that the quantitative outputs

of differential models would have no experimental counterpart for comparison.

4.5. CONCLUSION

In this paper we have described a succession of approaches aiming to build Boolean models able to reproduce a set of spatio-temporal gene expression patterns, whilst complying with prior knowledge on the regulatory topology. Starting from a brute force approach exhaustively enumerating a list of candidate models, we have been led to more sophisticated developments based on genetic programming. The latter were required by this case study. It seems likely that other systems involving cell differentiation and tissue patterning would require similar refinements, but it might be, in cases where prior biological knowledge is detailed enough, that the simplest approach leads to relevant conclusions. Therefore, the results have included all the different steps with some details, as summarized now.

The most naïve search strategy, exhaustive search, can only be carried out on very simple models, though it can be improved upon by restricting the search space to models conforming to a predetermined network topology. This drastically simplifies the problem, however, it might still not be enough if the network topology is too complex. Another issue is that it requires a sufficiently comprehensive network topology, which might not be available. However, even if these two problems do not arise, solutions generated this way may not

be satisfying, as they are likely to involve complex, unlikely regulation mechanisms.

These three problems are addressed by the genetic programming algorithm used here. The family of genetic algorithms is known to be efficient at exploring high-dimensional spaces, such as the space of all Boolean models involving a set of nodes. Genetic programming has the added benefit of being able to generate Boolean equations directly, which makes it easier to target models involving simpler, more plausible regulatory interactions. This algorithm has successfully been applied to the regulatory network controlling cell identity in the SAM, resulting on the formulation of several plausible models and the suggestion of novel regulatory interactions absent from the starting network topology, but confirmed by independent laboratory work.

4.6. MATERIAL AND METHODS

4.6.1. Data

4.6.1.1. Classical 3-identity model. There are traditionally three characterized identities for cells constituting the SAM: vegetative, inflorescence and floral (Adrian et al., 2009). Some genes are commonly considered as characteristic of these profiles (Table 4.1). The vegetative profile represents any cell of the vegetative (pre-transition) SAM, as they do not seem to differ in the expression of any of the considered genes. The inflorescence profile represents cells of the main shoot of the inflorescence meristem (i.e.: primordia are excluded). The floral profile represents cells of the floral primordia. FT is necessary to induce the shift from vegetative to inflorescence in the OC and CZ, but once the inflorescence identity of CZ cells is acquired, FT is no longer required (memory effect).

4.6.1.2. Developmental transformations. The development of the SAM is assumed to take place through the occurrence of perturbations making the system transition from one steady state to another (Table 4.3).

4.6.1.3. Mutant phenotypes. In *tfl1* mutants, a terminal flower develops at the apex of the meristem. Another interesting case is the *ap1*/*cal* double mutant. *CAL* is a close homolog of *AP1*. When both are knocked out, the inflorescence develops into a cauliflower shape, where meristem primordia turn into inflorescence meristems and recursively generate new primordia. This information is summarized in Table 4.4.

4.6.2. Genetic programming

Three criteria come into play in the fitness function, listed below in order of priority.

1. $n_{violated}$: the sum of the XOR distances between the required end steady states and the end steady states reached by the model, for the species deemed relevant (lower is better, always 0 for solutions to the problem); For each (I, P, F, C, M) transition (see S3 Text), attractor(P(I)) is calculated. If the latter is a steady state, the distance between attractor(P(I)) and F is the number of non-zero values in (P(I) XOR F) AND C. Otherwise, if attractor(P(I)) is a cycle, the model is rejected and the distance is set to the number of non-zero values in C;
2. $n_{interactions}$: the number of novel (i.e. not present in the data, see details below) interactions in the model (lower is better). For efficiency reasons, this is based on the equations of the model rather than its truth table. A novel interaction ij is considered included in a

model if and only if i appears in the equation of j and interaction ij is not in the prior knowledge.

3. n_{terms} : the number of terms in the equations (including operators, lower is better). It is given by the number of nodes in the tree of the model. In order to optimize the fitness function, genetic programming algorithms produce successive generations of offspring.

The formula of the fitness function is presented in Equation 4.2.

$$n_{violated} - \frac{1}{1 + n_{interactions} - \frac{1}{1 + n_{terms}}} \qquad (4.2)$$

This function does not allow any kind of trade-off: criteria with lower ranks always have priority over those with higher ranks.

As genetic algorithms can potentially get stuck in local minima of fitness functions, the scheme devised here mitigates this issue by running the algorithm multiple times and introducing transition data both progressively and in a different order each time. Each run follows the following process:

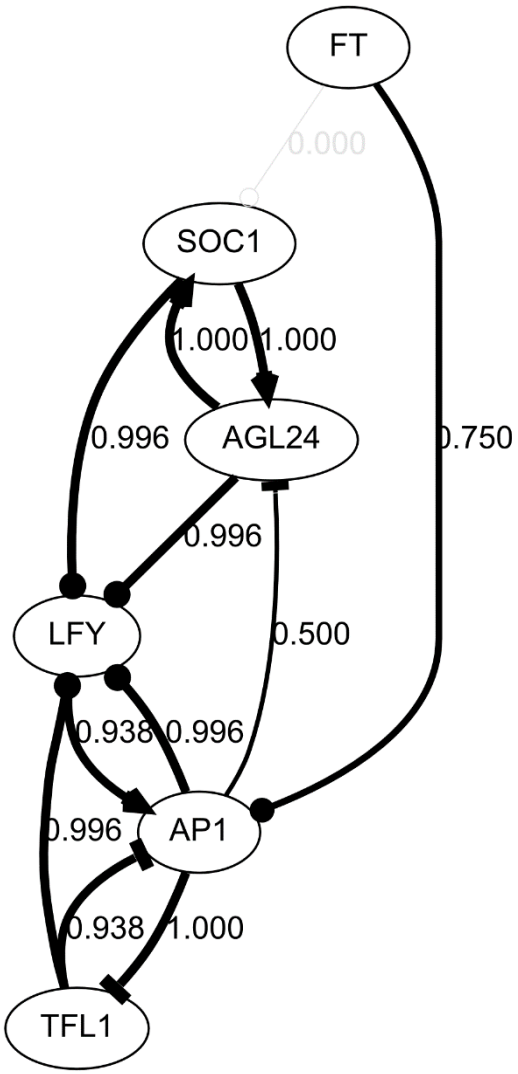
1. Establish a dataset D of known transitions;
2. Create an empty dataset D';
3. Pick a transition in D randomly, and move it into D';
4. Run the genetic programming algorithm until a solution that does not violate any transition in D' is found or the algorithm times out (i.e. no solutions could be found in a preset number of generations after the

latest transition was added). Repeat from step 3 until D is empty and a solution compatible with D' is found (unless a time-out occurs);

5. If such a solution is found, keep running the genetic programming algorithm for a fixed number of iterations to come up with a simplified form. Save the best individual as a solution.

Running this algorithm multiple times generates different solutions.

4.7. SUPPORTING INFORMATION



S1 Fig. Aggregate graph of the models generated by exhaustive search on the topology reported by Fornara and colleagues. Nodes are genes. Edges represent regulatory interactions. Edge labels and edge thicknesses denote the occurrence frequencies of the associated interactions. V-, T- and O-shaped arrowheads indicate positive, negative and ambiguous interactions, respectively.

S4.1 Table. List of the genes and time points extracted from *in situ* hybridization images, and their sources. Dates are expressed as days after germination (dag), days after induction (dai) or as developmental stages when no other information was available (vegetative, transition or inflorescence).

Genes and times	Reference
FD (6, 8, 10 dag) *SOC1* (6, 10 dag)	(Searle et al., 2006)
TFL1 (7, 14, 17 dag; inflorescence) *AP1* (inflorescence) *LFY* (inflorescence)	(Liu et al., 2013)
FD (0, 4, 5, 6 dai) *AP1* (0, 4, 5, 6 dai)	(Wigge et al., 2005)
TFL1 (12 dag) TFL1 protein (12, 16 dag)	(Conti and Bradley, 2007)
AP1 (inflorescence) *SOC* (inflorescence)	(Liu et al., 2007)
SOC1 (0, 1, 3, 5 dai) *AP1* (0, 3, 5 dai)	(Wang et al., 2009)
AGL24 (inflorescence)	(Michaels et al., 2003)
LFY (inflorescence)	(Blazquez et al., 1997)

S1 Text. Boolean modeling.

Boolean networks

Boolean network models represent genes as binary variables (either on or off) that influence each other dynamically, following a specified set of logical rules that can be written using combinations of the AND, OR and NOT operators

(Kauffman, 1969). The state of the network (i.e.: of each gene) at a given time point depends on the state of the network at the previous time point. The function that yields the next state of the model when given any state as an input is called the successor function, and that next state is called the successor of the state given as input.

As the number of states of a network is finite, a chain of successors starting at any state will sooner or later include at least one state more than once and initiate a periodic pattern. That periodic pattern constitutes an attractor. If the chain ends with the repetition of a single successor, this attractor is a steady state. If the chain ends with the repetition of multiple states, this attractor is a cycle.

Another consequence of the number of states being finite is that it is possible to establish an exhaustive list of states and their successors, for any model (or part of a model). This list is usually presented as a truth table, which is a table divided into a left-hand side and a right hand side. The left hand side lists all the possible states of the regulators of the genes of interest, while the right-hand side contains the matching states of the genes of interests at the following time step.

Updating scheme

Synchronous updating (i.e.: multiple variables can change their values per time step) was chosen over asynchronous updating (i.e.: only one variable can change its value per time step). Asynchronous updating is generally considered

more realistic, as, in reality, time is continuous, so multiple genes are unlikely to change their states at once. However, asynchronous updating has a drawback: when multiple genes might change their states at once, the Boolean states typically have multiple potential successors, leading to non-deterministic outcomes. Furthermore, steady states are independent of the updating scheme.

S2 Text. Exhaustive search.

Since each model is characterized by a truth table, exhaustive search works by enumerating all possible truth tables. Truth tables with empty right-hand sides are first generated for each internal node of the regulatory network. The right-hand sides of these truth tables are then filled as much as possible with information extracted from the steady states, using the fact that, for a steady state A, successor(A) = A (i.e. if the left side of a row matches A, then its right side should also match A). At this stage, an incompatibility between two steady states can occur (i.e. the left side of a row matches two steady states A and B, however the right side cannot match A and B at the same time), in which case the search problem has no solution. If all steady states are compatible, we proceed to iterate over all possible values for the empty cells of the truth tables, and record the models that have all the required transitions as valid, if any have been defined. If not, then all models are considered valid.

S3 Text. Verifying if a model can explain a transition.

Transitions are implemented as tuples (I, P, F, C, M) where:

- I is an initial steady state
- P is a perturbation to be applied to I. The resulting state is P(I).
- F is the steady state that P(I) is supposed to lead to, if it is left to evolve spontaneously. For a given model m, if $\text{attractor}_m(P(I)) = F$, then the model can explain this transition. If $\text{attractor}_m(P(I))$ is a cycle, model m is rejected, even if $\text{attractor}_m(P(I))$ contains F.
- C is a certainty mask used to modulate the comparison between $\text{attractor}_m(P(I))$ and F. It is a Boolean vector of the same size as I and F. 1 values in C indicate the associated variables in $\text{attractor}_m(P(I))$ and F should be taken into account for the comparison, 0 values mean they should not. This means model m can explain the transition if all variables in ($\text{attractor}_m(P(I))$ XOR F) AND C are 0.
- M is a list of mutations. If it is not empty, the transition should apply to a "mutant variant" of model m (instead of applying to model m directly).

S4 Text. Modeling of mutants.

Let f be the successor function of a WT model, and f_i the function giving its i-th component. Let f^j be the successor function of the same model, with a knock-out mutation of species j, and f_i^j the function yielding its i-th component. Let X be a state of the system.

$$\forall i \neq j, \qquad \forall X, \qquad f_i^j(X) = f_i(X)$$

$$f_j^j(X) = 0$$

S5 Text. Genetic programming.

Genetic programming is a method to generate structured sequences like computer code or mathematic equations using a genetic algorithm. We applied this method to the generation of Boolean models using the DEAP module (Fortin et al., 2012) in Python.

Structured sequences can be written as a tree. In the case of equations, a node is a function, and the children of that node are its arguments.

In genetic programming, the nodes of the trees are called primitives.

Primitives

In our case, the leaf primitives are the nodes of the GRN. They can be combined into Boolean expressions using AND, OR and NOT nodes. Finally, at the top level, Boolean expressions are aggregated into a list of Boolean expressions (one expression for each state variable of the model).

Primitive	Type	Arguments
List maker	List	6 bools
And	Bool	2 bools
Or	Bool	2 bools
Not	Bool	1 bool
Nodes of the GRN	Bool	None

Offspring generation

At each iteration of the algorithm, for a population of n individuals, n offspring individuals are generated and added to the population. Each of the offspring individual is generated randomly by either mutation, mating or reproduction. The respective probabilities of these events were chosen arbitrarily, and are given below, as well as descriptions of the processes.

Mutation

A branch of the tree is replaced with a random branch. This occurs with probability 0.4.

Mating

Exchange of branches of the tree related to the same genes. This occurs with probability 0.4.

Reproduction

An individual is copied as-is. This occurs with probability 0.2.

Constraint on the trees

The maximal depth of the tree is capped to 11 in order to avoid bloat. Lower values reduce the size of the search space, but if they are too low, they can prevent solutions from being found.

Selection

n individuals are selected out of the n original individuals and their n offspring indivuals using n 2-invidual tournaments. The whole population is split randomly into n pairs, and the better individual of each pair is selected to be part of the next generation.

S6 Text. Model analysis.

Adjacency matrix

For each state of the model, the effect of switching on inactive genes one at a time was recorded in an adjacency matrix whose values (-1, 0, 1 or 2) indicate the type of each regulatory interaction.

Let n be the number of species in the model. Let A be a $n \times n$ matrix. Let $X = (X_1, \dots, X_i, \dots, X_n)$ be a model state. Let $X_i^* = (X_1, \dots, 1, \dots, X_n)$ be a state derived from X by setting the value of the i-th node to 1. Let f be the successor function of the model. Let $Y = f(X) = (Y_1, \dots, Y_j, \dots, Y_n)$ and $Y_i^* = f(X_i^*) = (Y_i^*{}_1, \dots, Y_i^*{}_j, \dots, Y_i^*{}_n)$.

For each (i, j) pair:

- If, for all X, $Y_j = Y_i^*{}_j$, then $A_{ij} = 0$
- Else, if, for all X, $Y_j \leq Y_i^*{}_j$, then $A_{ij} = 1$
- Else, if, for all X, $Y_j \geq Y_i^*{}_j$, then $A_{ij} = -1$
- Else, $A_{ij} = 2$

Aggregation of multiple models

The following method was used to aggregate the adjacency matrices of a set of models K into a single matrix.

Let A be the aggregated adjacency matrix and A^k the adjacency matrix of model k.

- If $\forall k \in K,\ A_{ij}^k = 0$, then $A_{ij} = 0$.
- If $\forall k \in K,\ A_{ij}^k \in \{0, 1\}$, then $A_{ij} = 1$.
- If $\forall k \in K,\ A_{ij}^k \in \{-1, 0\}$, then $A_{ij} = -1$.
- If $\exists k \in K,\ A_{ij}^k = -1$ and $\exists k' \in K,\ A_{ij}^{k'} = 1$, then $A_{ij} = 2$.
- If $\exists k \in K,\ A_{ij}^k = 2$, then $A_{ij} = 2$.

Graph representation

Let A be the adjacency matrix of a model and Gits graph.

- If $A_{ij} \neq 0$, G includes the regulatory edge $i \to j$.
- If $A_{ij} = 1$, i is an activator of j.
- If $A_{ij} = -1$, i is a repressor of j.
- If $A_{ij} = 2$, the effect of i on j is ambiguous.

Principal components analysis (PCA)

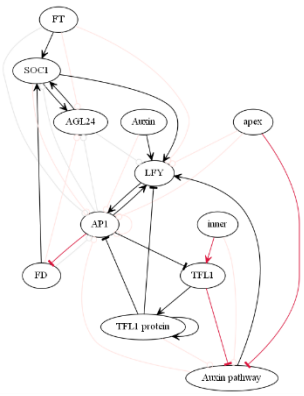
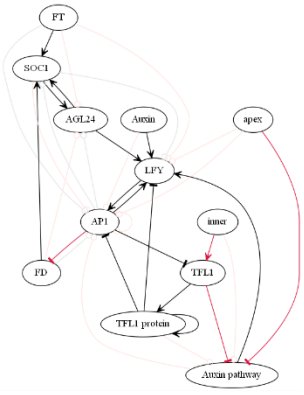
PCA was carried out on sets of models to assess their diversity.

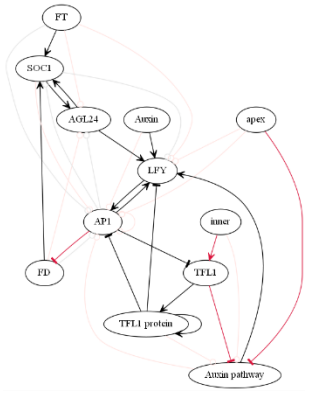
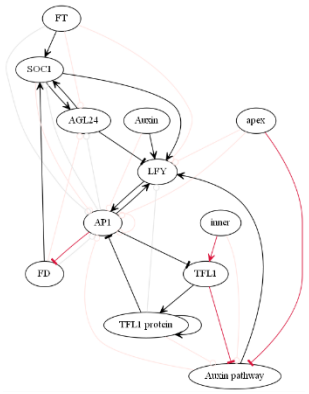
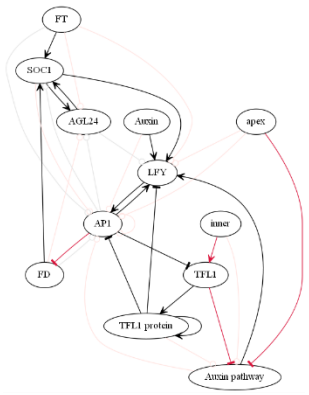
First, a matrix of Boolean values was built, where:

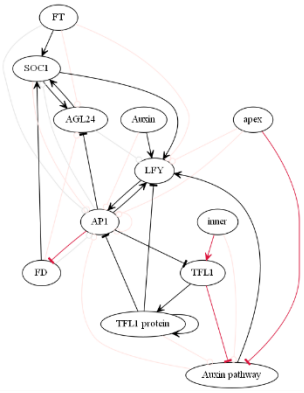
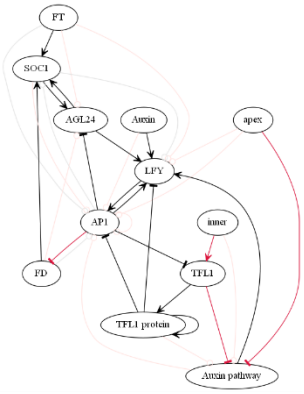
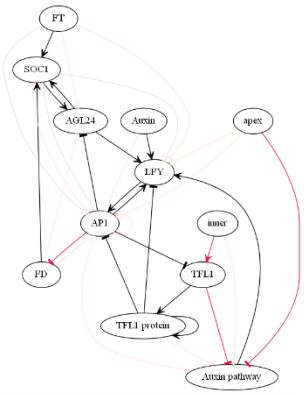
- Each row is a row vector of Boolean variables indicating which interactions are present in a model;
- Each column corresponds to a directed interaction edge in the GRN. Only interactions that vary in the set of models are retained (the variance of the column vector is greater than 0).

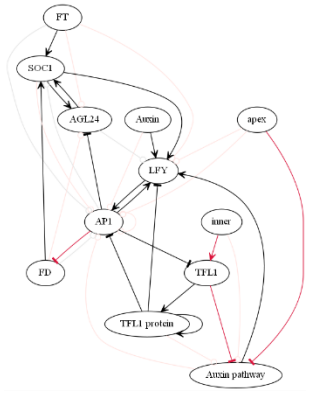
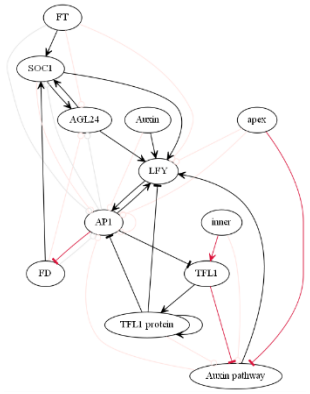
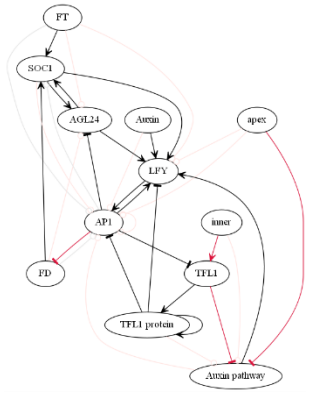
PCA was then performed using the Scikit-learn module (Pedregosa et al., 2011) for Python.

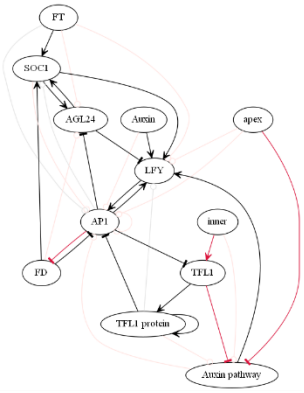
S4.2 Table. Equations and graphs of the twelve best models from the genetic programming search.

Rank	Equations	Graph
1a	TFL1 protein' = or(TFL1 protein, TFL1) Auxin pathway' = not(or(apex, TFL1)) FD' = not(AP1) SOC1' = and(or(AGL24, FT), FD) AGL24' = SOC1 LFY' = and(or(AP1, and(SOC1, Auxin)), or(not(TFL1 protein), Auxin pathway)) AP1' = and(LFY, not(TFL1 protein)) TFL1' = and(inner, not(AP1))	
1b	TFL1 protein' = or(TFL1, TFL1 protein) Auxin pathway' = not(or(TFL1, apex)) FD' = not(AP1) SOC1' = and(or(FT, AGL24), FD) AGL24' = SOC1 LFY' = and(or(not(TFL1 protein), Auxin pathway), or(and(AGL24, Auxin), AP1)) AP1' = and(LFY, not(TFL1 protein))	

	TFL1' = and(inner, not(AP1))	
1c	TFL1 protein' = or(TFL1 protein, TFL1) Auxin pathway' = not(or(TFL1, apex)) FD' = not(AP1) SOC1' = and(or(AGL24, FT), FD) AGL24' = SOC1 LFY' = or(AP1, and(and(Auxin, AGL24), or(not(TFL1 protein), Auxin pathway))) AP1' = and(not(TFL1 protein), LFY) TFL1' = and(inner, not(AP1))	
2a	TFL1 protein' = or(TFL1, TFL1 protein) Auxin pathway' = not(or(apex, TFL1)) FD' = not(AP1) SOC1' = or(and(FD, FT), and(FD, AGL24)) AGL24' = SOC1 LFY' = or(and(or(not(AGL24), Auxin pathway), and(SOC1, Auxin)), AP1) AP1' = and(not(TFL1 protein), LFY) TFL1' = and(inner, not(AP1))	
2b	TFL1 protein' = or(TFL1, TFL1 protein) Auxin pathway' = not(or(TFL1, apex)) FD' = not(AP1) SOC1' = or(and(FD, AGL24), and(FT, FD)) AGL24' = SOC1 LFY' = or(and(SOC1, and(Auxin, or(not(TFL1 protein), Auxin pathway))), AP1) AP1' = and(not(TFL1 protein), LFY) TFL1' = and(inner, not(AP1))	

3a	TFL1 protein' = or(TFL1 protein, TFL1) Auxin pathway' = not(or(TFL1, apex)) FD' = not(AP1) SOC1' = or(AGL24, and(FT, FD)) AGL24' = and(not(AP1), SOC1) LFY' = or(AP1, and(or(Auxin pathway, not(TFL1 protein)), and(SOC1, Auxin))) AP1' = and(LFY, not(TFL1 protein)) TFL1' = and(inner, not(AP1))	
3b	TFL1 protein' = or(TFL1, TFL1 protein) Auxin pathway' = not(or(TFL1, apex)) FD' = not(AP1) SOC1' = or(and(FT, FD), AGL24) AGL24' = and(not(AP1), SOC1) LFY' = and(or(AP1, and(Auxin, AGL24)), or(not(TFL1 protein), Auxin pathway)) AP1' = and(not(TFL1 protein), LFY) TFL1' = and(inner, not(AP1))	
3c	TFL1 protein' = or(TFL1 protein, TFL1) Auxin pathway' = not(or(TFL1, apex)) FD' = not(AP1) SOC1' = or(AGL24, and(FD, FT)) AGL24' = and(SOC1, not(AP1)) LFY' = or(AP1, and(and(AGL24, or(Auxin pathway, not(TFL1 protein))), Auxin)) AP1' = and(not(TFL1 protein), LFY) TFL1' = and(inner, not(AP1))	

3d	TFL1 protein' = or(TFL1 protein, TFL1) Auxin pathway' = not(or(TFL1, apex)) FD' = not(AP1) SOC1' = or(and(FT, FD), AGL24) AGL24' = and(SOC1, not(AP1)) LFY' = and(or(and(Auxin, SOC1), AP1), or(Auxin pathway, not(TFL1 protein))) AP1' = and(LFY, not(TFL1 protein)) TFL1' = and(inner, not(AP1))	
4	TFL1 protein' = or(TFL1 protein, TFL1) Auxin pathway' = not(or(apex, TFL1)) FD' = not(AP1) SOC1' = and(FD, or(FT, AGL24)) AGL24' = SOC1 LFY' = and(not(not(or(and(Auxin, SOC1), AP1))), or(and(AGL24, Auxin pathway), not(TFL1 protein))) AP1' = and(not(TFL1 protein), LFY) TFL1' = and(inner, not(AP1))	
5	TFL1 protein' = or(TFL1 protein, TFL1) Auxin pathway' = not(or(TFL1, apex)) FD' = not(AP1) SOC1' = or(and(FD, FT), AGL24) AGL24' = and(not(AP1), SOC1) LFY' = or(and(Auxin, or(and(Auxin pathway, AGL24), and(SOC1, not(TFL1 protein)))), AP1) AP1' = and(not(TFL1 protein), LFY) TFL1' = and(inner, not(AP1))	

6	TFL1 protein' = or(TFL1, TFL1 protein) Auxin pathway' = not(or(TFL1, apex)) FD' = not(AP1) SOC1' = or(and(FT, FD), AGL24) AGL24' = and(SOC1, not(AP1)) LFY' = and(or(AP1, and(Auxin, SOC1)), or(Auxin pathway, not(AGL24))) AP1' = not(and(FD, or(not(LFY), TFL1 protein))) TFL1' = and(inner, not(AP1))	

5. 4D MODEL OF THE FLORAL TRANSITION IN ARABIDOPSIS THALIANA

5.1. INTRODUCTION

This chapter was realized in collaboration with the Virtual Plants team (Inria, CIRAD, University of Montpellier 2), in particular Eugenio Azpeitia and Christophe Godin.

The previous chapter demonstrated the feasibility of designing a set of logical rules resulting in the spatial self-organization of a meristem, although space was only modelled through proxy variables. In this chapter, the possibility of implementing a model in a 3D tissue structure, resulting in the self-organization of a meristem, was studied.

The emergence of spatial patterns from homogeneous systems, as a response to the diffusion and the reaction of biochemical species in living organisms, has famously been theorized by Turing (Turing, 1952). Gierer and Meinhardt later proposed equations to describe the mechanisms underlying the formation of various patterns, including dots and stripes (Gierer and Meinhardt, 1972; Meinhardt and Gierer, 1974). They describe patterns formed by one category of Turing mechanism resulting from the interplay between a short-range activator and a long-range inhibitor, in a continuous medium. However, it is also possible to generate spatial patterns in discrete media, such as a tissue with individual cells. In plant systems, Jönsson and colleagues, whose work this chapter builds upon, have modelled the polar transport of auxin in the SAM. It

belongs to another category of Turing mechanism: substrate depletion (Cheong et al., 2010). Auxin is transported from cell to cell by PIN1 transporters, predominantly towards cells with higher auxin concentrations, creating a positive feedback loop, and simultaneously draining auxin away from lower concentration areas. Auxin is a prime example of a species whose spatial distribution is important to model, because of its role in cell elongation (Rayle and Cleland, 1992), and therefore, morphogenesis. As a consequence, its patterning has also been studied in other organs, such as roots (Band et al., 2014).

Here, the patterning of auxin was studied in conjunction with that of meristem identity genes to understand how the spatial organization of the SAM is achieved during the floral transition.

5.2. MATERIAL AND METHODS

5.2.1. Tissue structure

The structure used in this model is based on a real inflorescence meristem (referenced YR01), imaged by confocal microscopy by Yassin Refahi, Lisa Willis, Raymond Wightman and Henrik Jönsson (Sainsbury Laboratory, Cambridge), using a protocol described by Willis and colleagues (Willis et al., 2016). It was then segmented and meshed by Sophie Ribes and Guillaume Cerutti (Virtual Plants, Montpellier).

As it is an inflorescence SAM structure, it includes all the domains of the post-floral-transition SAM studied in Chapter 4. Some of these domains are

temporally disjoint for a given cell, but due to the coexistence of cell populations of various developmental stages in the SAM, all domains can be observed simultaneously across the same SAM. This inflorescence SAM structure could also technically be used as a template for simulations of the pre-floral-transition SAM, but the vegetative and inflorescence SAM do have morphological differences (Liu et al., 2013) that make in-depth interpretations of such simulations more difficult.

5.2.2. Model hypotheses

The SAM tissue is subdivided into cells, and each cell has its own biochemical profile, which can be defined as the given of the concentrations in its modelled chemical species. Those concentrations are different for each cell, and vary over time due to synthesis, degradation, and intercellular exchanges. For the sake of simplicity, the intracellular diffusion of species is modelled as instant and unaffected by organelle boundaries, i.e. concentrations are always uniform within a cell. This is the usual assumption for such multicellular models (Jönsson et al., 2003; Angel et al., 2011; Band et al., 2012).

5.2.3. Equations of the ODE model

An ODE model describing the evolution of the quantities of matter of the relevant species inside of each cell was developed. The decision to model quantities of matter rather than concentrations, as is often the case, was made because the cells are of different sizes and mechanistic arguments suggesting that gene transcription should scale with cell size are absent.

The equations describing the regulatory mechanisms were derived from model 1a of Chapter 4, which was one of the three best (tied with 1b and 1c in fitness values). Model 1a was chosen over model 1b and 1c because it does not involve *AGL24* in the regulation of *LFY*. *AGL24* is only involved in a positive feedback loop with *SOC1*, therefore it can be removed from the network altogether and replaced with a *SOC1* self-activation.

Multiple formalisms were considered for the transformation of Boolean equations into ODEs, including SQUAD (Mendoza and Xenarios, 2006; Cara et al., 2007) and Odefy (Wittmann et al., 2009). Those had the benefit of being implementable in an automated way. They however had the drawback of having forms rather far removed from traditional ODE models, especially in the way they deal with OR operators. Therefore, the Hill and Shea-Ackers formalisms (Hill, 1910; Ackers et al., 1982; Alon, 2006) were used to translate the Boolean equations into the synthesis parts of the ODEs, using principles presented in Table 5.1. As a general rule, Hill and Shea-Ackers formulae are used at the top level of the functions to bound synthesis between 0 and 1, but are not used at deeper levels to keep the equations as simple and legible as possible. The AND operator is translated using multiplication, while the OR operator is translated using addition, as in the Shea-Ackers formula.

The most complex cases encountered are presented in Table 5.2. More complicated cases could theoretically arise and might then be hard to translate simply. However, it did not happen in model 1a, most likely because the genetic programming algorithm that generated it aimed at providing the

simplest equation possible. The deepest level of nesting for AND and OR operators is 3, and this only occurs once, in the equation of LFY, so this was not a concern. It could also be argued that deeper levels of nesting would anyway result in biologically implausible functions, as they would have to involve many regulators, or the regulators would have to interact in very irregular (i.e. non-factorable) ways. The former is limited by the topology of the network. As for the latter, the irregularity of the interactions is limited by the physicochemical properties of the interacting species.

Table 5.1. Principles guiding the transformation of Boolean equations into ODEs.

Boolean	ODE
x	$\frac{x^n}{\theta^n + x^n}$
$NOT\ x$	$\frac{\theta^n}{\theta^n + x^n}$
$x\ OR\ y$	$\frac{x^n + y^n}{\theta^n + x^n + y^n}$
$x\ AND\ y$	$\frac{x^n}{\theta^n + x^n} \cdot \frac{y^n}{\theta^n + y^n}$

Table 5.2. Examples of more complicated transformations of Boolean equations into ODEs.

Boolean	ODE
$x\ OR\ (y\ AND\ z)$	$\frac{x^n + (y.z)^n}{\theta^n + x^n + (y.z)^n}$
$x\ AND\ (NOT\ y\ OR\ z)$	$\frac{x^n}{\theta^n + x^n} \cdot \frac{\theta^n + z^n}{\theta^n + y^n + z^n}$

The transformation of Boolean equations to ODEs also requires scaling: the kind of input functions described in Table 5.1 and Table 5.2 only assume values between 0 and 1, but should lead to concentration ranges appropriate as inputs (e.g. x, y, z in that table) of the same functions. For cell autonomous species, at the steady state in a cell i, the relationship between an input function f and the concentration c_{A_i} it leads to is as follows:

$$c_{A_i} = \frac{A_i}{V_i} = \frac{a.f(x, y, z, \dots)}{k.V_i}$$

- c_{A_i}: concentrations in species A in cell i
- A_i: quantity of matter of species A in cell i
- V_i: volume of cell i
- a: synthesis coefficient
- $f(x, y, z, \dots)$: value of the input function, which depends on some concentrations $x, y, z\dots$
- k: degradation coefficient

As a convention, it was decided that higher concentration values should be on the order of magnitude of 1, so the θ parameter, which defines activation or

repression thresholds, was set to 0.3 for all species. The n parameter, which defines the steepness of the activation or repression, also has a single value (n=3) for all species. Using only one value per parameter for all equations was the most parsimonious option and yielded satisfactory results. The particular values of these parameters (and all others) were selected by trial and error and are given in Table 5.3.

For higher concentration values to actually be on the order of magnitude of 1, parameters a and k need to be chosen appropriately. As only the ratio $\frac{a}{k}$ matters, k was set to 1 for simplicity. a therefore has to take into account the size of cells. Setting a to the volume of the largest cell would ensure no concentration is higher than 1. However, there are some outliers in the distribution of cell volumes in the tissue (Figure 5.1), which means it would also result in most cells always having very high concentrations. This could be addressed by using the median volume instead. It would result in some concentrations being greater than 1, but the input functions can accommodate this, due to their saturating nature. However, another factor to take into consideration was the heterogeneity of cell volumes in the tissue. The distribution of cell volumes in the L1 is markedly different from the general one, with smaller cells on average (Figure 5.1). As this model was mainly aiming at predicting the identities of cells on the surface of the SAM (where lateral organs are initiated), the value of a was chosen to be the median volume of L1 cells, resulting in the higher concentrations of L1 cells being on the order of magnitude of 1.

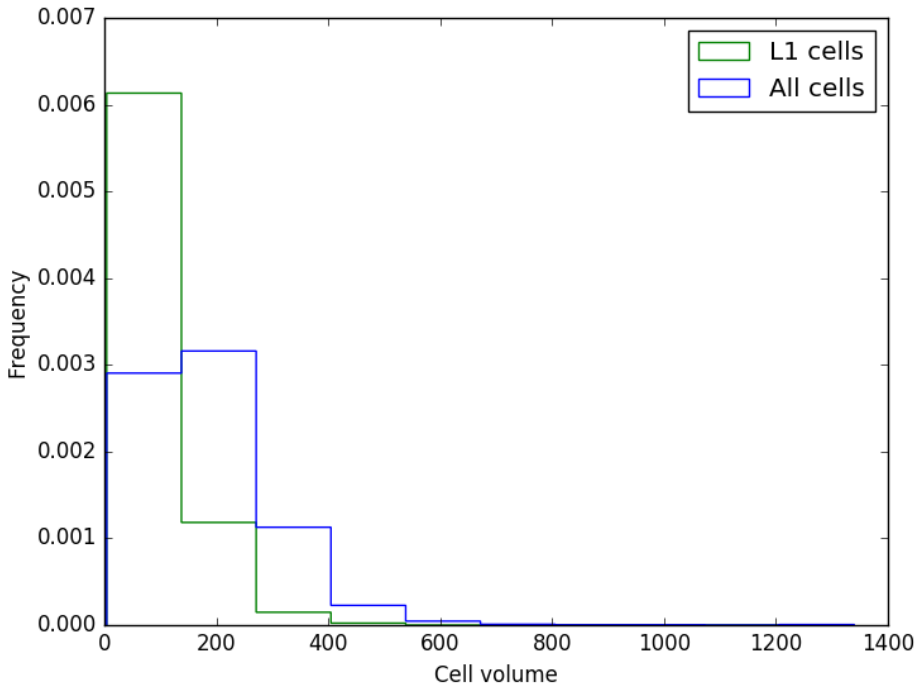


Figure 5.1. Histograms of the distributions of cell volumes in the whole tissue and in the L1 layer. The two populations have different numbers of cells, therefore the heights of the bars were normalized so that the integral of each distribution is 1 (similar to probability distribution functions).

Other parts of the ODEs had to be written from scratch, because they had no counterpart in the Boolean model. This is the case of the transport terms and degradation terms.

In the Boolean model, mobile species such as auxin and the TFL1 protein were simply considered as input variables. In a spatio-temporal model, it is however possible to model organ-scale transport phenomena explicitly. Passive transport between cells was therefore modelled as a diffusion-like process occurring through the interfaces between cells, using Fick's law (Attwood et

al., 2006). The interfaces between cells were modelled as membranes of uniform thickness, therefore the term describing the thickness of the membrane in Fick's law was assimilated into the diffusion coefficient. In the case of auxin, polar transport was computed using a model derived from previous work by Jönsson and colleagues (Jönsson et al., 2006). The model assumes that auxin is transported actively by a membrane-bound outflux transporter (PIN1) whose distribution is dynamic and favours the exportation of auxin towards cells with higher auxin concentrations. In the retained variant of the models proposed by Jönsson and colleagues, the distribution is not modelled explicitly as a state variable. It is assumed at a steady state with respect to auxin concentrations, and can therefore be computed directly from auxin concentrations. The efflux of auxin through an interface is then proportional to the number of PIN1 transporters on that interface and to the concentration of auxin in the source cell, following a mass action law.

Technically, FT is also a mobile species, however its transport was not modelled explicitly, as with auxin and the TFL1 protein. Instead, it was kept as an input variable, as it is synthesized outside of the SAM (in leaves), and there are no data detailing its distribution in the SAM. Its concentrations were therefore modelled as being solely time-dependent, and uniform throughout the SAM, at any given time. The activation of FT is represented in the equations by a Hill equation whose activator is time. This is to leave time for the auxin patterning to establish, which takes about 50 time units, before FT is introduced. It is meant to mimic the biological process as closely as possible.

This might be particularly important because the floral identity genes (*LFY* and *AP1*) constitute a positive feedback loop activated (indirectly) by a combination of FT and auxin. This means that the activation of *LFY* and *AP1* might not be reversible if the pattern of auxin changes after FT is introduced. In this particular case, using constant concentrations for FT ($c_{FT_i} = 1$) turned out to result in the same patterns as using the Hill equation, but there is no guarantee it would apply to other models.

Finally, degradation terms were simply modelled as exponential decays.

The resulting equations are described in Equation 5.1, and the associated parameters are presented in Table 5.3.

Equation 5.1. ODE model derived from Boolean model 1a generated by genetic programming.

$$\frac{dFT_i}{dt} = V_i \frac{t^n}{t^n + \tau^n} - FT_i$$

$$\frac{dFD_i}{dt} = V_{median_{L1}} \frac{\theta^n}{\theta^n + c_{AP1_i}^n} - FD_i$$

$$\frac{dSOC1_i}{dt} = V_{median_{L1}} \frac{c_{SOC1_i}^n + c_{FT_i}^n}{\theta^n + c_{SOC1_i}^n + c_{FT_i}^n} \frac{c_{FD_i}^n}{\theta^n + c_{FD_i}^n} - SOC1$$

$$\frac{dLFY_i}{dt} = V_{median_{L1}} \frac{c_{AP1_i}^n + c_{auxin_i}^n . c_{SOC1_i}^n}{\theta^n + c_{AP1_i}^n + c_{auxin_i}^n . c_{SOC1_i}^n} \frac{\theta^n + c_{ARF_i}^n}{\theta^n + c_{ARF_i}^n + c_{TFL1protein_i}^n} - LFY_i$$

$$\frac{dAP1_i}{dt} = V_{median_{L1}} \frac{c_{LFY_i}^n}{\theta^n + c_{LFY_i}^n} \frac{\theta^n}{\theta^n + c_{TFL1protein_i}^n} - AP1_i$$

$$\frac{dTFL1_i}{dt} = V_{median_{L1}} \frac{\theta^n}{\theta^n + c_{AP1_i}^n} I_{inner_i} - TFL1_i$$

$$\frac{d}{dt} TFL1protein_i = 100\, V_{median_{L1}} \frac{c_{TFL1_i}}{c_{TFL1_i} + \theta^n} - TFL1protein_i + D_{TFL1protein} \sum_{j \in neighbours(i)} S_{ij} \left(c_{TFL1protein_j} - c_{TFL1protein_i} \right)$$

$$\frac{d}{dt} auxin_i$$

$$= V_{median_{L1}} I_{L1_i} - \ auxin_i$$

$$+ 0.1\ D_{auxin} \sum_{j \in neighbours(i)} S_{ij} \left(c_{auxin_j} - c_{auxin_i}\right)$$

$$+ D_{auxin} \sum_{j \in neighbors(i) \cap L1} S_{ij} \left(c_{auxin_j} - c_{auxin_i}\right)$$

$$+ \ T_{auxin} \sum_{j \in neighbors(i) \cap L1} \left(c_{auxin_j} \frac{c_{auxin_i} S_{ij}}{\sum_{k \in neighbours(j) \cap L1} c_{auxin_k} S_{jk}} \right.$$

$$\left. - c_{auxin_i} \frac{c_{auxin_j} S_{ij}}{\sum_{k \in neighbours(i) \cap L1} c_{auxin_k} S_{ik}} \right)$$

$$\frac{dARF_i}{dt} = V_{median_{L1}} \left(1 - I_{apex_i}\right) \frac{\theta^n}{\theta^n + c_{TFL1_i}^n} - ARF_i$$

- FT_i, FD_i, $SOC1_i$, LFY_i, $AP1_i$, $TFL1_i$, $TFL1protein_i$, $auxin_i$, ARF_i: quantities of matter of *FT* mRNA, *FD* mRNA, *SOC1* mRNA, *LFY* mRNA, *AP1* mRNA, *TFL1* mRNA, TFL1 protein, auxin and *ARF* mRNA, respectively, in cell i (arb. unit).
- c_{FT_i}, c_{FD_i}, c_{SOC1_i}, c_{LFY_i}, c_{AP1_i}, c_{TFL1_i}, $c_{TFL1protein_i}$, c_{auxin_i}, c_{ARF_i}: concentrations in *FT* mRNA, *FD* mRNA, *SOC1* mRNA, *LFY* mRNA, *AP1* mRNA, *TFL1* mRNA, TFL1 protein, auxin and *ARF* mRNA, respectively, in cell i (arb. unit). For any species S, $c_{S_i} = \frac{S_i}{V_i}$.
- θ: threshold concentration used in the Hill functions regulating all species (arb. unit).
- n: Hill coefficient used in the Hill functions regulating all species (dimensionless).

- V_i: volume of cell i (arb. unit)
- $V_{median_{L1}}$: median volume of L1 cells (arb. unit)
- τ: time of the floral transition (arb. unit)
- D_{auxin} , $D_{TFL1protein}$: diffusion coefficients of auxin and the TFL1 protein, respectively (arb. unit)
- T_{auxin}: active transport coefficient of auxin (arb. unit)
- I_{L1_i}, I_{inner_i}, I_{apex_i}: indicator variables that are 1 if cell i is part of the L1 layer, the "inner" zone, the "apex" zone, respectively, and 0 otherwise.
- $neighbours(i)$: set of cells adjacent to cell i
- $L1$: set of cells belonging to the L1 layer

Table 5.3. Parameter values.

Parameter	Value
τ	100
n	3
θ	0.3
$V_{median_{L1}}$	95.15
$D_{TFL1protein}$	250
D_{auxin}	1000
T_{auxin}	100000

5.2.4. Fixed zones of the tissue

In order to achieve a plausible auxin pattern, several zones were defined. A first zone called L1 defines the outermost layer of cells of the meristem (Figure 5.2). It is the set of cells among which the active transport of auxin is assumed to take place. In reality, polar auxin transport might occur under the L1 as well, but it could not be observed (Vernoux et al., 2011). The border of the L1 constitutes a second zone (Figure 5.3). In this border, auxin concentrations are fixed. This is to help the auxin pattern anchor to the geometry of the SAM. Without this, auxin maxima are not able to form on the edges of the L1. The third zone is the apex. It corresponds to the centre of the L1 (Figure 5.4). Cells of this zone do not form lateral organ primordia in response to auxin (Reinhardt et al., 2000), possibly due to the absence of some genes of the ARF family (Vernoux et al., 2011). The final zone is the inner part of the meristem, situated in the centre of the bottom side of the SAM (Figure 5.5). This is where *TFL1* transcription and translation occur (Conti and Bradley, 2007).

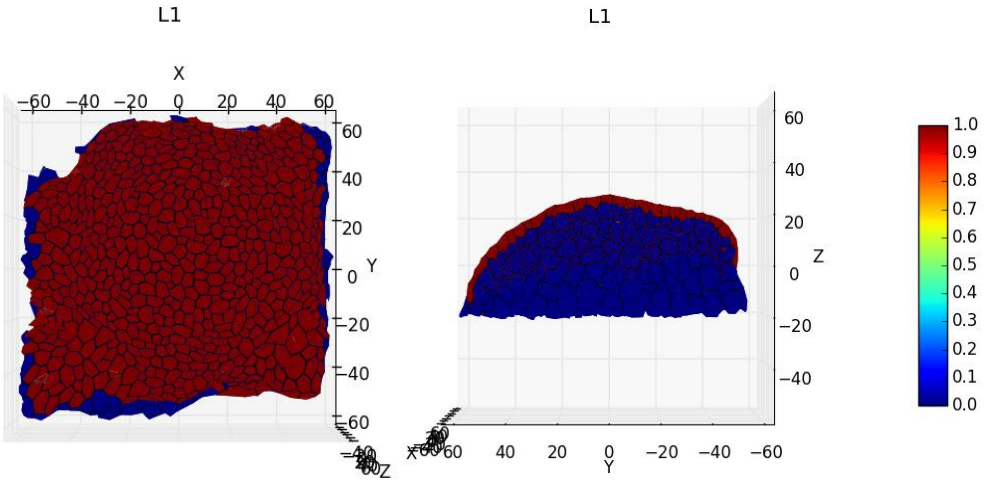


Figure 5.2. L1 layer (top view and cross-section). The L1 is where polar auxin transport takes place.

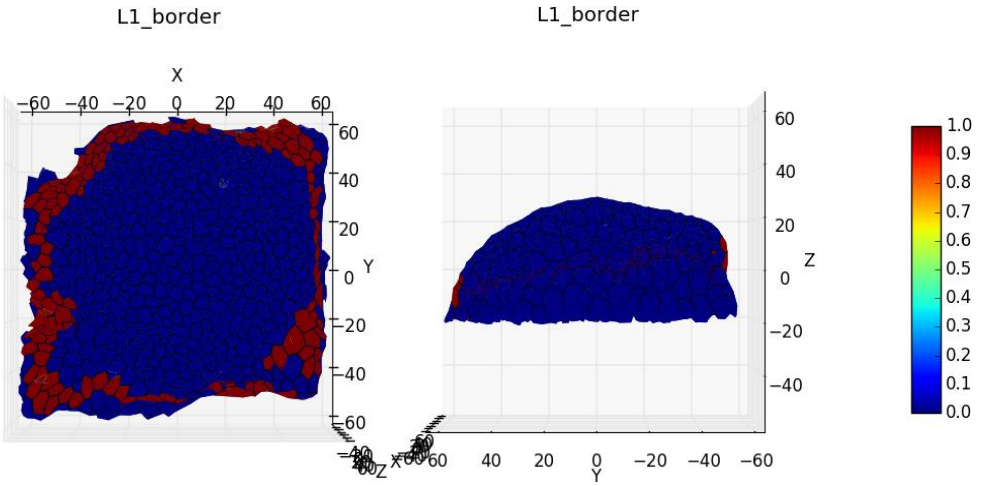


Figure 5.3. Border of the L1 layer (top view and cross-section). Those cells are on the edge of the imaged section of the SAM. Their auxin concentrations are fixed to anchor the auxin pattern to the geometry of the meristem.

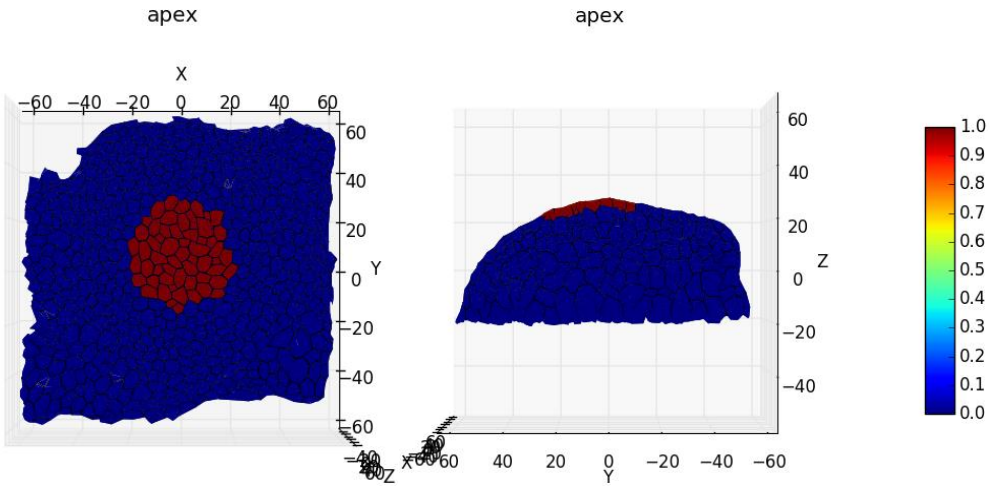


Figure 5.4. Apex zone (top view and cross-section). This zone is insensitive to auxin and cannot form new primordia in WT *A. thaliana* plants.

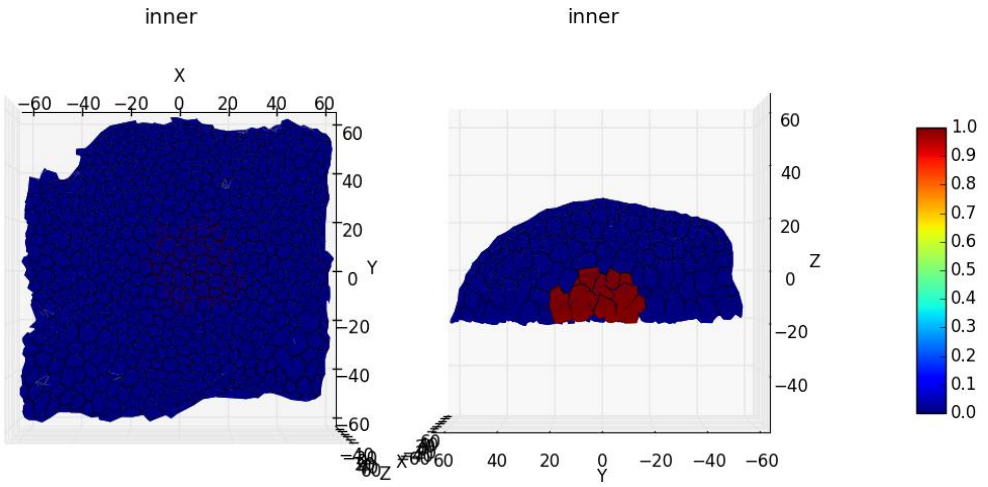


Figure 5.5. Inner zone (top view and cross-section). It is where *TFL1* is expressed.

5.2.5. Initial concentrations

All initial concentrations except those of auxin were initialized to 0 plus random uniform noise of amplitude 0.001. In the case of auxin, visible primordia were marked, and auxin concentrations were defined in primordia

cells following radial, linear concentration gradients centred on each primordium (Figure 5.6).

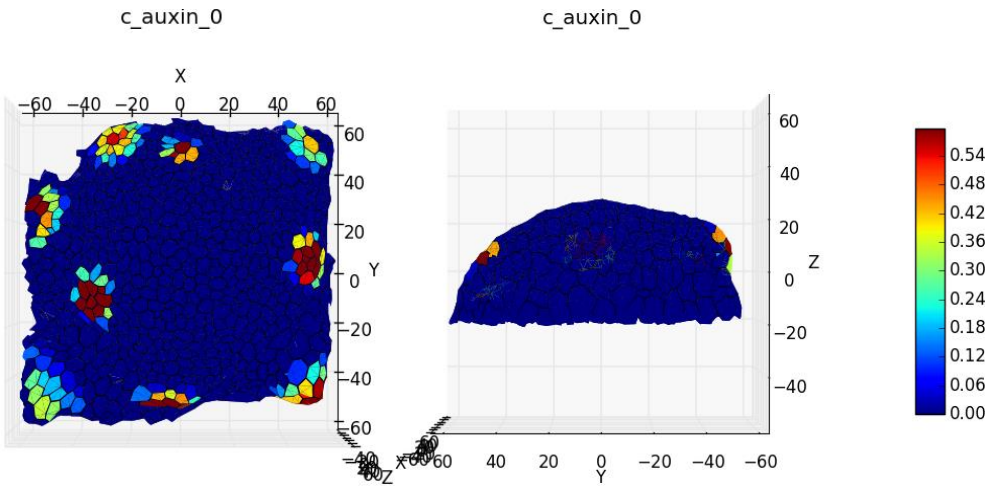


Figure 5.6. Initial auxin concentrations (t=0, top view and cross-section). The locations of primordia were marked approximately, and auxin concentrations were set according to a linear gradient, decreasing from the centres of primordia to their edges.

5.2.6. Software implementation

The model was implemented in Python using the Multicell framework, developed during this PhD project. Multicell is open source (MIT license) and available from Github (Dinh, 2016).

5.2.6.1. Tissue structure

Multicell uses tissue structures stored in the Topomesh format defined by OpenAlea (Dufour-Kowalski et al., 2007; Pradal et al., 2008; Cokelaer et al., 2009).

5.2.6.2. ODE integration

The equations presented in Equation 5.1 have to be integrated for each cell of the tissue, and the behaviour of each cell is dependent on those of its neighbours, due to transport phenomena (both passive and active). This results in large systems of ODEs that cannot be integrated independently of each other.

Under the assumption that all cells depend on every other cell, the integration of such a system can be done using standard solvers such as LSODA from the ODEPACK collection (Hindmarsh, 1982), which is the default solver in the Scipy module in Python (Jones et al., 2001) and the deSolve package in R (Soetaert et al., 2016). However, using LSODA results in very long simulation times, which can be drastically improved by using another solver. The core idea is that only cells that are adjacent depend on each other, therefore each ODE only depends directly on a few others. ODEPACK includes an efficient solver for this type of problems, LSODES, which specialises in systems of ODEs with a sparse Jacobian matrix (i.e. systems of ODEs where each ODE only depends directly on a few others). A Python wrapper for LSODES was developed by John Fozard (formerly University of Nottingham, now John Innes Center, Norwich) (Fozard, 2015), and is used by Multicell. A simple benchmarking case (passive transport of a species from the left-most cell in a row of 1000 cubic cells to the others, until homogenization) showed that using LSODES instead of LSODA resulted in a 31-fold speed increase. Such a speed-up is particularly relevant in real simulations, which already take 4 or more hours with LSODES.

5.3. RESULTS

5.3.1. Induction of FT in WT plants results in the expected expression patterns

The simulation results presented hereafter were obtained after running the simulation for 300 time units, which corresponds to an FT concentration of 0.99. At this time, the patterns of every species have stabilized, and this marks the end of the floral transition.

The objective was to reproduce expression or distribution patterns similar to those observed by ISH or other methods in real plants, in published experiments. The experimental results used as reference have been adapted and are depicted in Figure 5.8, Figure 5.10, Figure 5.12, Figure 5.14, Figure 5.16, Figure 5.18, Figure 5.20 and Figure 5.22.

FT concentrations are only time-dependent and simulate an influx of FT proteins from the leaves. They are homogeneous across the whole meristem (Figure 5.7). *FD*, the interaction partner of FT, is expressed in the inflorescence meristem, but not in the primordia (Figure 5.9). The pattern of *SOC1* is similar to that of FD (Figure 5.11), but those of *LFY* and *AP1* are opposite (Figure 5.13 and Figure 5.15). *TFL1* is only expressed in the inner zone of the meristem (Figure 5.17), but its protein diffuses to a larger zone and forms a concentration gradient in the SAM (Figure 5.19). Auxin forms local concentration maxima, which match the expression pattern of *LFY* and *AP1*, except for the centremost

maximum (Figure 5.21). *ARF* is expressed in all of the L1, except the apex (Figure 5.23).

Overall, the simulated patterns match those observed in the published ISH experiments.

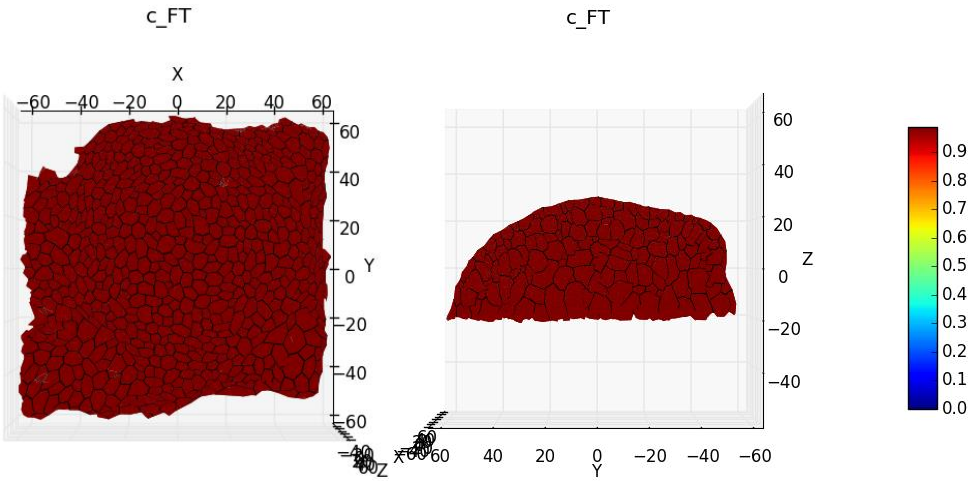


Figure 5.7. FT protein concentrations at t=300 (top view and cross-section).

FT is distributed uniformly across the whole SAM.

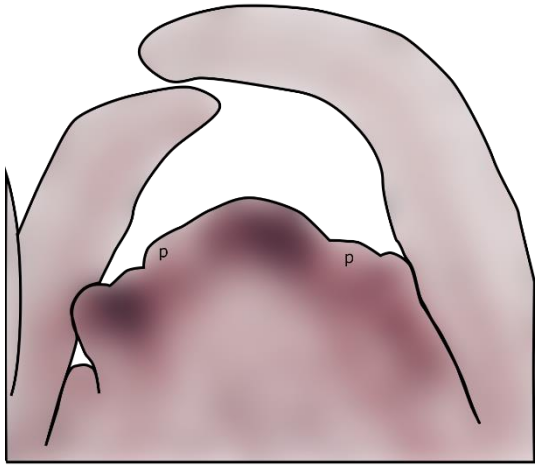


Figure 5.8. Expression of *FD* in *A. thaliana* inflorescence SAM (after Wigge et al., 2005). Colours were derived from the original figure and indicate expression intensity (darker is more intense). "p" denotes primordia.

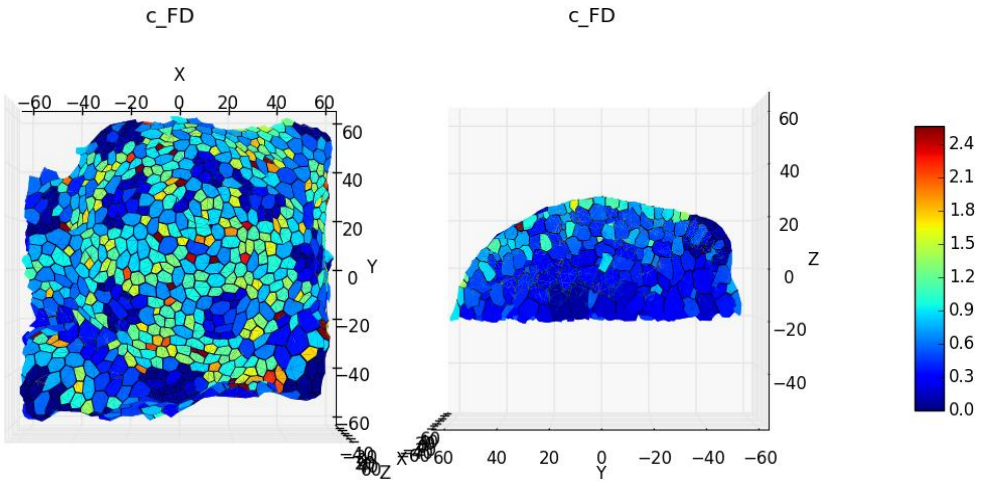


Figure 5.9. *FD* mRNA concentrations at t=300 (top view and cross-section). *FD* is expressed in the inflorescence meristem, but not in the primordia.

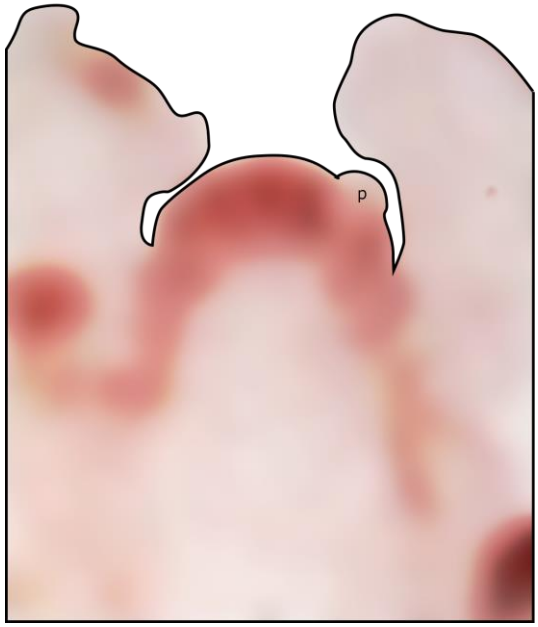


Figure 5.10. Expression of *SOC1* in *A. thaliana* inflorescence SAM (after Wang et al., 2009). Colours were derived from the original figure and indicate expression intensity (darker is more intense). "p" denotes primordia.

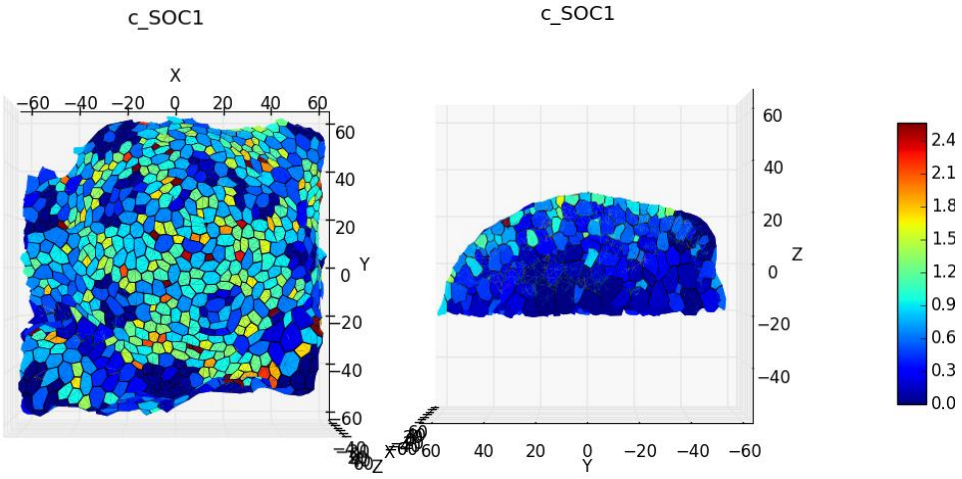


Figure 5.11. *SOC1* mRNA concentrations at t=300 (top view and cross-section). *SOC1* is expressed in the inflorescence meristem, but not in primordia.

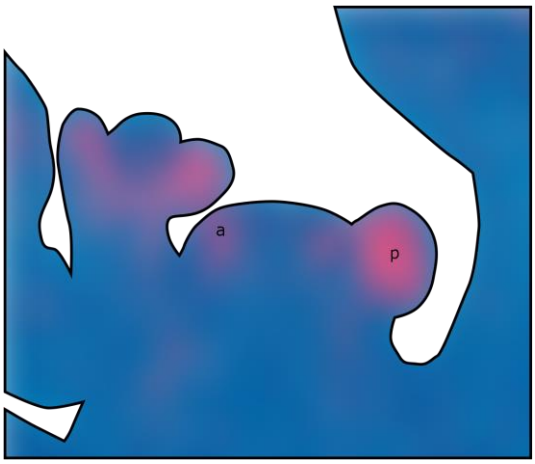


Figure 5.12. Expression of *LFY* in *A. thaliana* inflorescence SAM (after Blazquez et al., 1997). Colours were derived from the original figure and indicate expression intensity (expression appears pink, background is blue). “a” and “p” denote anlagen and primordia, respectively.

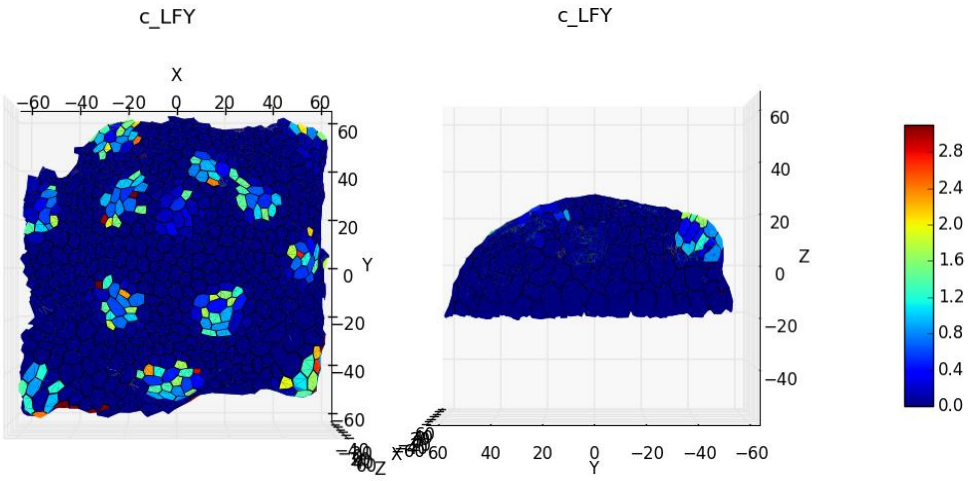


Figure 5.13. *LFY* mRNA concentrations at t=300 (top view and cross-section). *LFY* is expressed in primordia.

Figure 5.14. Expression of *AP1* in *A. thaliana* inflorescence SAM (after Wigge et al., 2005). Colours were derived from the original figure and indicate expression intensity (darker is more intense). "p" denotes primordia.

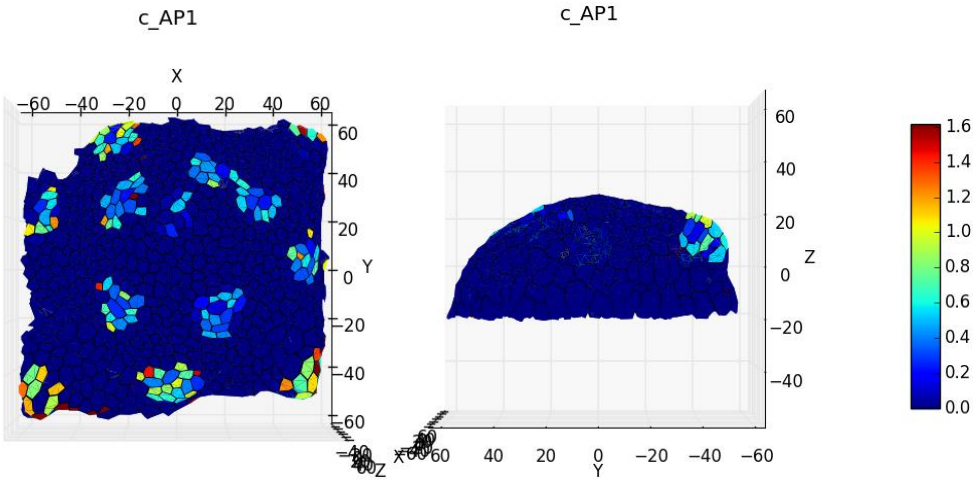


Figure 5.15. *AP1* mRNA concentrations at t=300 (top view and cross-section). *AP1* is expressed in primordia. Older primordia exhibit higher concentrations.

Figure 5.16. Expression of *TFL1* in cross-section of *A. thaliana* inflorescence SAM (after Liu et al., 2013). Colours were derived from the original figure and indicate expression intensity (darker is more intense). "p" denotes primordia.

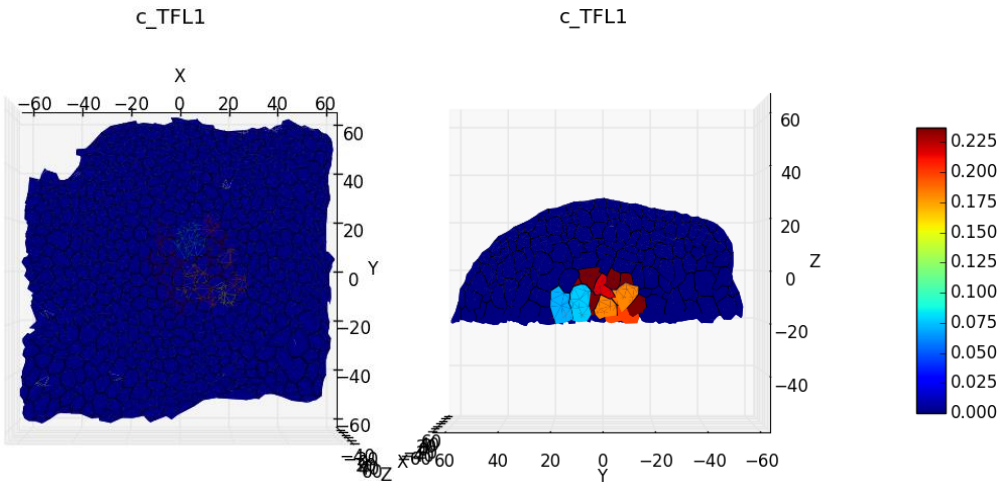


Figure 5.17. *TFL1* mRNA concentrations at t=300 (top view and cross-section). *TFL1* is transcripted in the inner part of the SAM. No *TFL1* expression is visible from the top.

Figure 5.18. Distribution of TFL1 protein in cross-section of *A. thaliana* inflorescence SAM (after Conti and Bradley, 2007). Colours were derived from the original figure and indicate protein presence (proteins are black, background is blue). "p" denotes primordia.

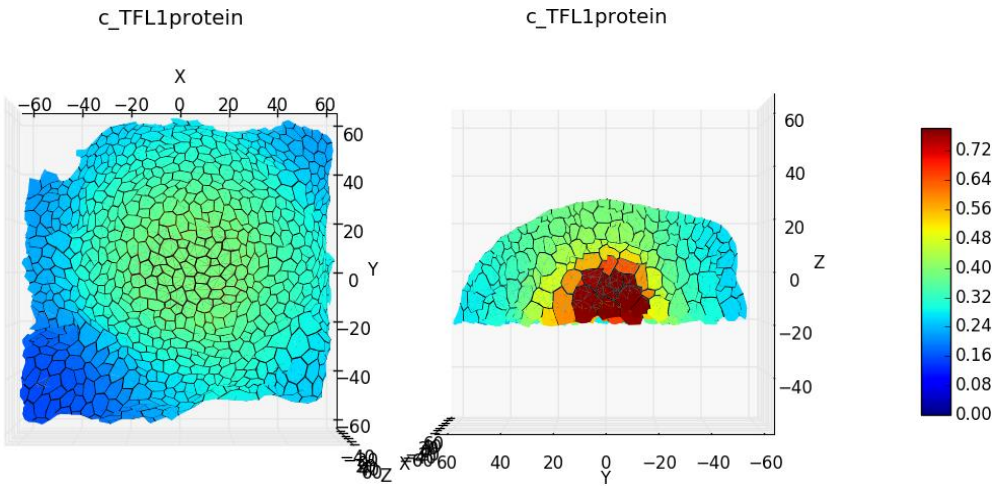


Figure 5.19. TFL1 protein concentrations at t=300 (top view and cross-section). The TFL1 protein forms a concentration gradient around its expression domain.

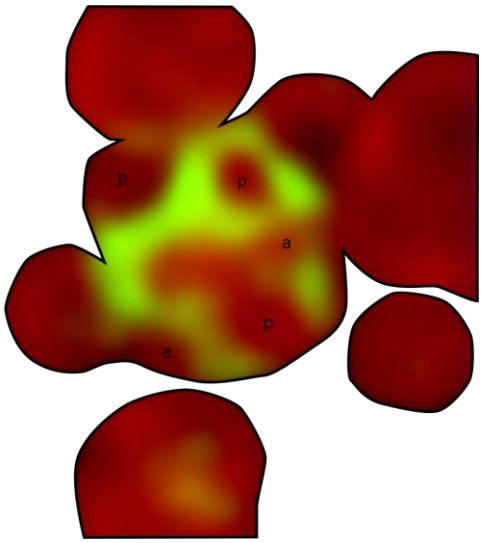


Figure 5.20. Distribution of auxin in top-view of *A. thaliana* inflorescence SAM (after Vernoux et al., 2011). Colours were derived from the original figure and indicate auxin presence (auxin is red, background is green). "p" denotes primordia.

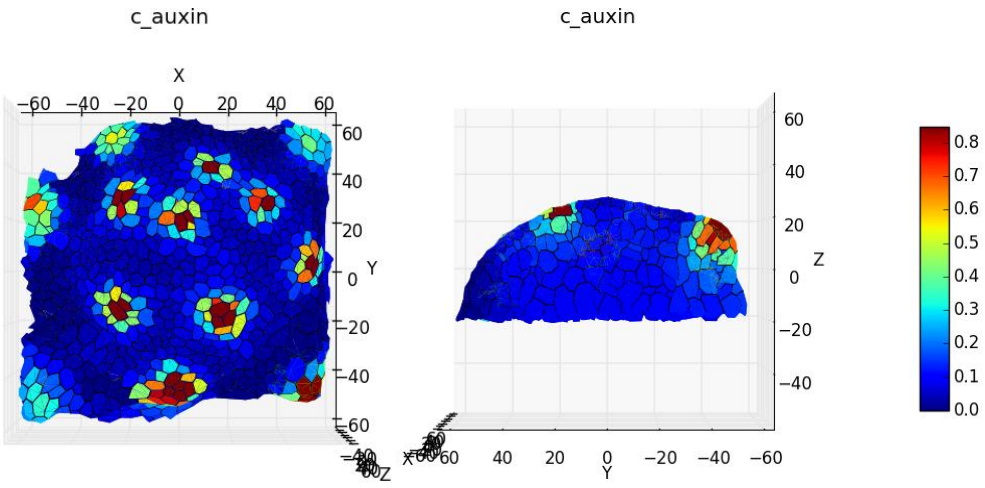


Figure 5.21. Auxin concentrations at t=300 (top view and cross-section). Auxin forms local maxima as a result of polar transport.

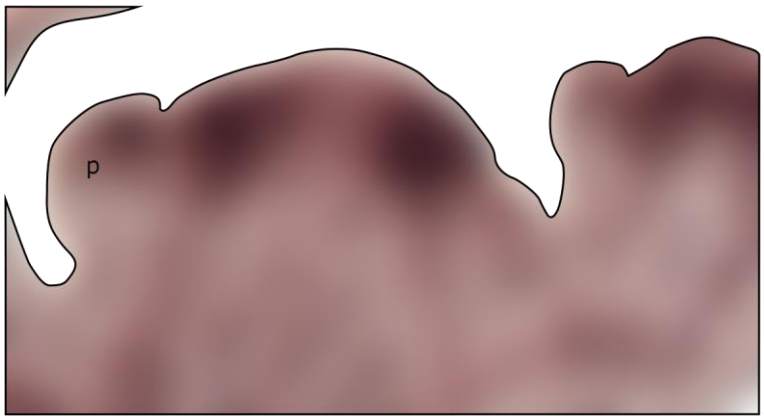


Figure 5.22. Expression of *ARF5* in cross-section of *A. thaliana* inflorescence SAM (after Vernoux et al., 2011). Colours were derived from the original figure and indicate expression intensity (darker is more intense). "p" denotes primordia.

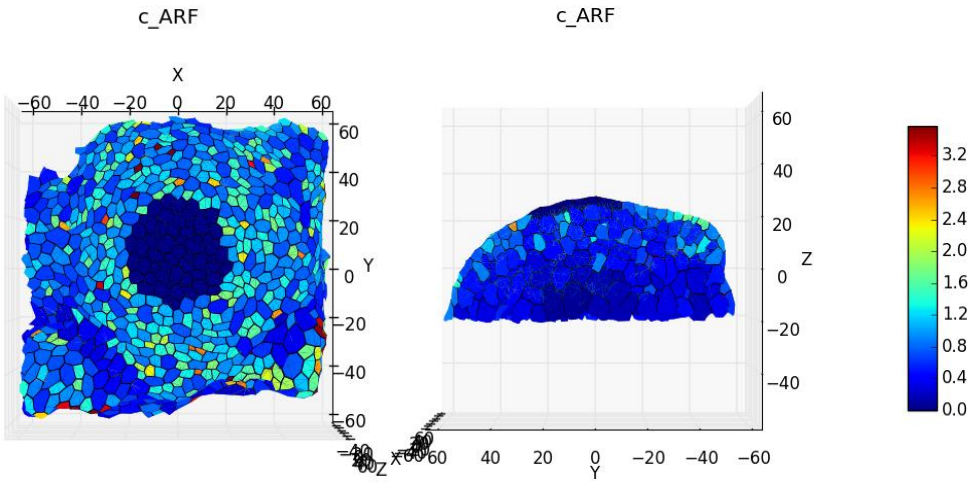


Figure 5.23. *ARF* mRNA concentrations at t=300 (top view and cross-section). *ARF* are not expressed at the apex of the SAM.

5.3.2. In the absence of FT induction, inflorescence and floral identity genes are not expressed

When FT is not induced (FT=0 for the whole simulation), inflorescence and floral identity genes are not expressed, resulting in expression profiles reminiscent of a vegetative SAM (Figure 5.24, and Figure 5.26 to Figure 5.28). Due to the non-expression of *AP1*, FD does not get repressed in primordia and is therefore expressed throughout the SAM (Figure 5.25). Other species (*TFL1*, TFL1 protein, auxin and *ARF*) were not affected.

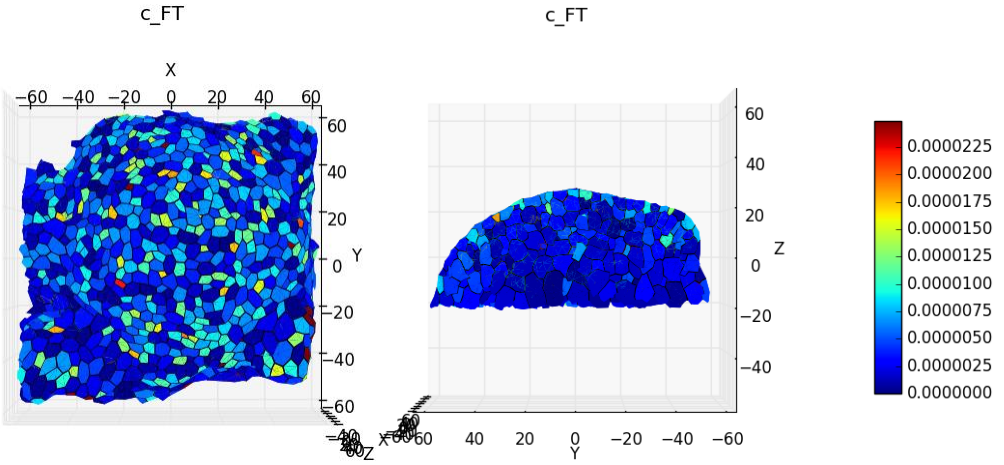


Figure 5.24. FT protein concentrations at t=300 without induction of FT (top view and cross-section). FT is not present in the SAM, except for the initialization noise.

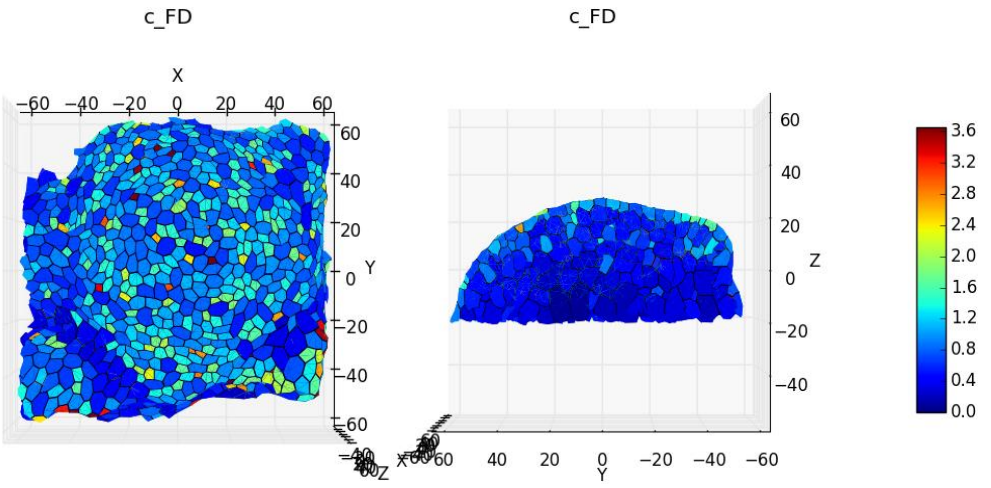


Figure 5.25. *FD* mRNA concentrations at t=300 without induction of FT (top view and cross-section). FD is expressed throughout the meristem, even in primordia.

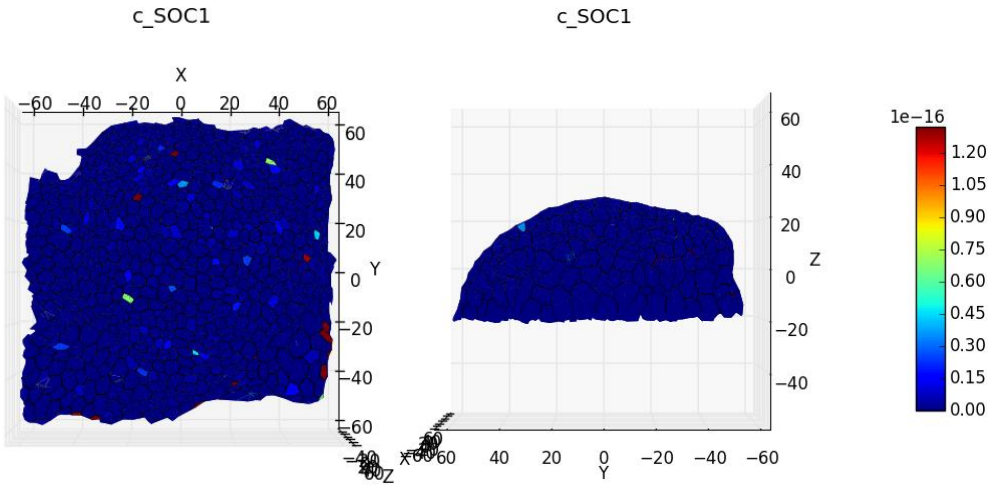


Figure 5.26. *SOC1* mRNA concentrations at t=300 without induction of FT (top view and cross-section). SOC1 is not expressed in the SAM.

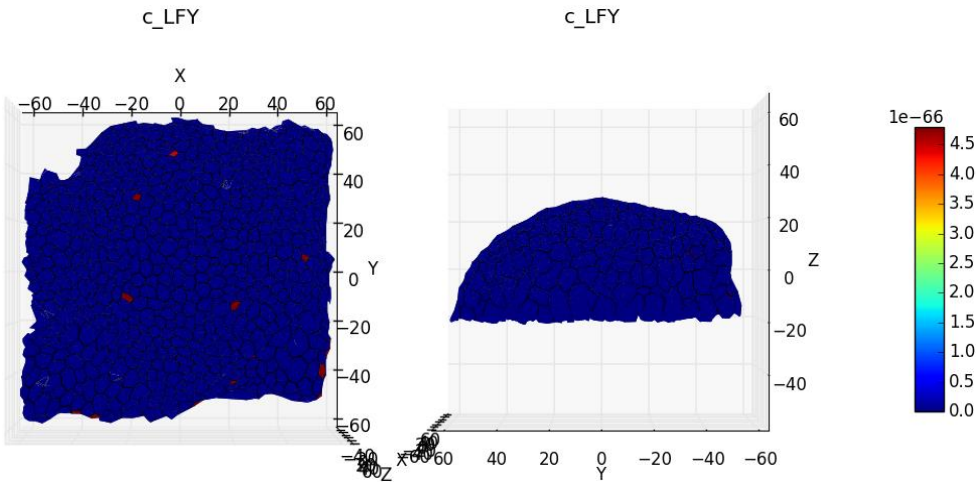


Figure 5.27. *LFY* mRNA concentrations at t=300 without induction of FT (top view and cross-section). LFY is not expressed in the SAM.

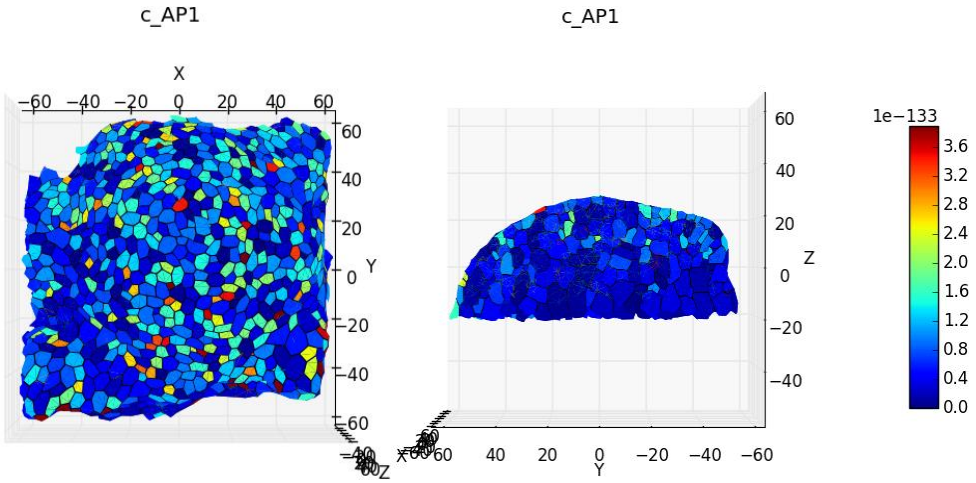


Figure 5.28. *AP1* mRNA concentrations at t=300 without induction of FT (top view and cross-section). AP1 is not expressed in the SAM.

5.3.3. Induction of FT in *ap1* and *tfl1* mutants results in expression patterns compatible with their respective inflorescence architectures

Simulations were also carried out for the *ap1* and *tfl1* mutants, which had also been studied in Chapter 4, by keeping the initial values of the relevant variables

close to 0 and setting the associated ODEs to always return 0. In the *ap1* mutant, *SOC1* and *FD* patterns do not show any holes where the primordia should be (Figure 5.29 and Figure 5.30). This can potentially explain why the *ap1/cal* mutant exhibits a recursive, cauliflower-like inflorescence structure, where inflorescence meristems generate additional inflorescence meristems (Smyth, 1995). In the *tfl1* mutant, *AP1* expression is observed in the apex zone (Figure 5.31). This is consistent with the fact that the inflorescence of the *tfl1* mutant ends with a flower (Shannon and Meeks-Wagner, 1991).

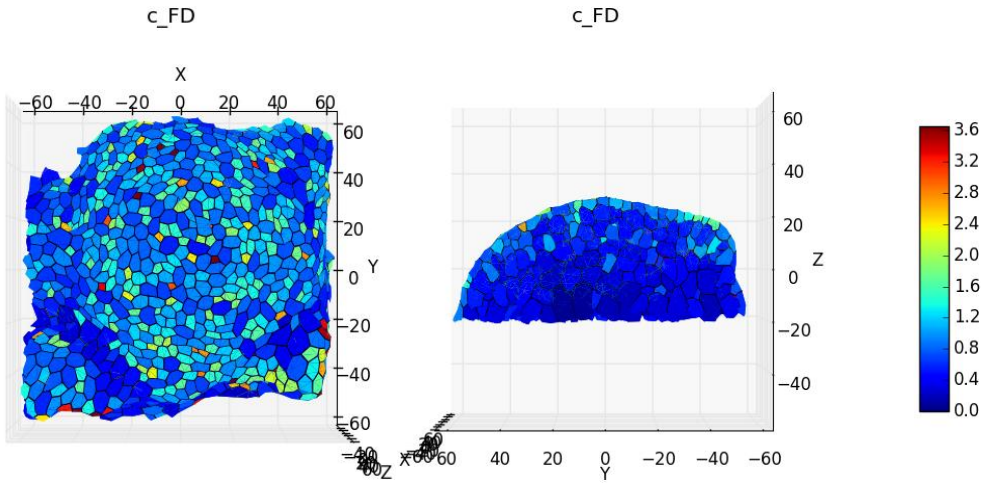


Figure 5.29. *FD* mRNA concentrations at t=300 in the *ap1* mutant. There are no holes corresponding to the primordia in the pattern.

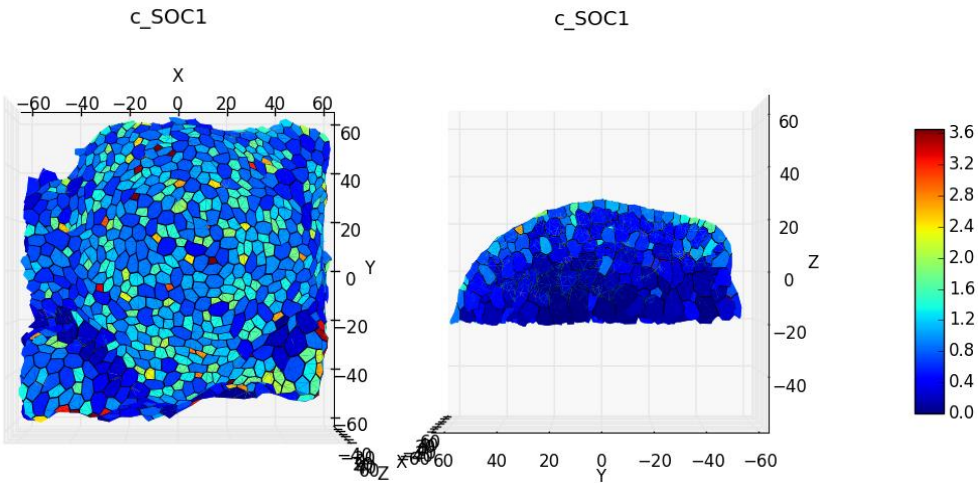


Figure 5.30. *SOC1* mRNA concentrations at t=300 in the *ap1* mutant (top view and cross-section). There are no holes corresponding to the primordia in the pattern.

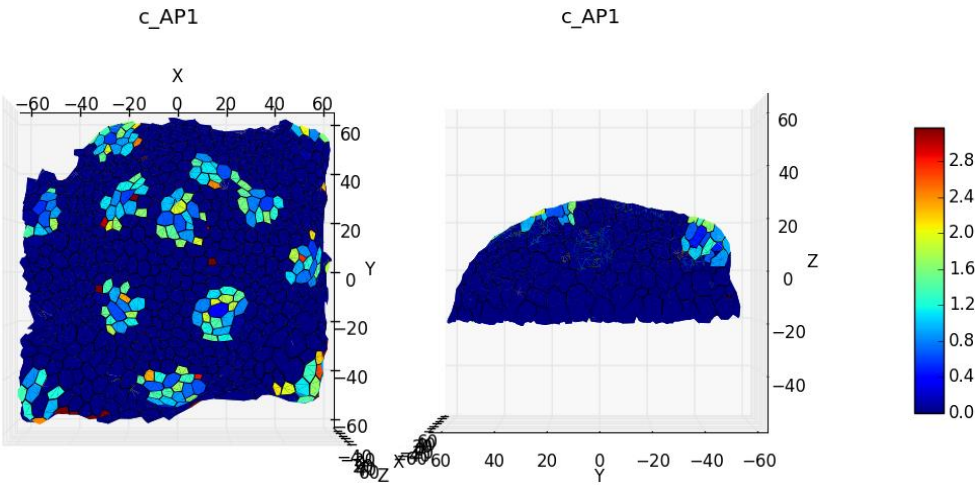


Figure 5.31. *AP1* mRNA concentrations at t=300 in the *tfl1* mutant (top view and cross-section). *AP1* gets expressed in the apex zone.

5.4. DISCUSSION

The results of this chapter have shown that the regulatory network proposed in Chapter 4 is not only viable as a Boolean model where space is abstracted

as a set of compartments differing by the values of their input variables, but also as a more realistic ODE model implemented in a 3D tissue structure, where the intercellular transport of species is modelled explicitly. This can be seen in the patterns generated by the ODE model.

5.4.1. An ODE model derived from a Boolean model generated by genetic programming is able to replicate the patterning of the SAM

The simulated patterns match the ISH observations reviewed in the previous chapter (S4.1 Table). In the WT, they are a result of the interplay between the mobile species (FT, auxin and the TFL1 protein) and the species whose patterns are affected by the predefined inner and apex zones. FT triggers the expression of inflorescence genes (in this model, *SOC1*). Where *SOC1* is accompanied by auxin and *ARF* (i.e. in the primordia), *LFY* also gets activated. In the apex zone however, auxin is present but *ARF* is absent, therefore *LFY* is not strongly expressed (Figure 5.13). If the concentration in TFL1 protein is low enough, the positive feedback loop between *LFY* and *AP1* gets activated, which in turns activates the negative feedback of *AP1* on *FD* and *SOC1*. Figure 5.15 shows that *AP1* is expressed in the same cells as *LFY*, but the intensity of AP1 expression is higher in the older meristems. This can be attributed to the fact that the TFL1 protein – an inhibitor of *AP1* – is present in higher concentrations in the younger primordia, due to the concentration gradient (Figure 5.19).

In the simulations of the *ap1* mutant, *AP1* does not get expressed in any of the primordia, preventing them from downregulating *FD* (Figure 5.29) and *SOC1*

(Figure 5.30). The resulting expression profile of primordia is therefore similar to that of the inflorescence meristem, which could explain how the recursive structure of the cauliflower-like *ap1/cal* mutant (Kempin et al., 1995) establishes.

In the *tfl1* mutant, the auxin maximum in the apex zone is no longer repressed by the TFL1 protein, causing it to start expressing *AP1* (Figure 5.31). This might explain why the inflorescence of the *tfl1* mutant loses its indeterminate trait and turns into a flower (Liljegren et al., 1999). Surprisingly, however, *AP1* expression in the apex zone is conditioned by the presence of auxin, but is not affected by the absence of *ARF*. Additional experiments would be required to determine whether this is biologically relevant or not, but the regulation of *ARF* is known to require additional information, as its proposed inhibitor is the *TFL1* mRNA, not the TFL1 protein, which would be biologically surprising.

5.4.2. Updating quantities of matter in the ODEs result in noisy patterns

The simulated concentrations of cell autonomous species are very noisy. This is a consequence of the assumptions made regarding the synthesis and the transport of the modelled chemical species. Synthesis is assumed to result in the same amount of molecules regardless of the size of the cell where it is taking place. Therefore, smaller cells will develop stronger concentrations if all other factors are equal.

Other assumptions could be made. First, synthesis could be proportional to cell volume. This is commonly seen in non-spatialized models where $\frac{dc_i}{dt} =$

$f(c_1, \ldots, c_n)$ (as opposed to $\frac{di}{dt} = f(c_1, \ldots, c_n)$ in this model). While the former is simpler, there does not seem to be any biological principle supporting an effect of cell size on synthesis rate. Second, the dilution volume used in the calculation of concentrations may not be the volume of the whole cell, but the volume of one or several organelles, such as the cytoplasm. If the modelled chemical species are contained within the cytoplasm, it could then be safe to consider that all cells have an equal "volume", as far as the calculation of concentrations is concerned. Indeed, meristematic cells have no vacuoles (only prevacuoles), but in mature plant cells, up to 90% of cell volume is taken up by the vacuole (Wink, 1993). Therefore, the volume of the cytoplasm might be less variable, or even constant. However, the main drawback of using a constant volume for all cells comes from the artefacts it generates when combined with intercellular passive transport and the hypothesis of instant diffusion within cells. It results in mobile species moving abnormally fast through large cells, because diffusion from one side of the cell to the other is instant, and this is not compensated by the inertia to changes in concentrations that a larger cell volume would normally afford.

To test whether the first hypothesis was plausible, a model updating concentrations instead of quantities of matter was implemented. Equations are given in Equation 5.2 and parameters are still the same as in the previous model (Table 5.3). In practice, concentration updating was not found to affect simulation results much (Figure 5.32 to Figure 5.40), although it does reduce the variability of concentrations across the SAM. The small extent of the

changes is probably due to the fact that the disparities in cell sizes are not too pronounced in this tissue structure, although SOC1 concentration is noticeably higher in the inner cells of the SAM with the concentration-updating model (Figure 5.11 and Figure 5.34).

Equation 5.2. Variant of the ODE model with concentration updating.

$$\frac{dc_{FT_i}}{dt} = \frac{t^n}{t^n + \tau^n} - c_{FT_i}$$

$$\frac{dc_{FD_i}}{dt} = \frac{\theta^n}{\theta^n + c_{AP1_i}^n} - c_{FD_i}$$

$$\frac{dc_{SOC1_i}}{dt} = \frac{c_{SOC1_i}^n + c_{FT_i}^n}{\theta^n + c_{SOC1_i}^n + c_{FT_i}^n} \frac{c_{FD_i}^n}{\theta^n + c_{FD_i}^n} - c_{SOC1_i}$$

$$\frac{dc_{LFY_i}}{dt} = \frac{c_{AP1_i}^n + c_{auxin_i}^n . c_{SOC1_i}^n}{\theta^n + c_{AP1_i}^n + c_{auxin_i}^n . c_{SOC1_i}^n} \frac{\theta^n + c_{ARF_i}^n}{\theta^n + c_{ARF_i}^n + c_{TFL1protein_i}^n} - c_{LFY_i}$$

$$\frac{dc_{AP1_i}}{dt} = \frac{c_{LFY_i}^n}{\theta^n + c_{LFY_i}^n} \frac{\theta^n}{\theta^n + c_{TFL1protein_i}^n} - c_{AP1_i}$$

$$\frac{dc_{TFL1_i}}{dt} = 0.175 \frac{\theta^n}{\theta^n + c_{AP1_i}^n} I_{inner_i} - c_{TFL1_i}$$

$$\frac{dc_{TFL1protein_i}}{dt} = 100 \frac{c_{TFL1_i}}{c_{TFL1_i} + \theta^n} - c_{TFL1protein_i} + D_{TFL1protein} \sum_{j \in neighbours(i)} S_{ij} \left(c_{TFL1protein_j} - c_{TFL1protein_i} \right)$$

$$\frac{dc_{auxin_i}}{dt}$$

$$= I_{L1_i} - c_{auxin_i}$$

$$+ \frac{1}{V_i}\Bigg(0.1\, D_{auxin} \sum_{j \in neighbours(i)} S_{ij} \left(c_{auxin_j} - c_{auxin_i} \right)$$

$$+ D_{auxin} \sum_{j \in neighbors(i) \cap L1} S_{ij} \left(c_{auxin_j} - c_{auxin_i} \right)$$

$$+ T_{auxin} \sum_{j \in neighbors(i) \cap L1} \Bigg(c_{auxin_j} \frac{c_{auxin_i} S_{ij}}{\sum_{k \in neighbours(j) \cap L1} c_{auxin_k} S_{jk}}$$

$$- c_{auxin_i} \frac{c_{auxin_j} S_{ij}}{\sum_{k \in neighbours(i) \cap L1} c_{auxin_k} S_{ik}} \Bigg) \Bigg)$$

$$\frac{dc_{ARF_i}}{dt} = \left(1 - I_{apex_i}\right) \frac{\theta^n}{\theta^n + c_{TFL1_i}^n} - c_{ARF_i}$$

- c_{FT_i}, c_{FD_i}, c_{SOC1_i}, c_{LFY_i}, c_{AP1_i}, c_{TFL1_i}, $c_{TFL1protein_i}$, c_{auxin_i}, c_{ARF_i}: concentrations in *FT* mRNA, *FD* mRNA, *SOC1* mRNA, *LFY* mRNA, *AP1* mRNA, *TFL1* mRNA, TFL1 protein, auxin and *ARF* mRNA, respectively, in cell i (arb. unit).
- θ: threshold concentration used in the Hill functions regulating all species (arb. unit).
- n: Hill coefficient used in the Hill functions regulating all species (dimensionless).
- V_i: volume of cell i (arb. unit)
- τ: time of the floral transition (arb. unit)

- D_{auxin} , $D_{TFL1protein}$: diffusion coefficients of auxin and the TFL1 protein, respectively (arb. unit)
- T_{auxin}: active transport coefficient of auxin (arb. unit)
- I_{L1_i}, I_{inner_i}, I_{apex_i}: indicator variables that are 1 if cell i is part of the L1 layer, the “inner” zone, the “apex” zone, respectively, and 0 otherwise.
- $neighbours(i)$: set of cells adjacent to cell i
- $L1$: set of cells belonging to the L1 layer

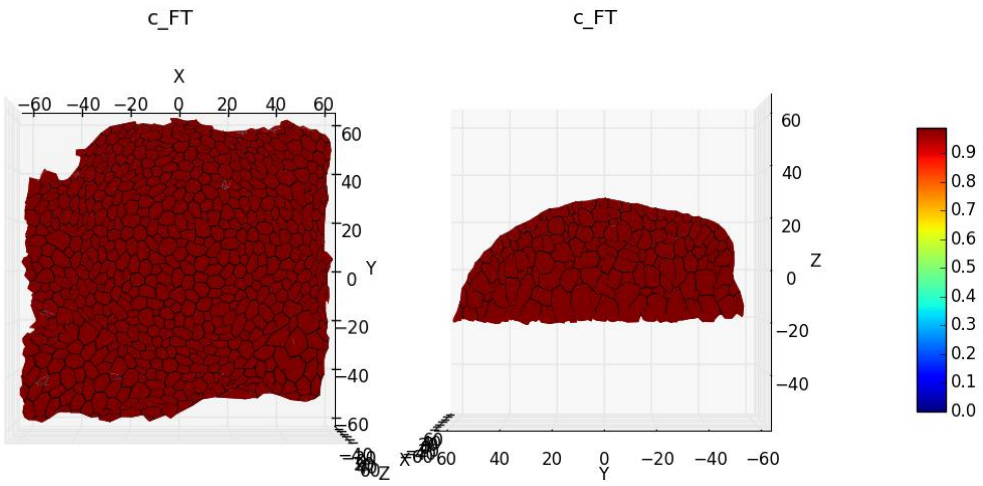


Figure 5.32. FT protein concentrations at t=300 (concentration-updating model, top view and cross-section). FT is present across the whole SAM.

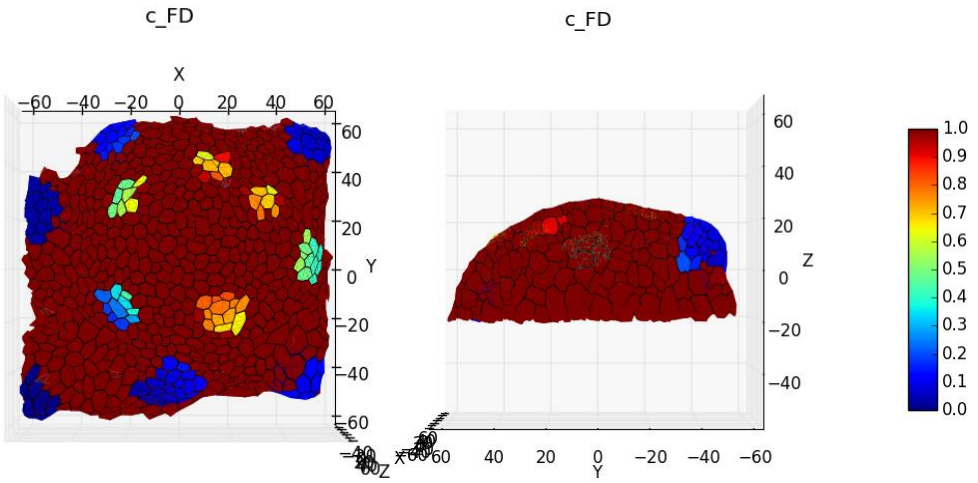


Figure 5.33. *FD* mRNA concentrations at t=300 (concentration-updating model, top view and cross-section). *FD* is expressed throughout the inflorescence meristem. It is not expressed in older primordia, but it is at a lower level in anlagen.

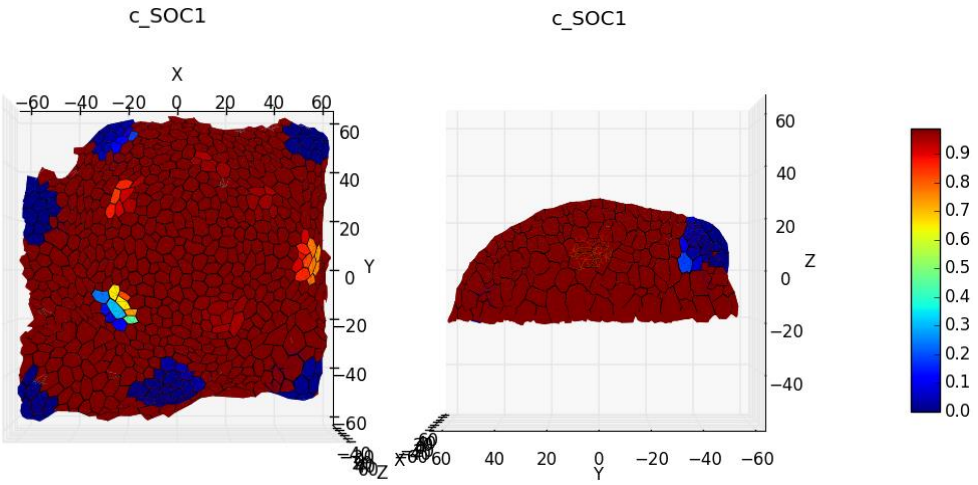


Figure 5.34. *SOC1* mRNA concentrations at t=300 (concentration-updating model, top view and cross-section). *SOC1* is expressed in the inflorescence meristem and in anlagen, but not in primordia.

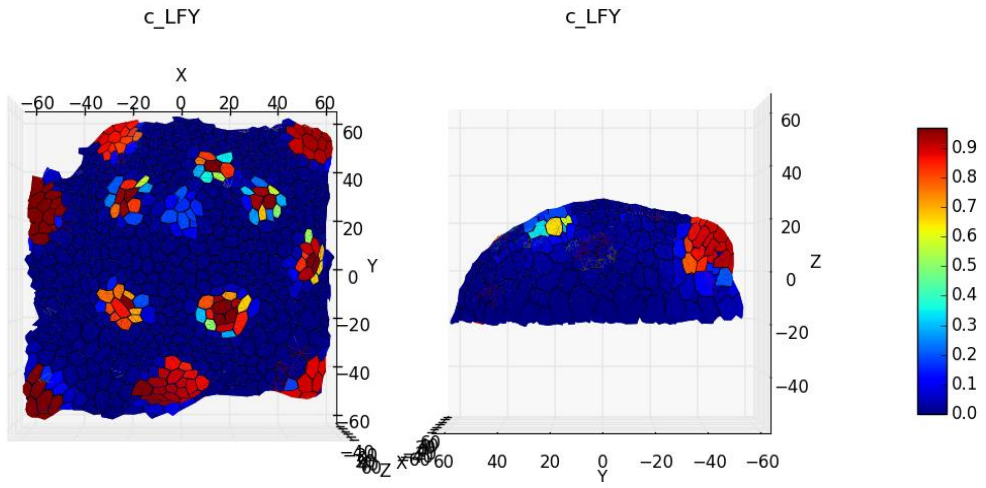


Figure 5.35. *LFY* mRNA concentrations at t=300 (concentration-updating model, top view and cross-section). *LFY* is expressed in anlagen and primordia.

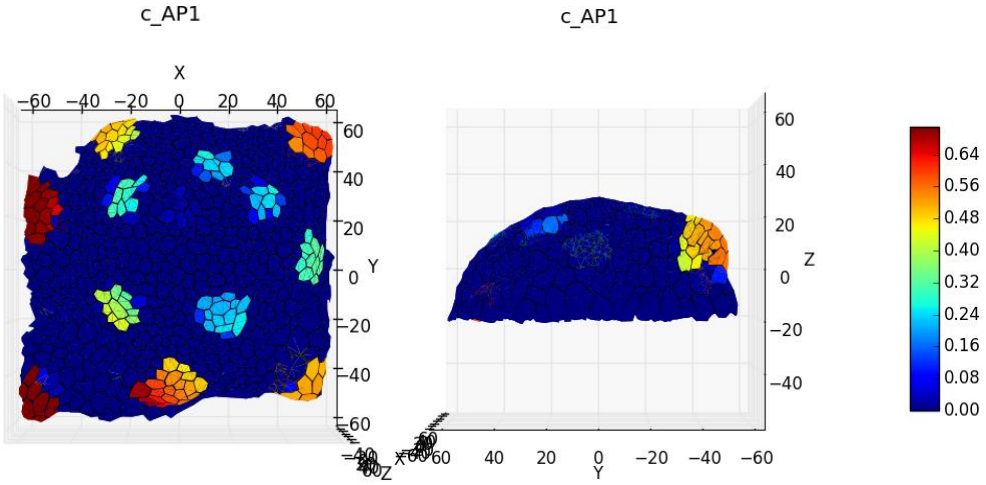


Figure 5.36. *AP1* mRNA concentrations at t=300 (concentration-updating model, top view and cross-section). AP1 is expressed in primordia, and – to a lesser extent – in anlagen.

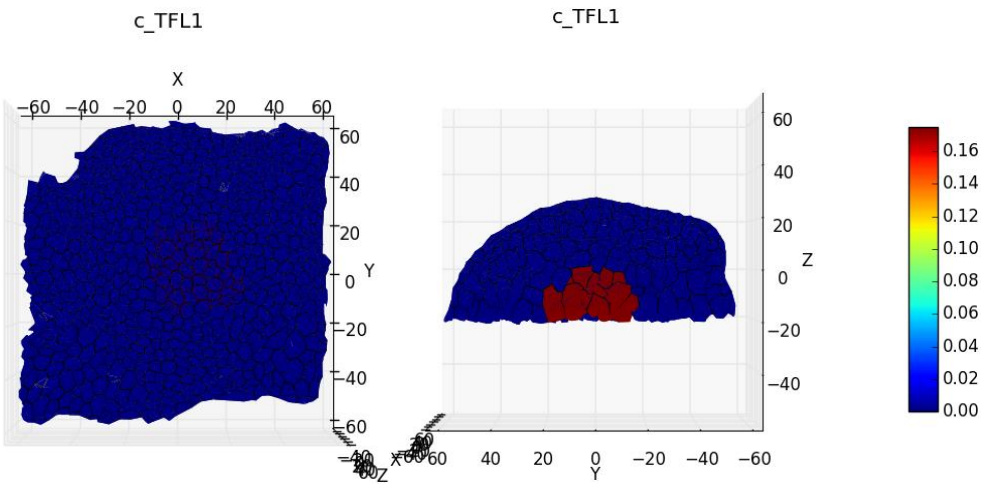


Figure 5.37. *TFL1* mRNA concentrations at t=300 (concentration-updating model, top view and cross-section). *TFL1* is expressed in the inner cells of the SAM.

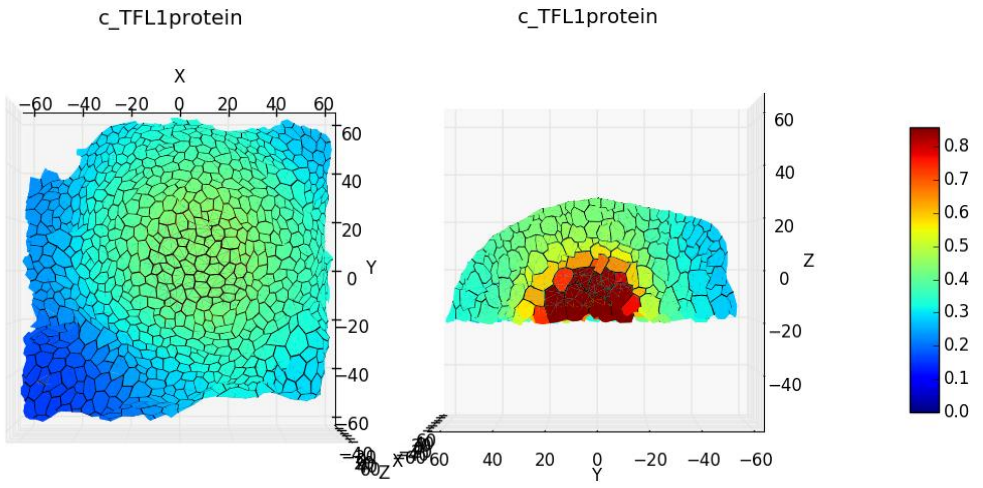


Figure 5.38. TFL1 protein concentrations at t=300 (concentration-updating model, top view and cross-section). TFL1 diffuses outward from its synthesis zone, resulting in a concentration gradient in the SAM.

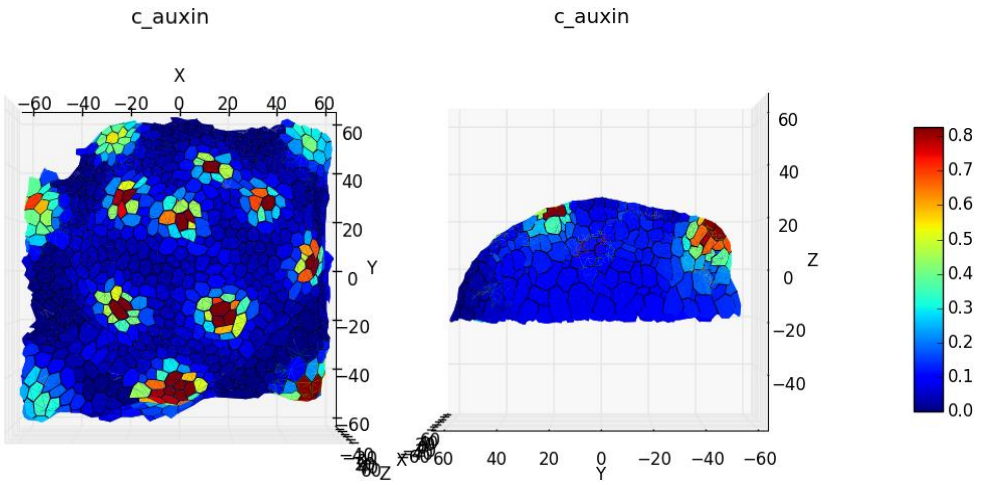


Figure 5.39. Auxin concentrations at t=300 (concentration-updating model, top view and cross-section). Auxin is actively transported in polar fashion, resulting in the formation of islands of higher concentrations, corresponding to anlagen and primordia.

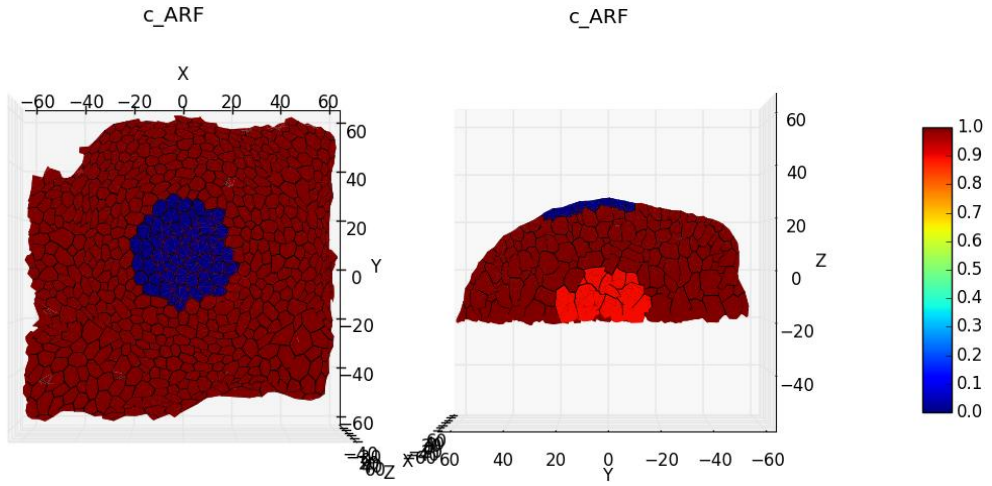


Figure 5.40. *ARF* mRNA concentrations at t=300 (concentration-updating model, top view and cross-section). *ARF* is expressed throughout the SAM, except in the central zone of the L1 layer. It is however partially repressed in the inner cells of the SAM.

5.4.3. Growth could be integrated into the model

The ultimate goal of this 4D model of the floral transition was to include growth, either kinematically, as a pre-recorded evolution of the geometry of the tissue, or dynamically, in response to stimuli like auxin. The dynamic approach was pursued using the Sofa modelling framework (Allard et al., 2007), already used by partners from Inria Montpellier to model sepal formation during early flower development (Boudon et al., 2015). One of the main benefits of modelling growth dynamically is that it might allow the simulation of some mutant phenotypes, such as the fractal, cauliflower-like inflorescence of the *ap1/cal* mutant (Kempin et al., 1995). However, modelling the floral transition proved more challenging than sepal formation, as it requires cell divisions to prevent growing cells from getting too large. Oversized cells lead to issues several natures: geometrical (plausibility of the simulated tissue), chemical (over-dilution of modelled species) and mechanical (larger cells are weaker than smaller cells).

5.5. CONCLUSIONS

A Boolean model generated in the previous chapter could easily be translated into an ODE model using Hill and Shea-Ackers functions, due to the simplicity of the models generated by genetic programming. The ODE model was implemented in a 3D structure of SAM using the Multicell framework, developed during this PhD project. It was able to reproduce the establishment of the patterns observed in the SAM by ISH with quantitative – though relative

– predictions. This confirmed that the regulatory network proposed in Chapter 4 was plausible not only in a simple Boolean modelling framework, but also in more realistic 3D, ODE modelling framework.

6. GENERAL DISCUSSION

This thesis has explored multiple approaches to model the floral transition and understand it in a systemic way. Those approaches fall into two categories: those based on quantitative time series of gene expression and those based on qualitative data. It stands to reason that, all other things kept equal, quantitative data should be superior to qualitative data. However, this work has shown that, in practice, qualitative data may be more appropriate, because of other characteristics.

6.1. GENERATING SUITABLE QUANTITATIVE DATA IS DIFFICULT

Chapters 2 and 3 have both shown that time series of gene expressions are not necessarily sufficient to allow for an accurate reconstruction of gene regulation dynamics on their own.

6.1.1. Gene regulatory events are not visible at the time resolution used by studies of the floral transition

Intuitively, one might think that time series contain the necessary information to infer gene regulations. For instance, if the expression level of a gene starts increasing and is promptly followed by another, it would be reasonable to hypothesize the first is an activator of the second. However, the time resolution used in floral transition studies is too low to make such observations. The delay between the activation of a regulator and that of its target is only a few hours (Rosenfeld and Alon, 2003). In the rice data presented in Chapter 2, the resolution was one measurement per 14 days, and

in the *A. thaliana* data of Chapter 3, it was one measurement per 1 or 3 days, depending on the data set. This is because floral experiments need to span over several weeks or months, and acquiring samples every hour would require unrealistic amounts of plants and labour.

However, even if time series of gene expression could be acquired at high temporal resolution, they might still present other issues.

6.1.2. Some phenomena are not observed in the data

ODE models often make use of saturating functions (Michaelis-Menten or Hill functions) (Hill, 1910; Michaelis et al., 2011; Alon, 2006). The intent is to acknowledge that there is an upper bound to the synthesis rate of any mRNA, no matter how much transcription factors stimulate it, dictated by physical limitations, such as the number of RNA polymerases that can fit on a gene and the transcription rate of a single polymerase. There is however no guarantee that transcription takes place anywhere near these bounds during an experiment, as shown in Chapters 2 and 3. When this is the case, the parameters of a Hill or Michaelis-Menten function cannot be estimated.

In Chapters 2 and 3, this problem was circumvented by using linear and polynomial (i.e. non-saturating) functions. Although they may not be mechanistically accurate, they are suitable approximations of the pre-plateau sections of Michaelis-Menten and Hill curves.

6.1.3. Some phenomena are not identifiable

Even when the modelled phenomena do take place in the experiments that generated the data, it may not be possible to separate their contributions to the observations. As seen in Chapters 2 and 3, this is the case of synthesis and degradation. These phenomena both contribute to the variation of observed expression levels. As a consequence, their parameters are strongly correlated when performing parameter estimation on gene expression time series. In practice, this means that changes in degradation rates can be compensated by changes in synthesis rates, with little consequences to the fit of the model to the data. In Chapter2, this problem was addressed by not modelling synthesis and degradation separately. Instead, the concentration of a gene's mRNA was simply represented as a function of the concentrations of its regulators' mRNAs. In Chapter 3, degradation was modelled separately from synthesis, but a single parameter was used for all genes, and information from another experiment (Narsai et al., 2007), where mRNA abundance was tracked in cells treated with transcription inhibitors, was introduced.

Modelling approaches relying on the available quantitative data were therefore not very successful in the case of the floral transition.

In the particular case of the determination of meristematic identity (Chapter 3), it is also likely due to the low spatial resolution of the measurement method (i.e. one measurement representative of the whole apex). Chapter 4 indeed demonstrated that data acquired at a higher spatial

resolution could be used successfully to infer gene regulation networks, despite limitations such as a low temporal resolution and the qualitative nature of the measurement method (ISH). Chapter 4 thereby also demonstrates that qualitative data are useful and should not be hastily dismissed.

6.2. QUALITATIVE DATA CAN BE USEFUL

Qualitative data, by definition, do not contain as much information as quantitative data. However, this can be compensated by other characteristics, such as the number of observations available, as qualitative data are often easier to acquire. This was illustrated by ISH data. Even though each ISH picture is only a snapshot of SAM development at a given time, a single picture actually contains observations about multiple parts of the SAM. From there, combining knowledge about the stages of SAM development (Traas and Vernoux, 2002; Carles and Fletcher, 2003), it is possible to piece together a rough timeline of gene regulatory events. However, the sequence of regulatory events is not the only valuable information. Their spatial location is important in itself, because the spatial organization of a SAM is not constant, but dynamic. Therefore, models of the development of the SAM should be able to predict the evolution of this spatial organization caused by the floral transition.

In the case of the floral transition, the spatial organization of the SAM is mediated by mobile factors (proteins and hormones). Through passive and active transport mechanisms, those factors establish patterns in the SAM, and

the ensuing combinations nudge the regulatory network of the floral transition towards diverse steady states, leading to the expression patterns observed in ISH pictures.

6.3. ONE MODEL OR MANY MODELS?

It is natural to aim at finding the model that would match reality perfectly, however this goal may be unattainable, due to incomplete data and the limitations of modelling formalisms. In that case, it makes sense to generate as many models as possible that fit the constraints set by all of the existing data. That approach is similar in spirit to Monte Carlo simulations (Harrison, 2010), as the underlying idea is to sample the variability of the solutions to a problem.

This is the idea behind the genetic programming approach of Chapter 4. It implies generating and testing large numbers of models, and was made possible due to the simplicity of Boolean models, as their simulation is computationally cheap, and they do not require any parameter optimization. Where just the parameter optimization of a single ODE model of Chapter 3 would take several days, testing one Boolean model only takes a fraction of a second. Therefore, it is possible to screen many models to find multiple suitable ones.

However, even among suitable models (i.e. models matching all observations), there may be some that are more plausible than others. Exhaustive search approaches, which iterate over all mathematically possible models, have

shown that they can result in models with very complex, unlikely-looking regulatory functions. Therefore, minimizing the complexity of regulatory functions appears desirable. This was easily implemented in genetic programming, because it works directly on equations. Independent runs of the algorithm resulted in multiple plausible solutions (Figure 4.8 and S4.2 Table), demonstrating that multiple regulatory scenarios are possible with the available data.

6.4. THE NETWORK OF THE FLORAL TRANSITION

Several topologies of the network of the floral transition have been proposed in Chapter 4. As new data become available in the future, this set of proposed networks might be narrowed down and provide a more precise picture of the real network. In the meantime, it is already possible to draw conclusions based on the common features of these networks. They for instance indicate that the switch behaviours of the floral transition network can be explained by several feedback loops, which are discussed below, in conjunction with other data sources.

6.4.1. The vegetative-inflorescence switch

The vegetative inflorescence switch needs to be activated by FT-FD dimers, but does not require them to remain activated (Adrian et al., 2009). The crucial role of FT is backed by the flowering time data used in Chapter 3. *ft* mutants are extremely late flowering (Figure 3.3), however the fact that they do eventually flower suggests that there are back-up activation pathways for

inflorescence genes. This role might be filled by the aging pathway and by *FT* homologs such as *TSF* (Yamaguchi et al., 2005), not present in that model. *fd* mutants are not as late flowering as *ft* mutants (Figure 3.3), even though the DNA-binding domains are not located on FT but on FD (Abe et al., 2005). This might be due to *FDP*, an *FD* paralog (Abe et al., 2005).

The vegetative inflorescence switch consists of a positive feedback loop including at least *SOC1,* and possibly *AGL24*. The involvement of *AGL24* is not supported by flowering time data, as the flowering time of the *agl24* mutant is not affected, compared to the WT (Figure 3.3).

6.4.2. The inflorescence-floral switch

According to Chapter 4, the inflorescence-floral switch consists of two loops: a positive feedback loop between *LFY* and *AP1*, and a negative feedback loop where *AP1* downregulates its activators, *FD* and *SOC1*. This negative feedback loop is consistent with results reported by Kaufmann and colleagues (Kaufmann et al., 2010), although they hypothesize that *AP1* directly down-regulates both of *FD* and *SOC1*. This would constitute an uncommon, coherent type 2 feedforward loop (Alon, 2006), but might result in faster down-regulation of *SOC1*. The Boolean models used in Chapter 4 were however not able to discern between these scenarios, as they do not represent time quantitatively.

According to the network reviewed by Liu and colleagues (Liu et al., 2009) and supported by the modelling work of Chapter 4, the *LFY-AP1* loop is activated

by a combination of *SOC1* and auxin signalling, while *AP1* expression is repressed by *TFL1*. The profiles resulting from the superposition of the patterns of auxin, its signalling partners (*ARF*) and TFL1 determine the locations of floral primordia. *AP1* activation has been confirmed to be *LFY*-dependent (Benlloch et al., 2011), but flowering time data seem to conflict with that observation, as the flowering time of *lfy* mutants is not affected (compared to the WT; **Error! Reference source not found.**). A possible explanation is that AP1 does not actually get activated in *lfy* mutants. This conflicts with the assumptions made in Chapter 3, but the fact that even the *ap1* mutant can flower (Irish and Sussex, 1990) indicates *AP1* is not actually required for flowering. Alternatively, if *AP1* does get activated in *lfy* mutants, then it would suggest there is probably a parallel activation pathway for AP1, as suggested by the fact that the *soc1* mutants are also only moderately late flowering. This second explanation corresponds to the ODE model of Chapter 3, which does not feature the *LFY-AP1* loop, but does include the controversial activation of *AP1* by FT that bypasses *SOC1* (Wigge et al., 2005; Benlloch et al., 2011).

6.5. FUTURE WORK

Answering the new questions raised by this work would require additional data. The problems with the currently available quantitative data and the benefits of spatial qualitative data have already been described previously. However, ideally, data should be simultaneously temporal, spatial and

quantitative. This can be done using several methods, such as the confocal imaging of fluorescent proteins, or single-cell RNA sequencing. These methods are at the time of writing still labour-intensive and expensive, but biological modelling would surely greatly benefit if they were somehow automated.

It would also be useful not to limit data generation to a single genetic background (typically the WT) and set of growing conditions, but to extend it to mutants, across a range of growing conditions. This would enable the construction of models that behave correctly in multiple conditions, and are therefore likely to be closer to the truth.

Finally, there needs to be more data about crop species for meaningful real-world applications. Currently, most of the available data comes from *A. thaliana*, which is a model plant in biology, but a common weed outside of the laboratory. *A. thaliana*, as a *Brassicacea*, is more closely related to the cabbage family than to most crops, therefore knowledge gained on its floral transition might transfer to crops such as cauliflower. However, it is quite distant from cereal species, which have rather different inflorescence architectures (Tanaka et al., 2013). Chapter 2 showed that data are scarce even for rice, which is one of the easiest cereal species to work with due to its relatively small genome (Jackson, 2016).

Aside from the generation of new data, the predictions of the 4D model in Chapter 5 show promise. The next logical step in the development of this model would be to integrate it with growth. Growth is a crucial phenomenon

in the establishment of the identity of lateral organs. It is not so much because of increases in the size of the SAM, as its domains are able to scale adaptively (Gruel et al., 2016), but because it is growth that drives the transitions of SAM cells between the identities described in Chapter 4 by pushing older cells away from the apex (or moving the apex away from older cells, depending on the frame of reference). The growth simulation method developed by Boudon and colleagues is a good candidate (Boudon et al., 2015), as Frédéric Boudon (Virtual Plants, Montpellier) has recently developed a way to include cell divisions in the simulations, thereby removing the previously mentioned roadblock.

In the longer term, Chapter 4 showed data from heterogeneous sources (ISH and gene interactions) could be successfully integrated to infer the dynamics of regulatory networks. There are however many other types of data available, such as qRT-PCRs (even if they only cover a few genes), transcriptomic and proteomic data, degradation rate experiments and confocal microscopy images. Specifying how they should all connect to each other is an ample task, but succeeding would probably result in an even better understanding of the floral transition. If not, it would at least have the benefit of formalising how raw data are interpreted, thereby making interpretation issues more traceable. A possible approach could be a two-level model, where the first level decomposes all types of data into collections of single-cell expression time-series, and the second level is a regulatory network model to be fitted on these time series. Depending on the flexibility required from the second-level

model and the availability of data, machine learning algorithms such as recurrent support vector machines (SVM) and recurrent neural networks (RNN) could be used (Graves, 2012; Schmidhuber et al., 2007).

6.6. GENERAL CONCLUSIONS

The goal of this thesis was to provide insight into the floral transition through mathematical modelling. In *A. thaliana*, multiple models were developed following different approaches. However, they are not in perfect agreement, as the various sources of data used do not seem entirely consistent. Some of it can be put down to suboptimal experimental designs (e.g. qRT-PCRs of tissues of varying compositions), but it is also possible that some are due to misinterpretations of experimental results (e.g. incorrect estimation of the importance of a regulatory interaction or improper analysis of ISH pictures). Additional data, preferably of better quality, would be useful in resolving those conflicts.

The modelling methods developed in this PhD project are most likely applicable to other cases, in particular the automatic generation of ISH-based Boolean models using genetic programming, and their translation into 4D ODE models.

SOURCES

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., Araki, T., 2005. FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex. Science 309, 1052–1056. doi:10.1126/science.1115983

Ackers, G.K., Johnson, A.D., Shea, M.A., 1982. Quantitative model for gene regulation by lambda phage repressor. Proc. Natl. Acad. Sci. 79, 1129–1133.

Adrian, J., Torti, S., Turck, F., 2009. From Decision to Commitment: The Molecular Memory of Flowering. Mol. Plant 2, 628–642. doi:10.1093/mp/ssp031

Akaike, H., 1973. Information theory and an extension to the maximum likelihood principle, in: B.N. Petrov, F. Csàki (Eds.), 2nd International Symposium on Information Theory. Akademiai Kiàdo, Budapest.

Allard, J., Cotin, S., Faure, F., Bensoussan, P.-J., Poyer, F., Duriez, C., Delingette, H., Grisoni, L., 2007. SOFA - an Open Source Framework for Medical Simulation. Presented at the MMVR 15 - Medicine Meets Virtual Reality, IOP Press, pp. 13–18.

Alon, U., 2006. An Introduction to Systems Biology: Design Principles of Biological Circuits. CRC Press.

Amasino, R., 2004. Vernalization, Competence, and the Epigenetic Memory of Winter. Plant Cell 16, 2553–2559. doi:10.1105/tpc.104.161070

Angel, A., Song, J., Dean, C., Howard, M., 2011. A Polycomb-based switch underlying quantitative epigenetic memory. Nature 476, 105–108. doi:10.1038/nature10241

Attwood, T., Campbell, P., Parish, H., Smith, A., Vella, F., Stirling, J., 2006. Oxford Dictionary of Biochemistry and Molecular Biology. OUP Oxford.

Ballerini, E.S., Kramer, E.M., 2011. In the Light of Evolution: A Reevaluation of Conservation in the CO-FT Regulon and Its Role in Photoperiodic Regulation of Flowering Time. Front. Plant Sci. 2. doi:10.3389/fpls.2011.00081

Band, L.R., Úbeda-Tomás, S., Dyson, R.J., Middleton, A.M., Hodgman, T.C., Owen, M.R., Jensen, O.E., Bennett, M.J., King, J.R., 2012. Growth-induced hormone dilution can explain the dynamics of plant root cell elongation. Proc. Natl. Acad. Sci. 109, 7577–7582. doi:10.1073/pnas.1113632109

Band, L.R., Wells, D.M., Fozard, J.A., Ghetiu, T., French, A.P., Pound, M.P., Wilson, M.H., Yu, L., Li, W., Hijazi, H.I., Oh, J., Pearce, S.P., Perez-Amador, M.A., Yun, J., Kramer, E., Alonso, J.M., Godin, C., Vernoux, T.,

Hodgman, T.C., Pridmore, T.P., Swarup, R., King, J.R., Bennett, M.J., 2014. Systems Analysis of Auxin Transport in the *Arabidopsis* Root Apex. Plant Cell Online 26, 862–875. doi:10.1105/tpc.113.119495

Bartoń, K., 2016. MuMIn: Multi-Model Inference.

Benlloch, R., Kim, M.C., Sayou, C., Thévenon, E., Parcy, F., Nilsson, O., 2011. Integrating long-day flowering signals: a LEAFY binding site is essential for proper photoperiodic activation of APETALA1. Plant J. 67, 1094–1102. doi:10.1111/j.1365-313X.2011.04660.x

Blazquez, M.A., Soowal, L.N., Lee, I., Weigel, D., 1997. LEAFY expression and flower initiation in Arabidopsis. Development 124, 3835–3844.

Bomblies, K., Wang, R.-L., Ambrose, B.A., Schmidt, R.J., Meeley, R.B., Doebley, J., 2003. Duplicate FLORICAULA/LEAFY homologs zfl1 and zfl2 control inflorescence architecture and flower patterning in maize. Development 130, 2385–2395. doi:10.1242/dev.00457

Boudon, F., Chopard, J., Ali, O., Gilles, B., Hamant, O., Boudaoud, A., Traas, J., Godin, C., 2015. A Computational Framework for 3D Mechanical Modeling of Plant Morphogenesis with Cellular Resolution. PLOS Comput Biol 11, e1003950. doi:10.1371/journal.pcbi.1003950

Brambilla, V., Fornara, F., 2013. Molecular Control of Flowering in Response to Day Length in Rice. J. Integr. Plant Biol. 55, 410–418. doi:10.1111/jipb.12033

Cara, A.D., Garg, A., Micheli, G.D., Xenarios, I., Mendoza, L., 2007. Dynamic simulation of regulatory networks using SQUAD. BMC Bioinformatics 8, 462. doi:10.1186/1471-2105-8-462

Carles, C.C., Fletcher, J.C., 2003. Shoot apical meristem maintenance: the art of a dynamic balance. Trends Plant Sci. 8, 394–401. doi:10.1016/S1360-1385(03)00164-X

Cavanaugh, J.E., 1997. Unifying the derivations for the Akaike and corrected Akaike information criteria. Stat. Probab. Lett. 33, 201–208. doi:10.1016/S0167-7152(96)00128-9

Chelliah, V., Juty, N., Ajmera, I., Ali, R., Dumousseau, M., Glont, M., Hucka, M., Jalowicki, G., Keating, S., Knight-Schrijver, V., Lloret-Villas, A., Natarajan, K.N., Pettit, J.-B., Rodriguez, N., Schubert, M., Wimalaratne, S.M., Zhao, Y., Hermjakob, H., Le Novère, N., Laibe, C., 2015. BioModels: ten-year anniversary. Nucleic Acids Res. 43, D542–D548. doi:10.1093/nar/gku1181

Cheong, R., Paliwal, S., Levchenko, A., 2010. Models at the Single Cell Level. Wiley Interdiscip. Rev. Syst. Biol. Med. 2, 34–48. doi:10.1002/wsbm.49

Chickarmane, V.S., Gordon, S.P., Tarr, P.T., Heisler, M.G., Meyerowitz, E.M., 2012. Cytokinin signaling as a positional cue for patterning the apical–

basal axis of the growing Arabidopsis shoot meristem. Proc. Natl. Acad. Sci. 109, 4002–4007. doi:10.1073/pnas.1200636109

Clark, S., 1997. Organ Formation at the Vegetative Shoot Meristem. Plant Cell 9, 1067–1076.

Cockram, J., Jones, H., Leigh, F.J., O'Sullivan, D., Powell, W., Laurie, D.A., Greenland, A.J., 2007. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. J. Exp. Bot. 58, 1231–1244. doi:10.1093/jxb/erm042

Cokelaer, T., Pradal, C., Fournier, C., 2009. Plant modelling with Python components in OpenAlea. Presented at the EuroSciPy 2009.

Colasanti, J., Coneva, V., 2009. Mechanisms of Floral Induction in Grasses: Something Borrowed, Something New. Plant Physiol. 149, 56–62. doi:10.1104/pp.108.130500

Conti, L., Bradley, D., 2007. TERMINAL FLOWER1 Is a Mobile Signal Controlling Arabidopsis Architecture. Plant Cell 19, 767–778. doi:10.1105/tpc.106.049767

de Jong, H., 2002. Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. J. Comput. Biol. 9, 67–103. doi:10.1089/10665270252833208

Dinh, J.-L., 2016. jldinh/multicell: Python library to run biological simulations of 3D, multicellular tissues. [WWW Document]. GitHub. URL https://github.com/jldinh/multicell/ (accessed 2.22.17).

Doi, K., Izawa, T., Fuse, T., Yamanouchi, U., Kubo, T., Shimatani, Z., Yano, M., Yoshimura, A., 2004. Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. Genes Dev. 18, 926–936. doi:10.1101/gad.1189604

Dong, Z., 2003. Incorporation of Genomic Information Into the Simulation of Flowering Time in Arabidopsis Thaliana. Kansas State University.

Dong, Z., Danilevskaya, O., Abadie, T., Messina, C., Coles, N., Cooper, M., 2012. A Gene Regulatory Network Model for Floral Transition of the Shoot Apex in Maize and Its Dynamic Modeling. PLoS ONE 7, e43450. doi:10.1371/journal.pone.0043450

Dufour-Kowalski, S., Pradal, C., Dones, N., Reuille, P.B. de, Boudon, F., Chopard, J., Silva, D.D., Durand, J.-B., Ferraro, P., Fournier, C., Guédon, Y., Ouangraoua, A., Smith, C., Stoma, S., Théveny, F., Sinoquet, H., Godin, C., 2007. OpenAlea: An open-source platform for the integration of heterogeneous FSPM components. Presented at the FSPM07 - 5th International Workshop on Functional-Structural Plant Models, p. P36:1-2.

Engelmann, K., Purugganan, M., 2006. The Molecular Evolutionary Ecology of Plant Development: Flowering Time in Arabidopsis thaliana, in: Research, B.-A. in B. (Ed.), Developmental Genetics of the Flower. Academic Press, pp. 507–526. doi:10.1016/S0065-2296(06)44013-1

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 Online 16, 2923–2939. doi:10.1105/tpc.104.021725

Fields, S., Song, O., 1989. A novel genetic system to detect protein-protein interactions. Nature 340, 245–246. doi:10.1038/340245a0

Fornara, F., de Montaigu, A., Coupland, G., 2010. SnapShot: Control of Flowering in Arabidopsis. Cell 141, 550–550.e2. doi:10.1016/j.cell.2010.04.024

Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A.G., Parizeau, M., Gagné, C., 2012. DEAP: Evolutionary Algorithms Made Easy. J Mach Learn Res 13, 2171–2175.

Fozard, J., 2015. odesparse Summary · CPIB / CVL Code Repository [WWW Document]. URL https://code.plant-images.org/odesparse (accessed 2.27.17).

Frugis, G., Giannino, D., Mele, G., Nicolodi, C., Chiappetta, A., Bitonti, M.B., Innocenti, A.M., Dewitte, W., Van Onckelen, H., Mariotti, D., 2001. Overexpression of KNAT1 in Lettuce Shifts Leaf Determinate Growth to a Shoot-Like Indeterminate Growth Associated with an Accumulation of Isopentenyl-Type Cytokinins. Plant Physiol. 126, 1370–1380.

Gade, P., Kalvakolanu, D.V., 2012. Chromatin Immunoprecipitation Assay as a Tool for Analyzing Transcription Factor Activity. Methods Mol. Biol. Clifton NJ 809, 85–104. doi:10.1007/978-1-61779-376-9_6

Geier, F., Lohmann, J.U., Gerstung, M., Maier, A.T., Timmer, J., Fleck, C., 2008. A Quantitative and Dynamic Model for Plant Stem Cell Regulation. PLoS ONE 3, e3553. doi:10.1371/journal.pone.0003553

Gelman, A., Roberts, G.O., Gilks, W.R., 1996. Efficient Metropolis jumping rules 599–607.

Gierer, A., Meinhardt, H., 1972. A theory of biological pattern formation. Kybernetik 12, 30–39. doi:10.1007/BF00289234

Golembeski, G.S., Kinmonth-Schultz, H.A., Song, Y.H., Imaizumi, T., 2014. Photoperiodic flowering regulation in Arabidopsis thaliana. Adv. Bot. Res. 72, 1–28. doi:10.1016/B978-0-12-417162-6.00001-8

Gómez-Ariza, J., Galbiati, F., Goretti, D., Brambilla, V., Shrestha, R., Pappolla, A., Courtois, B., Fornara, F., 2015. Loss of floral repressor function

adapts rice to higher latitudes in Europe. J. Exp. Bot. 66, 2027–2039. doi:10.1093/jxb/erv004

Goretti, D., Martignago, D., Landini, M., Brambilla, V., Gómez-Ariza, J., Gnesutta, N., Galbiati, F., Collani, S., Takagi, H., Terauchi, R., Mantovani, R., Fornara, F., 2017. Transcriptional and Post-transcriptional Mechanisms Limit Heading Date 1 (Hd1) Function to Adapt Rice to High Latitudes. PLOS Genet. 13, e1006530. doi:10.1371/journal.pgen.1006530

Graves, A., 2012. Supervised Sequence Labelling with Recurrent Neural Networks. Springer Science & Business Media.

Gruel, J., Landrein, B., Tarr, P., Schuster, C., Refahi, Y., Sampathkumar, A., Hamant, O., Meyerowitz, E.M., Jönsson, H., 2016. An epidermis-driven mechanism positions and scales stem cell niches in plants. Sci. Adv. 2, e1500989. doi:10.1126/sciadv.1500989

Gustafson-Brown, C., Savidge, B., Yanofsky, M.F., 1994. Regulation of the arabidopsis floral homeotic gene APETALA1. Cell 76, 131–143.

Harrison, R.L., 2010. Introduction To Monte Carlo Simulation. AIP Conf. Proc. 1204, 17–21. doi:10.1063/1.3295638

Hayama, R., Yokoi, S., Tamaki, S., Yano, M., Shimamoto, K., 2003. Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature 422, 719–722. doi:10.1038/nature01549

Heo, J.B., Sung, S., 2011. Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. Science 331, 76–79. doi:10.1126/science.1197349

Higgins, J.A., Bailey, P.C., Laurie, D.A., 2010. Comparative Genomics of Flowering Time Pathways Using Brachypodium distachyon as a Model for the Temperate Grasses. PLoS ONE 5, e10065. doi:10.1371/journal.pone.0010065#pone.0010065-Alexandre1

Hill, A.V., 1910. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J. Physiol. 40, iv–vii. doi:10.1113/jphysiol.1910.sp001386

Hindmarsh, A.C., 1982. ODEPACK, a systematized collection of ODE solvers. Lawrence Livermore National Laboratory.

Hodgman, T.C., Ajmera, I., 2015. The successful application of systems approaches in plant biology. Prog. Biophys. Mol. Biol., Multi-scale Systems Biology 117, 59–68. doi:10.1016/j.pbiomolbio.2015.01.002

Irish, V.F., Sussex, I.M., 1990. Function of the apetala-1 gene during Arabidopsis floral development. Plant Cell 2, 741–753. doi:10.1105/tpc.2.8.741

Jackson, S.A., 2016. Rice: The First Crop Genome. Rice 9. doi:10.1186/s12284-016-0087-4

Jaeger, K.E., Pullen, N., Lamzin, S., Morris, R.J., Wigge, P.A., 2013. Interlocking Feedback Loops Govern the Dynamic Behavior of the Floral Transition in Arabidopsis. Plant Cell Online 25, 820–833. doi:10.1105/tpc.113.109355

Jaeger, K.E., Wigge, P.A., 2007. FT Protein Acts as a Long-Range Signal in Arabidopsis. Curr. Biol. 17, 1050–1054. doi:10.1016/j.cub.2007.05.008

Jones, E., Oliphant, T., Peterson, P., others, 2001. SciPy: Open source scientific tools for Python.

Jönsson, H., Heisler, M.G., Shapiro, B.E., Meyerowitz, E.M., Mjolsness, E., 2006. An auxin-driven polarized transport model for phyllotaxis. Proc. Natl. Acad. Sci. U. S. A. 103, 1633–1638. doi:10.1073/pnas.0509839103

Jönsson, H., Shapiro, B.E., Meyerowitz, E.M., Mjolsness, E., 2003. 8 - Signalling in multicellular models of plant development, in: On Growth, Form and Computers. Academic Press, London, pp. 156–161. doi:10.1016/B978-012428765-5/50041-4

Jung, C., Müller, A.E., 2009. Flowering time control and applications in plant breeding. Trends Plant Sci. 14, 563–573. doi:10.1016/j.tplants.2009.07.005

Jung, J.-H., Ju, Y., Seo, P.J., Lee, J.-H., Park, C.-M., 2012. The SOC1-SPL module integrates photoperiod and gibberellic acid signals to control flowering time in Arabidopsis. Plant J. 69, 577–588. doi:10.1111/j.1365-313X.2011.04813.x

Kang, C.C., Chuang, Y.J., Tung, K.C., Chao, C.C., Tang, C.Y., Peng, S.C., Wong, D.S.H., 2011. A genetic algorithm-based boolean delay model of intracellular signal transduction in inflammation. BMC Bioinformatics 12, S17. doi:10.1186/1471-2105-12-S1-S17

Kauffman, S.A., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. 22, 437–467. doi:10.1016/0022-5193(69)90015-0

Kaufmann, K., Wellmer, F., Muiño, J.M., Ferrier, T., Wuest, S.E., Kumar, V., Serrano-Mislata, A., Madueño, F., Krajewski, P., Meyerowitz, E.M., Angenent, G.C., Riechmann, J.L., 2010. Orchestration of Floral Initiation by APETALA1. Science 328, 85–89. doi:10.1126/science.1185244

Kempin, S.A., Savidge, B., Yanofsky, M.F., 1995. Molecular basis of the cauliflower phenotype in Arabidopsis. Science 267, 522–525. doi:10.1126/science.7824951

Kerppola, T.K., 2008. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. Annu. Rev.

Biophys. 37, 465–487. doi:10.1146/annurev.biophys.37.032807.125842

Klintenäs, M., Pin, P.A., Benlloch, R., Ingvarsson, P.K., Nilsson, O., 2012. Analysis of conifer FLOWERING LOCUS T/TERMINAL FLOWER1-like genes provides evidence for dramatic biochemical evolution in the angiosperm FT lineage. New Phytol. 196, 1260–1273. doi:10.1111/j.1469-8137.2012.04332.x

Komiya, R., Yokoi, S., Shimamoto, K., 2009. A gene network for long-day flowering activates RFT1 encoding a mobile flowering signal in rice. Development 136, 3443–3450. doi:10.1242/dev.040170

Koo, B.-H., Yoo, S.-C., Park, J.-W., Kwon, C.-T., Lee, B.-D., An, G., Zhang, Z., Li, J., Li, Z., Paek, N.-C., 2013. Natural Variation in OsPRR37 Regulates Heading Date and Contributes to Rice Cultivation at a Wide Range of Latitudes. Mol. Plant 6, 1877–1888. doi:10.1093/mp/sst088

Kyozuka, J., Kobayashi, T., Morita, M., Shimamoto, K., 2000. Spatially and Temporally Regulated Expression of Rice MADS Box Genes with Similarity to Arabidopsis Class A, B and C Genes. Plant Cell Physiol. 41, 710–718. doi:10.1093/pcp/41.6.710

Kyozuka, J., Konishi, S., Nemoto, K., Izawa, T., Shimamoto, K., 1998. Down-regulation of RFL, the FLO/LFY homolog of rice, accompanied with panicle branch initiation. Proc. Natl. Acad. Sci. 95, 1979–1982.

La Rota, C., Chopard, J., Das, P., Paindavoine, S., Rozier, F., Farcot, E., Godin, C., Traas, J., Monéger, F., 2011. A Data-Driven Integrative Model of Sepal Primordium Polarity in Arabidopsis. Plant Cell Online 23, 4318–4333. doi:10.1105/tpc.111.092619

Le Novère, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J.L., Hucka, M., 2006. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res. 34, D689–D691. doi:10.1093/nar/gkj092

Lee, I., Michaels, S.D., Masshardt, A.S., Amasino, R.M., 1994. The late-flowering phenotype of FRIGIDA and mutations in LUMINIDEPENDENS is suppressed in the Landsberg erecta strain of Arabidopsis. Plant J. 6, 903–909. doi:10.1046/j.1365-313X.1994.6060903.x

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. doi:10.1111/j.1365-313X.2008.03552.x

Lee, J.H., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.S., Ahn, J.H., 2007. Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes Dev. 21, 397–402. doi:10.1101/gad.1518407

Lee, S., Kim, J., Han, J.-J., Han, M.-J., An, G., 2004. Functional analyses of the flowering time gene OsMADS50, the putative SUPPRESSOR OF OVEREXPRESSION OF CO 1/AGAMOUS-LIKE 20 (SOC1/AGL20) ortholog in rice. Plant J. 38, 754–764. doi:10.1111/j.1365-313X.2004.02082.x

Li, C., Donizelli, M., Rodriguez, N., Dharuri, H., Endler, L., Chelliah, V., Li, L., He, E., Henry, A., Stefan, M.I., Snoep, J.L., Hucka, M., Le Novère, N., Laibe, C., 2010. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. BMC Syst. Biol. 4, 92. doi:10.1186/1752-0509-4-92

Liljegren, S.J., Gustafson-Brown, C., Pinyopich, A., Ditta, G.S., Yanofsky, M.F., 1999. Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 Specify Meristem Fate. Plant Cell Online 11, 1007–1018. doi:10.1105/tpc.11.6.1007

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. doi:10.1242/dev.020255

Liu, C., Teo, Z.W.N., Bi, Y., Song, S., Xi, W., Yang, X., Yin, Z., Yu, H., 2013. A Conserved Genetic Pathway Determines Inflorescence Architecture in Arabidopsis and Rice. Dev. Cell 24, 612–622. doi:10.1016/j.devcel.2013.02.013

Liu, C., Thong, Z., Yu, H., 2009. Coming into bloom: the specification of floral meristems. Development 136, 3379–3391. doi:10.1242/dev.033076

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. doi:10.1242/dev.003103

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. doi:10.1038/360273a0

Mandel, M.A., Yanofsky, M.F., 1995. A gene triggering flower formation in Arabidopsis. Nature 377, 522–524. doi:10.1038/377522a0

Meinhardt, H., Gierer, A., 1974. Applications of a Theory of Biological Pattern Formation Based on Lateral Inhibition. J. Cell Sci. 15, 321–346.

Mendoza, L., Xenarios, I., 2006. A method for the generation of standardized qualitative dynamical systems of regulatory networks. Theor. Biol. Med. Model. 3, 13. doi:10.1186/1742-4682-3-13

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of State Calculations by Fast Computing Machines. J. Chem. Phys. 21, 1087–1092. doi:10.1063/1.1699114

Michaelis, L., Menten, M.L., Johnson, K.A., Goody, R.S., 2011. The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. Biochemistry (Mosc.) 50, 8264–8269. doi:10.1021/bi201284u

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. doi:10.1046/j.1365-313X.2003.01671.x

Narsai, R., Howell, K.A., Millar, A.H., O'Toole, N., Small, I., Whelan, J., 2007. Genome-Wide Analysis of mRNA Decay Rates and Their Determinants in Arabidopsis thaliana. Plant Cell 19, 3418–3436. doi:10.1105/tpc.107.055046

Nelder, J.A., Mead, R., 1965. A Simplex Method for Function Minimization. Comput. J. 7, 308–313. doi:10.1093/comjnl/7.4.308

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825−2830.

Pidkowich, M.S., Klenz, J.E., Haughn, G.W., 1999. The making of a flower: control of floral meristem identity in Arabidopsis. Trends Plant Sci. 4, 64–70. doi:10.1016/S1360-1385(98)01369-7

Pradal, C., Dufour-Kowalski, S., Boudon, F., Fournier, C., Godin, C., 2008. OpenAlea: a visual programming and component-based software platform for plant modelling. Funct. Plant Biol. 35, 751. doi:10.1071/FP08084

Prusinkiewicz, P., 2007. Modelling architecture of crop plants using L-systems. Frontis 22, 27–42.

Rayle, D.L., Cleland, R.E., 1992. The Acid Growth Theory of auxin-induced cell elongation is alive and well. Plant Physiol. 99, 1271–1274. doi:10.1104/pp.99.4.1271

Reece, J.B., Campbell, N.A., 2011. Campbell Biology, 9th ed. Benjamin Cummings : imprint of Pearson, Boston.

Reinhardt, D., Mandel, T., Kuhlemeier, C., 2000. Auxin Regulates the Initiation and Radial Position of Plant Lateral Organs. Plant Cell Online 12, 507–518. doi:10.1105/tpc.12.4.507

Reinhardt, D., Pesce, E.-R., Stieger, P., Mandel, T., Baltensperger, K., Bennett, M., Traas, J., Friml, J., Kuhlemeier, C., 2003. Regulation of phyllotaxis by polar auxin transport. Nature 426, 255–260. doi:10.1038/nature02081

Roli, A., Arcaroli, C., Lazzarini, M., Benedettini, S., 2011. Boolean Networks Design by Genetic Algorithms. ArXiv11016018 Cs Nlin.

Rosenfeld, N., Alon, U., 2003. Response Delays and the Structure of Transcription Networks. J. Mol. Biol. 329, 645–654. doi:10.1016/S0022-2836(03)00506-0

Scheidegger, A., 2012. adaptMCMC: Implementation of a generic adaptive Monte Carlo Markov Chain sampler.

Schmidhuber, J., Wierstra, D., Gagliolo, M., Gomez, F., 2007. Training recurrent networks by Evolino. Neural Comput. 19, 757–779. doi:10.1162/neco.2007.19.3.757

Schultz, E.A., Haughn, G.W., 1991. LEAFY, a Homeotic Gene That Regulates Inflorescence Development in Arabidopsis. Plant Cell 3, 771–781. doi:10.1105/tpc.3.8.771

Schwarz, G., 1978. Estimating the Dimension of a Model. Ann. Stat. 6, 461–464. doi:10.1214/aos/1176344136

Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Kröber, 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. doi:10.1101/gad.373506

Sekar, R.B., Periasamy, A., 2003. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. J. Cell Biol. 160, 629–633. doi:10.1083/jcb.200210140

Shannon, S., Meeks-Wagner, D., 1991. A Mutation in the Arabidopsis TFL1 Gene Affects Inflorescence Meristem Development. Plant Cell 3, 877–892.

Simon, R., Igeño, M.I., Coupland, G., 1996. Activation of floral meristem identity genes in Arabidopsis. Nature 384, 59–62. doi:10.1038/384059a0

Simpson, G.G., Dean, C., 2002. Arabidopsis, the Rosetta Stone of Flowering Time? Science 296, 285–289. doi:10.1126/science.296.5566.285

Sklyar, O., Murdoch, D., Smith, M., Eddelbuettel, D., Francois, R., Soetaert, K., 2015. inline: Functions to Inline C, C++, Fortran Function Calls from R.

Smith, T.D., 2014. tdsmith/eleven [WWW Document]. GitHub. URL https://github.com/tdsmith/eleven (accessed 1.30.17).

Smyth, D.R., 1995. Flower Development: Origin of the cauliflower. Curr. Biol. 5, 361–363. doi:10.1016/S0960-9822(95)00072-8

Soetaert, K., Petzoldt, T., Setzer, R.W., authors, odepack, 2016. deSolve: Solvers for Initial Value Problems of Differential Equations (ODE, DAE, DDE).

Tanaka, W., Pautler, M., Jackson, D., Hirano, H.-Y., 2013. Grass Meristems II: Inflorescence Architecture, Flower Development and Meristem Fate. Plant Cell Physiol. 54, 313–324. doi:10.1093/pcp/pct016

Torti, S., Fornara, F., Vincent, C., Andrés, F., Nordström, K., Göbel, U., Knoll, D., Schoof, H., Coupland, G., 2012. Analysis of the Arabidopsis Shoot Meristem Transcriptome during Floral Transition Identifies Distinct Regulatory Patterns and a Leucine-Rich Repeat Protein That Promotes Flowering. Plant Cell Online 24, 444–462. doi:10.1105/tpc.111.092791

Traas, J., Vernoux, T., 2002. The shoot apical meristem: the dynamics of a stable structure. Philos. Trans. R. Soc. B Biol. Sci. 357, 737–747. doi:10.1098/rstb.2002.1091

Turing, A.M., 1952. The Chemical Basis of Morphogenesis. Royal Society.

Valentim, F.L., Mourik, S. van, Posé, D., Kim, M.C., Schmid, M., van Ham, R.C.H.J., Busscher, M., Sanchez-Perez, G.F., Molenaar, J., Angenent, G.C., Immink, R.G.H., van Dijk, A.D.J., 2015. A Quantitative and Dynamic Model of the Arabidopsis Flowering Time Gene Regulatory Network. PLoS ONE 10, e0116973. doi:10.1371/journal.pone.0116973

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3, research0034. doi:10.1186/gb-2002-3-7-research0034

Vernoux, T., Brunoud, G., Farcot, E., Morin, V., Daele, H.V. den, Legrand, J., Oliva, M., Das, P., Larrieu, A., Wells, D., Guédon, Y., Armitage, L., Picard, F., Guyomarc'h, S., Cellier, C., Parry, G., Koumproglou, R., Doonan, J.H., Estelle, M., Godin, C., Kepinski, S., Bennett, M., Veylder, L.D., Traas, J., 2011. The auxin signalling network translates dynamic input into robust patterning at the shoot apex. Mol. Syst. Biol. 7. doi:10.1038/msb.2011.39

Vihola, M., 2012. Robust adaptive Metropolis algorithm with coerced acceptance rate. Stat. Comput. 22, 997–1008. doi:10.1007/s11222-011-9269-5

Wang, J.-W., Czech, B., Weigel, D., 2009. miR156-Regulated SPL Transcription Factors Define an Endogenous Flowering Pathway in Arabidopsis thaliana. Cell 138, 738–749. doi:10.1016/j.cell.2009.06.014

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.

Weigel, D., Nilsson, O., 1995. A developmental switch sufficient for flower initiation in diverse plants. Nature 377, 495–500. doi:10.1038/377495a0

Welch, S.M., Roe, J.L., Dong, Z., 2003. A genetic neural network model of flowering time control in Arabidopsis thaliana. Agron J 95, 71–81.

Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U., Weigel, D., 2005. Integration of Spatial and Temporal Information During Floral Induction in Arabidopsis. Science 309, 1056–1059. doi:10.1126/science.1114358

Willis, L., Refahi, Y., Wightman, R., Landrein, B., Teles, J., Huang, K.C., Meyerowitz, E.M., Jönsson, H., 2016. Cell size and growth regulation in the Arabidopsis thaliana apical stem cell niche. Proc. Natl. Acad. Sci. 113, E8238–E8246. doi:10.1073/pnas.1616768113

Wink, M., 1993. The plant vacuole: a multifunctional compartment. J. Exp. Bot., Supplement 44, 231–246.

Wittmann, D.M., Krumsiek, J., Saez-Rodriguez, J., Lauffenburger, D.A., Klamt, S., Theis, F.J., 2009. Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. BMC Syst. Biol. 3, 98. doi:10.1186/1752-0509-3-98

Yamaguchi, A., Kobayashi, Y., Goto, K., Abe, M., Araki, T., 2005. TWIN SISTER OF FT (TSF) Acts as a Floral Pathway Integrator Redundantly with FT. Plant Cell Physiol. 46, 1175–1189. doi:10.1093/pcp/pci151

Yang, Y., 2005. Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. Biometrika 92, 937–950. doi:10.1093/biomet/92.4.937