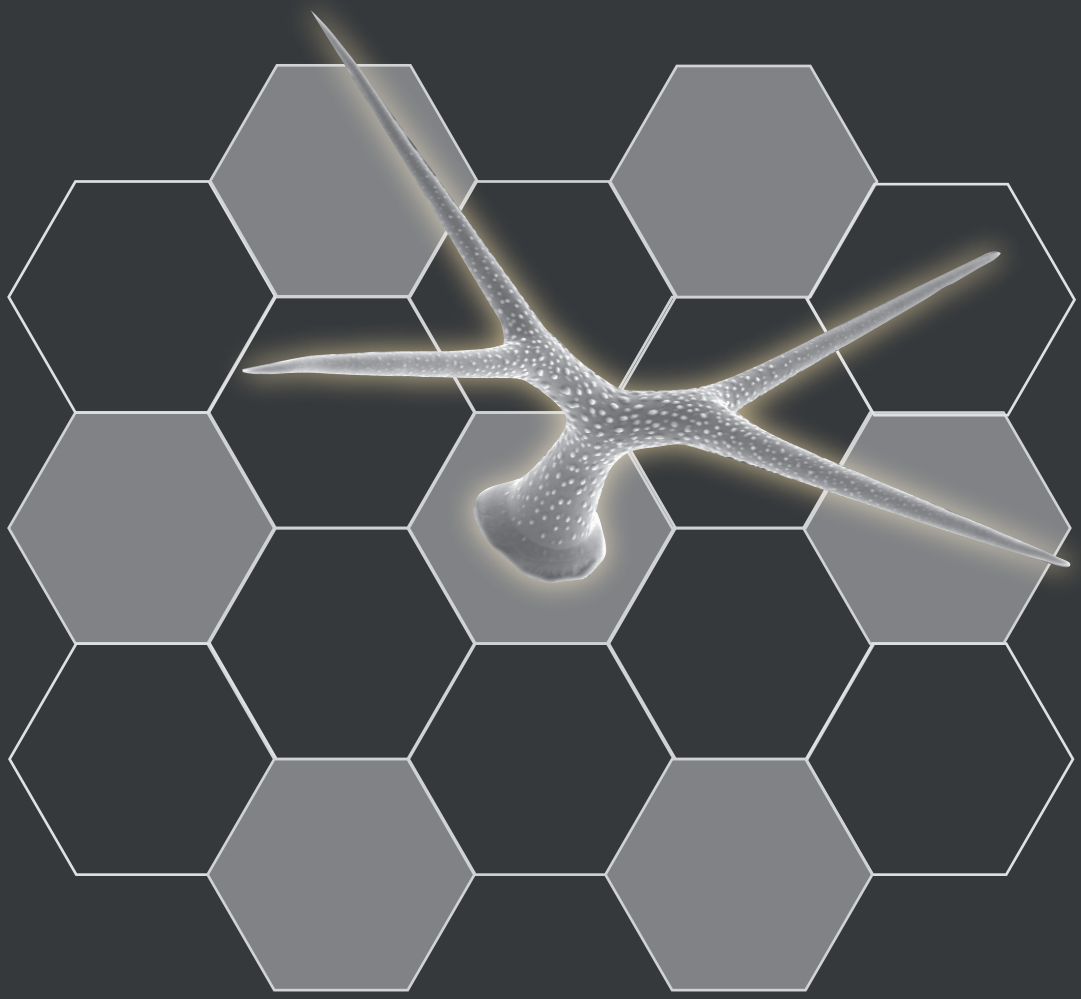
Mathematical Modelling of Trichome Patterning



Anna Deneer

Propositions

1. Even when all parameters of a model are unknown, it is possible to arrive at meaningful predictions.
   (this thesis)

2. The versatility and ease of genetic manipulation of the trichome system make it extremely suitable for studies on patterning and development.
   (this thesis)

3. The prevalent incentives in science revolve around building a reputation, which undermines two essential aspects in science, namely reproducibility and sharing of knowledge.

4. Introducing a publication bias by not publishing research that shows small effects or negative results leads to systematically unrepresentative scientific literature.

5. Advisory boards form the basis of self-criticism in a democracy, but this is only effective if they can function with complete independence.

6. The Turing test is outdated; a conversation held by an AI, no matter how convincing, is not proof of independent intelligence.

Propositions belonging to the thesis, entitled

Mathematical Modelling of Trichome Patterning

Anna Deneer
Wageningen, 5 December 2022

Mathematical modelling of trichome patterning

Anna Deneer

Thesis committee

Promotors
Prof. Dr J. Molenaar
Professor of Applied Mathematics
Wageningen University & Research

Prof. Dr V.A.P. Martins dos Santos
Personal chair, Bioprocess Engineering Group
Wageningen University & Research

Co-promotors
Prof. Dr M. Hülskamp
Professor of Plant Genetics, Institute of Botany
University of Cologne, Germany

Dr C. Fleck
Group head, Spatial Systems Biology Group
University of Freiburg, Germany

Other members
Prof. Dr C. S. Testerink, Wageningen University & Research
Prof. Dr B. J. G. Scheres, Rijk Zwaan Breeding B.V. & Wageningen University & Research
Prof. Dr R. M. H. Merks, Leiden University
Dr V. A. Grieneisen, Cardiff University, United Kingdom

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

Mathematical modelling of trichome patterning

A. Deneer

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 05 December 2022
at 1:30 pm in the Omnia Auditorium.

A. Deneer
Mathematical modelling of trichome patterning,
239 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2022)
With references, with summary in English and Dutch

ISBN: 978-94-6447-452-7
DOI: 10.18174/579174

Contents

Chapter 1

Introduction

How did the zebra get its stripes, the leopard its spots and the plant its haircut? Although the last question is seemingly different from the first two, theory suggests that many patterns in biology share common principles [1–5]. In this thesis, the focus is on the latter question of plant 'haircuts', or more precisely: the hair-like structures that are formed on the epidermis of plant leaves, called 'trichomes' [6]. The formation of patterns by trichomes may provide fundamental insight into how cells that are similar in early life stages differentiate over time into specific functions and shapes [7]. Given the existence of an intricate regulatory network underlying the regulation of trichome patterning [8–10], the use of mathematical modelling is essential in order to get a grasp on its inherent complexity. In this introductory chapter, an overview is given of the general theory behind models of pattern formation, after which we place this in the context of trichome patterning and concomitant challenges.

1.1 Patterns in nature

Although nature shows a great variety of shapes and structures, underneath all this diversity a surprising amount of regularity is found. Even between markedly different scales and contexts (e.g. sand dunes and pigment patterns of fish as in Figure 1.1), there is a similarity between natural patterns that strongly suggests that such phenomena could be explained by a common theoretical framework. The search for this theoretical framework has been led by mathematicians hoping to answer questions such as "*How do these patterns arise?*" and "*What are the common features?*".

The Scottish biologist and mathematician Sir D'Arcy Wentworth Thompson provided a first methodological answer in his book *On Growth and Form* in 1917 [2]. There, he combined natural history, biology, mathematics, and

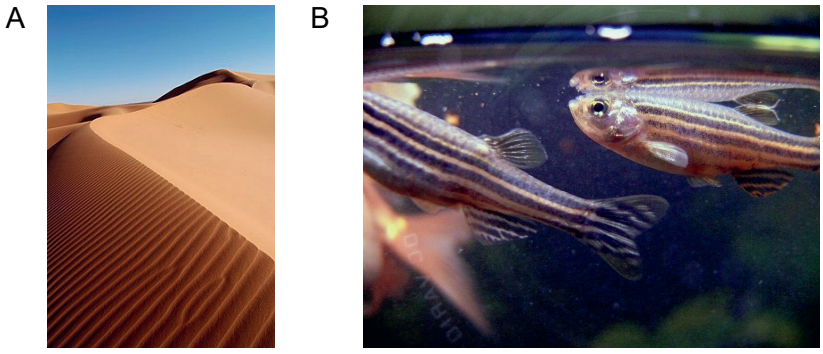


Figure 1.1: **Examples of patterns in nature.** A: Sand dunes in Morocco, self-organized into a regular pattern. B: *Danio rerio*, or zebrafish, known for its stripes and often used in patterning studies.
Images from Wikipedia Commons, distributed under a CC BY-SA 2.0 licence.

physics in search of a quantitative description of the mechanisms underlying plant and animal pattern formation. By applying tools from mathematics and physics, Thompson demonstrated how seemingly complex problems could be described by elegant and simple forms of reasoning. Early in the book, Thompson remarks that the phenomena of natural form can be found at many levels of organisation,

> "The waves of the sea, the little ripples on the shore, the sweeping curve of the sandy bay between the headlines, the outline of the hills, the shape of the clouds, all these are so many riddles of form, so many problems of morphology, and all of them the physicist can more or less easily read and adequately solve."

Thompson states his aim of finding a way of combining mathematics and physics for the purpose of explaining growth and form, as follows:

> "My sole purpose is to correlate with mathematical statement and physical law certain of the simpler outward phenomena of organic growth and structure or form, while all the while regarding the fabric of the organism, *ex hypothesi*, as a material and mechanical configuration."

Through his description of the "*mathematical beauty*" of nature, Thompson has inspired many scientists to explore a theoretical approach to explaining

natural form – among which Alan Turing, whose impact on the field of mathematical biology and the theory behind morphogenesis is substantial.

1.2 Turing patterns

Similar to Thompson's assumptions that physical forces can explain biological forms, Turing hypothesised that diffusing chemicals that react with one another could explain the formation of anatomical patterns in developing tissues, and that these simple chemical processes can be described precisely and mathematically [3]. In Turing's ground-breaking paper *The Chemical Basis of Morphogenesis* a set of chemical kinetic equations is introduced that may generate patterns resembling those seen in biology [3]. Since then, these patterns have been coined *Turing patterns*. These equations describe two mechanisms: reaction, where chemicals are produced and decayed; and diffusion, where chemicals spread across the tissue. This simple system consisting of only two species is capable of breaking the symmetry of a spatially homogeneous mixture of chemicals and leads to the development of structures, i.e. spatial patterns [3]. In view of their assumed form-producing properties, Turing termed these chemicals *morphogens* and the related chemical system is called a reaction-diffusion system. A surprising outcome of Turing's model is that he showed that a system which is stable in the absence of diffusion, could become unstable in the presence of diffusion. This finding was really counter-intuitive as diffusion is typically thought of as a homogenizing process. However, Turing showed that – given certain system kinetics – two stabilizing processes could lead to an instability and thus the formation of spatial patterns. Furthermore, these patterns are *self-organizing*, which means that the breaking of the initial symmetry follows from the system dynamics itself, triggered by random perturbations [3].

Since Turing's prediction of diffusion-driven instability, it took almost 40 years before Turing patterns were observed in a real chemical reaction [11]. Nowadays, such chemical or molecular signals have been frequently found in animals and plants and have been shown to spread across tissues describing patterns that are followed by processes like cell differentiation [12–16]. Crucially, Turing's model demonstrated that the dynamic changes in the distribution of a morphogen leads to tissue patterns. This evoked the question that is still relevant in developmental biology to this day: "*How are morphogen profiles shaped and interpreted in tissues?*".

Turing's approach exemplified the power of taking a mathematical and quantitative perspective on biological questions. This is reflected by the abundance of follow-up studies of morphogen-mediated patterning based on Turing's reaction-diffusion system [5, 15, 17]. The work of Turing and Thompson advocated the exploration of biological processes by finding simple mathematical descriptions based on physical principles. In the following section an overview is given of Turing's theory, which will play a central role in this thesis when dealing with the analysis of trichome patterns.

1.3 The theory behind Turing patterns

The Turing reaction-diffusion system [3] is of the form

$$\frac{\partial \mathbf{w}}{\partial t} = \mathbf{F}(\mathbf{w};\mu) + \Delta \mathbf{D} \Delta \mathbf{w}, \tag{1.1}$$

where $\mathbf{w}$ is a vector of chemical concentrations, $\mathbf{D}$ a matrix with diffusion constants (diagonal with strictly positive elements) and $\mathbf{F}(\mathbf{w};\mu)$ the (non-linear) reaction kinetics depending on (a number of) parameters μ. The simplest case that exhibits diffusion driven instability is a 2-component system. In that case, $\mathbf{w}$, $\mathbf{D}$ and $\mathbf{F}$ are written as

$$\mathbf{w} = \begin{pmatrix} u \\ v \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} f(u,v;\mu) \\ g(u,v;\mu) \end{pmatrix}, \tag{1.2}$$

with $u(\mathbf{r},t), v(\mathbf{r},t)$ concentrations of the two components at the spatial position $\mathbf{r} = (x,y)$ and time t. After scaling, $\mathbf{D}$ has diagonal elements 1 and d, the ratio of the diffusion coefficients of the components u and v.

A requirement is the existence of a stable homogeneous steady state $\bar{\mathbf{w}} = (\bar{u},\bar{v})$ of the system without diffusion, i.e. the solutions of the algebraic system $f(\bar{u},\bar{v};\mu) = g(\bar{u},\bar{v};\mu) = 0$, which are constant in time and homogeneous in space [18, 19]. The mechanisms by which $\bar{\mathbf{w}}$ may be destabilized determines the onset of the pattern. Usually, the initial conditions are taken to be random perturbations of $\bar{\mathbf{w}}$. It is the final pattern that is of interest, not the initialization of the pattern. In the case of diffusion driven instability, the homogeneous steady state must be linearly stable and destabilized upon the introduction of diffusion [18, 19]. To determine the stability of $\bar{\mathbf{w}}$ a small perturbation is introduced:

$$\bar{\mathbf{w}}' = \mathbf{w} - \bar{\mathbf{w}}. \tag{1.3}$$

If the perturbation $|\bar{\mathbf{w}}'|$ is sufficiently small, we may linearize (1.1) about the homogeneous steady state. This is achieved by writing a Taylor series expansion for $\mathbf{F}(\bar{\mathbf{w}})$ about $\bar{\mathbf{w}}$ and retaining the linear contributions [18, 19]. For (1.1) the linearized version in the absence of diffusion is

$$\frac{\partial \bar{\mathbf{w}}'}{\partial t} = \mathbf{J}\bar{\mathbf{w}}', \quad \mathbf{J} = \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}_{(\bar{u},\bar{v})}, \tag{1.4}$$

where $\mathbf{J}$ is the Jacobian matrix with $f_u = \frac{\partial f}{\partial u}(\bar{u},\bar{v})$, $g_u = \frac{\partial g}{\partial u}(\bar{u},\bar{v})$, etc. Solutions of (1.4) are of the form

$$\bar{\mathbf{w}}'(t) \propto e^{\lambda t}, \tag{1.5}$$

where, after substitution of (1.5) into (1.4), the eigenvalue λ can be determined from the associated characteristic polynomial,

$$\begin{aligned} |\mathbf{J} - \lambda \mathbf{I}| &= \left| \begin{pmatrix} f_u - \lambda & f_v \\ g_u & g_v - \lambda \end{pmatrix} \right| = 0 \\ \Rightarrow \quad & \lambda^2 - (f_u + g_v)\lambda + (f_u g_v - f_v g_u) = 0. \end{aligned} \tag{1.6}$$

From the roots of the characteristic polynomial it follows that

$$\lambda_1, \lambda_2 = \frac{1}{2}\left[(f_u + g_v) \pm \left((f_u + g_v)^2 - 4(f_u g_v - f_v g_u)^{1/2}\right)\right]. \tag{1.7}$$

The steady state is stable if $\mathrm{Re}\,\lambda < 0$, which is guaranteed if

$$\mathrm{tr}\,\mathbf{J} = f_u + g_v < 0, \quad |\mathbf{J}| = f_u g_v - f_v g_u > 0. \tag{1.8}$$

Having derived the conditions for linear stability in the absence of diffusion, the next goal is to find the necessary conditions upon which diffusion will drive the system to instability. Hence, the system in (1.1) is linearized about the steady state. This yields:

$$\frac{\partial \bar{\mathbf{w}}'}{\partial t} = \mathbf{J}\bar{\mathbf{w}}' + \mathbf{D}\nabla^2 \bar{\mathbf{w}}'. \tag{1.9}$$

The solutions of this linear problem are of the form

$$\bar{\mathbf{w}}'(\mathbf{r}, t) = \sum_k c_k e^{\lambda_k t} \mathbf{W}_k(\mathbf{r}), \tag{1.10}$$

where $\mathbf{W}_k(\mathbf{r})$ are the wave modes and c_k are constants to be determined by a Fourier expansion of the initial conditions. For example, in the case of a 1-dimensional system with a domain of $0 \le x \le a$ with zero-flux boundary conditions, $\mathbf{W} \propto \cos(n\pi x/a)$ where n is an integer and $k = n\pi/a$ is the wavenumber, which is a measure of the wavelike pattern and $1/k$ is proportional to the

wavelength that grows at rate λ_k. Now, the goal is to find which modes are unstable, i.e. $\mathrm{Re}\,\lambda_k > 0$, for any non-zero value of k. In the case where multiple modes are growing, small changes in the initial conditions can lead to different final outcomes. The integer values of k for which λ_k has a positive real part indicates the number of peaks the final pattern may form. For example, in the 1D case, if λ_4 is the only growth rate with positive real part, then the system will tend to a solution with a dominant $\cos(4\pi x/a)$ function, so the final pattern will have the corresponding number of peaks. If multiple growth rates are positive, then the final pattern is a superposition of modes. The precise form of this superposition cannot be predicted due to the initial perturbations being random. Furthermore, the nonlinearity of the system may give rise to unexpected patterning even if the initial conditions would be deterministic. To determine for which criteria the system has real positive eigenvalues, (1.10) is substituted into (1.9) and λ_k is determined by the roots of the characteristic polynomial

$$\begin{aligned} &\left|\lambda\mathbf{I} - \mathbf{J} + \mathbf{D}k^2\right| = \left|\begin{pmatrix} \lambda - f_u + k^2 & -f_v \\ -g_u & \lambda - g_v + dk^2 \end{pmatrix}\right| = 0 \\ &\Rightarrow \quad \lambda^2 + \lambda\left[k^2(1+d) - (f_u + g_v)\right] + h(k^2) = 0, \\ &h(k^2) = dk^4 - (df_u + g_v)k^2 + |\mathbf{J}(\bar{u}, \bar{v})| \end{aligned} \tag{1.11}$$

Together with the inequalities (1.8) derived for the stability of the homogeneous steady state, the conditions needed for a Turing instability are

$$df_u + g_v > 0, \quad (df_u + g_v)^2 - 4d(f_u g_v - f_v g_u) > 0. \tag{1.12}$$

These inequalities indicate that $d > 1$ is a necessary condition for diffusion-driven instability pattern formation [18, 19], so the species in the system need to have different diffusion coefficients.

The linear stability analysis described above is an efficient method for determining the bifurcation thresholds for diffusion-driven instability, for identifying the Turing space (the part of parameter space where patterns are formed), and to determine the characteristic length of the pattern (given by the mode with the largest real positive eigenvalue) [18, 19].

1.4 Extensions of Turing's model

As Turing himself already noted,

> "This model will be a simplification and an idealization, and consequently a falsification. It is to be hoped that the features retained for discussion are those of greatest importance in the present state of knowledge."

The simple two-component reaction-diffusion system was a simplification that would require further refinement when applied to the complexity behind patterning in developmental systems. As mentioned in the previous section, Turing patterns are sensitive to the initial random perturbation, indicating a lack of robustness [20–22]. In some cases in biology this is exactly what is observed, e.g. no two zebra show the same stripe pattern, they are unique like fingerprints. However, in other cases (e.g. digit development) the pattern is highly robustly reproduced. Extensions of Turing's model which include additional components, details like stochasicity, and domain growth have been shown to increase this robustness [21, 23–27].

Two particular variations on Turing's model are highlighted in the following in view of their prominence and relevance in numerous studies on biological patterns, including trichome patterning. These two variations are reaction-diffusion models with different reaction mechanics. The first one is the activator-inhibitor model, also called the Gierer-Meinhardt model, which relies on the principles of local induction and long-range inhibition by two biochemical reactions [5, 17, 28, 29]. In such a system, small deviations from the initially homogeneous field of activator concentrations will be amplified by auto-catalysis. This will lead to a local increase in the inhibitor concentration, which is being produced by the activator and diffuses into the surrounding tissue at a faster rate than the activator, where it will suppress activator production, leading to local peaks of activator and inhibitor concentrations [5]. Such a simple system of two interacting components is capable of forming patterns from an initially homogeneous field perturbed by small fluctuations, i.e. it is self-organizing as is the case in biological pattern formation. These interactions are described by the coupled differential equations

$$\frac{\partial u}{\partial t} = \alpha + \frac{u^2}{v(1+\kappa u^2)} - \mu u + D_u \nabla^2 u, \tag{1.13}$$

$$\frac{\partial v}{\partial t} = u^2 - v + D_v \nabla^2 v, \tag{1.14}$$

where α is the basal production rate of the activator, κ is a measure of the inhibition strength, μ is the decay rate of the activator, and D_u, D_v are the

diffusion constants for the activator and the inhibitor, respectively. An alternative form which relies on similar principles is the activator-depletion model [5], in which a substrate is depleted in the autocatalytic loop of the activator. This alternative form is

$$\frac{\partial u}{\partial t} = u^2 s - u + D_u \nabla^2 u, \tag{1.15}$$

$$\frac{\partial s}{\partial t} = \beta - u^2 s - \gamma s + D_s \nabla^2 v, \tag{1.16}$$

where β is the substrate basal production rate and γ the substrate decay rate.

One of the criteria for pattern formation for both models is a difference in diffusion rates. For the activator-inhibitor model it is necessary that $D_v > D_u$ and for the substrate-depletion model $D_s > D_u$. Gierer and Meinhardt have applied the model to a wide variety of biological patterning systems, often leading to a resemblance to the real pattern [17, 30]. However, it has been shown that the parameter sets that lead to patterns are often unrealistically restrictive [26]. In order to increase the range of patterning and improve robustness, some variations on the classical Gierer-Meinhardt models have been investigated [19, 26].

1.5 Trichome patterning

The pattern of hair cells on the leaf epidermis in *Arabidopsis thaliana* is regulated by a mechanism that is capable of forming the pattern without any pre-existing information like cell lineage or signalling from underlying cell layers [10, 31]. This mechanism robustly produces a pattern of hairs on leaves that maintains a consistent spacing between any pair of hair-cells, interspersed on average by the same amount of non-hair cells, with a certain degree of irregularity [32, 33]. The development of trichomes has become a well-known model of cell differentiation and growth due to the ease of accessibility and manipulability [7, 34–36]. In *Arabidopsis*, the trichomes are not essential for the plant under laboratory conditions, which has greatly facilitated the isolation and study of mutants [34]. These mutants show phenotypical defects that provide insight into the function of the underlying genotype. Based on the resulting patterning defect upon mutation, genes have been identified that fulfil the role of activator (induces trichome fate) and inhibitor (supresses trichome fate) [7, 31, 35]. Given such obvious classifications, it naturally follows that a model describing trichome patterning could be based on the Gierer-Meinhardt models, considering that the interactions included in those models

are activator-inhibitor reactions [5].

The network of genes involved in trichome development is extensive, involving around 40 genes that have been identified through mutagenesis screens [34, 37]. Out of all of these, a core network that is relevant for patterning has been identified and various subsets of that core have been used in modelling approaches to explain certain aspects of pattern formation [38–42]. An important mechanism of this core network is the formation of a complex by the designated activators [43, 44]. This complex is called the MBW-complex and consists of R2R3 MYB-type transcription factors encoded by the genes *GLABRA1* (GL1) and MYB23 [45], a bHLH protein encoded by *GLABRA3* (GL3) and *ENHANCER OF GLABRA3* (EGL3) [34, 44, 46] and finally, the WD40-protein encoded by *TRANSPARANT TESTA GLABRA1* (TTG1) [47–49]. All these genes fall in the group of positive regulators, based on their respective mutant phenotypes [34]. Mutations in GL1 and TTG1 each result in a complete absence of trichomes, whereas knockouts of GL3 and EGL3 results in leaves with less trichomes than wildtype – only when both GL3 and EGL3 are mutated the leaf is devoid of trichomes [34]. The group of negative regulators consist of *TRIPTYCHON* (TRY), *CAPRICE* (CPC) [34, 50–52], *ENHANCER OF TRY AND CPC 1*, *2* and *3* (ETC1, ETC2, ETC3) and *TRICHOMELESS 1* and *2* (TCL1, TCL2) [50, 53–59]. These negative regulators lack a transcriptional activation domain and seem to interfere with the function of the transcriptionally active MBW-complex through a competitive binding mechanism where they compete with the positive regulators for binding to GL3 [60]. Furthermore, in their absence, phenotypes show an increase in the number of trichomes or the formation of clusters where trichomes are closely packed together [34, 50]. Downstream of the interactions between activators and inhibitors is the gene *GLABRA2* (GL2), which is assumed to be responsible for translating the cues from the patterning genes into specific cell fate [61–63].

As mentioned before, the MBW complex plays a central role in the regulation of trichome patterning. Its function has been studied via yeast-two-hybrid, yeast-three-hybrid, promoter activity assays and LUMIER experiments [44, 64]. It has been found that the assumption of a trimeric complex is most likely too simple and that the actual mechanism relies on differential dimer formation as well as higher order complexes [64]. In Chapter 4, this binding behaviour is quantified and through protein binding models the different modes of competitive binding behaviour have been tested and used in making predictions about the relevant interactions and expected modes of function in terms of regulating cell fate. The MBW complex and its regulations are also

found to play a role in other plant developmental processes [43, 48, 65–68], making it a suitable system for studying general features of plant development. Especially in the case of root hairs, there is a significant overlap in the involved genes [8]; the system of root hairs and leaf hairs is almost completely analogous, yet with a few crucial distinctions. In the case of root hairs, there is a strong dependence on the signals coming from the underlying layer of cortex cells [69, 70]. There are also some functions that show the opposite effect, e.g. expression of GL2 on the leaf leads to trichomes, whereas GL2 expression in the root leads to non-hair cells [8]. Given the genetic and functional similarities between root hairs and trichomes, some assumptions made for the root system are taken into account for trichome patterning as well. One important assumption that is based on evidence found in the roots is the cell-to-cell movement of the inhibitor CPC [71]. These assumptions on mobility are of particular relevance for the patterning models as evident from the derived criteria for pattern formation described in the previous section.

The analysis of trichomes is not only interesting from a developmental perspective, but, given the finding of an evolutionary conserved gene cassette, also from the point of view of evolutional and functional diversification [72–74]. In addition to the model plant *Arabidopsis thaliana*, trichomes have been found in a variety of species [75], among which some hold interest for crop industry, e.g. trichomes in cotton for the production of cotton fibre [76], and the glandular trichomes in *Artemisia annua* that produce Artemisinin [77], the most well-known and potent medicine for malaria. As might be expected, the regulatory mechanism of trichomes varies between divergent plant species, for example, overexpression of *Arabidopsis* GL1 in tobacco does not affect the phenotype [78]. In order to discover where such differences come from it is helpful to approach this by studying species that are closely related enough to find similarities yet distant enough to see functional variation [79, 80]. In Chapters 6 and 7, we will consider a comparison between *Arabidopsis thaliana*, *Arabis alpina* and *Cardamine hirsuta*.

1.6 Mathematical modelling of trichome patterning

The pattern formation of trichomes has been modelled in the form of adaptations of the Gierer-Meinhardt models [38–42]. The development of these models has been guided by the attempt to reproduce certain phenotypical, biochemical, and molecular observations of mutant phenotypes with the goal to gain insight into the underlying regulatory mechanisms. These modelling

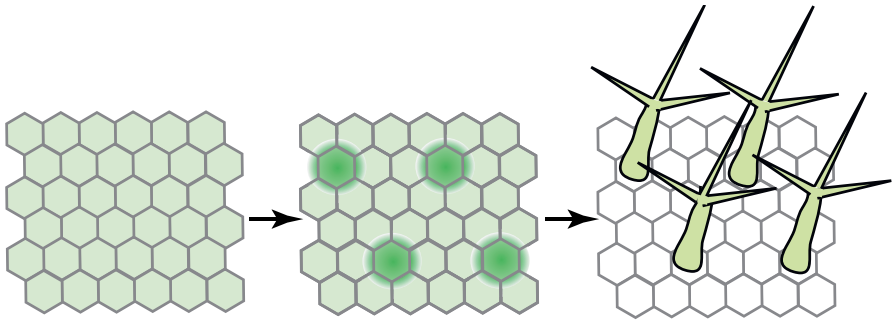


Figure 1.2: **Schematic representation of a simulation of trichome patterning and the grid.** In the simulations, cells are assumed to be organized on a hexagonal grid. Starting from homogeneous conditions (left-most figure), the system of equations is numerically integrated until steady state is reached (middle figure). Based on the final concentrations, trichomes are identified (right-most figure).

approaches allowed us to understand observations which were not immediately obvious, for example the apparent paradoxical behaviour of TTG1, which showed aspects of both inhibitor and activator behaviour in its mutant phenotype [40]. Experimental observations suggested a possible role of depletion of TTG1, which was corroborated by a substrate-depletion model involving TTG1 and GL3 [40].

Depending on the question at hand, models on trichome patterning may be as small as the original two-component system by Meinhardt and Gierer, or require the involvement of more components (the largest model in this thesis consists of 8 species). Each model is a representation of a subset of interactions from the full network described above, which is where both their weakness and strength comes from. What may be hard to track with intuition due to the underlying complexity, can be boiled down to the relevant interactions and components through the clever design of a model. This is true for most biological processes which typically show non-linear behaviour and where our understanding needs to be aided by mathematical modelling [81]. In the case of trichomes, there is a wealth of information available due to the ease of genetic manipulation and thus the wide array of mutant phenotypes available for study.

The models used in this thesis are based on simplifying assumptions around space and time. Typically, the patterns are investigated at steady state on a static (non-growing) domain which is discretized into units that represent cells which have some form of connectivity (Figure 1.2) and are assumed to be internally well-mixed, i.e. intracellular concentrations are assumed to be homogeneous across the cell. On leaves in *Arabidopsis thaliana*, the initiation of trichomes occurs at the base of the leaf, already showing a certain degree of regularity in distances between pairs. Further separation follows from cell division and growth [10, 82]. By ignoring division and growth in the model, it is assumed that the processes included in the model (protein binding, degradation etc.) operate on a much faster time scale than the growth and thus lay out the signals for patterning before processes relevant to cell differentiation are initiated at the appropriate positions. Usually, the relevant regulations are assumed to be on the level of proteins, which means that processes like translation and binding are assumed to be the rate-limiting factors.

Despite simplifying assumptions, our models contain parameters which are (mostly) unknown and in most cases cannot be estimated from the available data. In Chapter 4, some of these parameters are quantified through an approach combining biochemical experiments and mathematical modelling and in Chapter 2 a method is described which was developed to deal with such uncertainty when making predictions. In the following section, this issue of uncertainty is briefly introduced.

1.7 Uncertainty quantification

Modelling approaches typically include sources of uncertainty. Whether it stems from incomplete knowledge, lack of data or generally noisy systems - all of these sources have an effect on the predictions made by the model [83]. Specifically in the case of biological systems the natural variability due to intrinsic noise, e.g., due to stochastic gene expression, leads to uncertainties [84–87]. Therefore, it is crucial to quantify how model and data inaccuracies affect the model predictions.

A method that is well-known and often employed for uncertainty quantification is Monte Carlo (MC) sampling. In the literature, statistical methods based on (pseudo-)random numbers were used a long before the name "Monte Carlo" for these type of approaches was inspired by the Monte Carlo Casino [88]. MC methods are used in physics, chemistry, statistics, computer science,

to name just a few areas [89]. Fundamentally, MC methods rely on pseudo-random sampling of the unknown parameters in order to construct a set of input realisations. For each of these realisations there exists a corresponding solution of the model; all these solutions together give the sample solution set. Using this set one can estimate certain statistics of the model output, for example the expectation $\langle x \rangle$ of the model solution x can be estimated by:

$$\langle x \rangle = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} x^i w_i, \quad \sum_{i=1}^{N} w_i = N, \tag{1.17}$$

where x is the model outcome for input realisation i, N is the number of realisations and w_i is the weight associated with realisation i. In a non-biased sampling approach all $w_i = 1$. MC methods can be applied to any parametrised model, either stochastic or deterministic. MC methods are flexible and robust in that they do not depend on the model form or any regularity in the model solutions. MC methods will always provide some approximation of the quantity of interest. Convergence of the estimation can be assessed based on indicators that are related to the computed solutions. This convergence of MC methods is their main limitation, as the convergence rate depends on the number of realisations by $N^{-1/2}$. Several different sampling methods have been proposed to improve this rate like quasi-Monte Carlo (using low discrepancy sequences like Sobol sequences [90]), Latin hypercube, importance sampling, and variance reduction [91–93]. However, the large number of computations that are needed still remains a bottleneck for a substantial amount of uncertainty problems. Overall, Monte Carlo methods are widely applicable, however, the main disadvantage is the computational power required for a large number of repeated simulations such that the quantity approximated by MC simulations is sufficiently accurate. For some models a single simulation can already be computationally very costly; in these cases MC approaches are not feasible.

An interesting alternative to MC are spectral methods [94, 95], which will be treated in more detail in Chapter 2. In spectral methods, the expectation x is approximated by a series expansion or spectral expansion (SE) in terms of orthogonal functions Ψ of the uncertainty parameter θ:

$$x(\theta) \approx \sum_{k=0}^{N} c_k \Psi_k(\theta), \tag{1.18}$$

where c_k are expansion coefficients that may be time and space dependent, and N is the order of the expansion. This type of expansion resembles the

well-known Fourier expansion used in, e.g., signal analysis. The advantage of this representation is that an approximation of x is obtained for *all* values of θ at the same time. This form allows immediate evaluation of statistics of x, either analytically or through sampling of θ [95].

In a SE, the greatest computational costs comes from calculating the expansion coefficients c_k. In so-called *non-intrusive spectral projection*, the expansion coefficients c_k can be calculated using only straightforward deterministic model solutions and does not require any reformulation of the model. As a result of treating the model as a black-box, such expansions can be easily applied to very complex models (e.g., large or highly non-linear models). Although SE can become computationally expensive if the model itself is expensive to solve (in particular when a higher order of expansion is needed or when facing high dimensional problems [83, 95]), there are methods available to deal with this hurdle by reducing the complexity of the non-intrusive methods, such as adaptive sparse grids [96]. Spectral methods are increasingly popular in the field of uncertainty quantification and have proven very effective for physical and mechanical models (e.g. fluid dynamics) [94]. For these models SE outperform MC methods in terms of computational efficiency. For biological models spectral methods are still uncommon [97], however, given their success in applications that share the same mathematical principles and fundamentally similar sources of uncertainty, it seems a promising approach for biological problems as well; some examples of this will be treated in Chapter 2.

1.8 Thesis objectives and outline

In this thesis, biological experiments are combined with mathematical theory to model the link between genotype and phenotype in the context of trichome patterning. Genetic analysis has revealed a complex and interconnected underlying network which is impossible to explain intuitively, therefore, mathematical models are essential in elucidating this complexity. Both this level of complexity as well as the principles underlying the pattern formation are characteristic for many biological systems, making this an excellent model system. By modelling a variety of mutations, by quantifying fundamental interactions, and through evolutionary comparisons, our aim is to gain insight into the developmental processes involved. The development of the models is guided by experimental results and known biological functions. They are ultimately used not only as a working hypothesis, but also as a guideline for new experiments.

In **chapter 2** the problem of modelling under uncertainty is addressed and an efficient method based on spectral expansions is proposed for uncertainty quantification, specifically in the context of biological models. Through a series of examples, the method is demonstrated and it is shown how it can be adapted to overcome challenges that are typically found in biological systems, such as bifurcations. The method is also applied to the substrate-depletion model in the context of trichome patterning.

In **chapter 3** a mutation in one of the core components in the trichome network is analyzed. This mutation leads to weak *ttg1* alleles that – in contrast to strong alleles, which are devoid of trichomes – show clusters. The proteins in the weak *ttg1* alleles can not bind to GL3 and do not show depletion. In these mutants the formation of trichomes is still ongoing which suggests that parts of the network are functional, albeit aberrant. In addition, the pattern of trichomes appears to show a higher degree of irregularity than seen in wild-type. Aspects of the mutant pattern are quantified and, together with biochemical and molecular observations, are used to formulate a model that includes the essential features necessary to capture the mutant phenotype.

In **chapter 4** the interactions between proteins that are part of the MBW-complex are quantified. This MBW-complex plays a central role in trichome patterning. The formation of this complex is not as straight-forward as initially thought, as follows from previously published data, as well as the data shown in this chapter. Rather than forming a trimeric complex, there appear to be mechanisms at play that allow for higher-order complexes, as well as bilateral influences on distinct binding sites. Using experiments involving only two proteins in isolation as well as competition experiments, binding constants are estimated and used to make predictions on the composition of complexes found in plants and how this might impact and regulate trichome patterning.

In **chapter 5** a modelling approach is used to aid the elucidation of a previously unexplained phenotype, namely the double mutant *trycpc*. In this mutant, both the inhibitors TRY and CPC are non-functional, resulting in a highly-clustering phenotype. The single mutants exhibit either a high trichome density (*cpc*), or a low density in combination with small clusters (*try*). It was previously assumed that both the single and double mutants could be explained by a difference in range of effect of the inhibitors, where CPC was thought to repress trichome fate on a long range and TRY on a short range; the

double mutant was assumed to be the result of lacking both long-range and short-range effects. However, these simple assumptions alone are not enough to reproduce the observations of all three mutant phenotypes using any of the existing models. In this chapter, a new model is developed in order to determine which mechanisms can explain these kind of phenotypes and it is shown that the cause lies not only in the difference between mobility rates of inhibitors, but rather in a combination of effects that involve all other components in the derived network.

In **chapter 6** the trichome patterns of two related species, *Arabidopsis thaliana* and *Arabis alpina* are compared. Orthologues of the relevant genes in *Arabidopsis* were identified for *Arabis* through synteny and sequence similarity and mutants phenotypes were compared between the species. While in most cases the phenotypical defects were analogous, there were also striking differences. In one of these differences, the same mutation leads to an opposite phenotype. Guided by qPCR data, a mathematical model was developed that could help explain such non-intuitive behaviour by different ratios of the relevant proteins.

In **chapter 7** the expression levels of a large number of trichome patterning genes is compared between three closely related species: *Arabidopsis thaliana*, *Arabis Alpina* and *Cardamine hirsuta*. This data is used to fit a model consisting of the core components of trichome patterning and the resulting parameter distributions are compared, revealing how the same network of genes can be adapted to achieve the differences found between evolutionarily distinct species. Furthermore, a sensitivity analysis for each of the three species is used to make predictions on how each of the species show different forms of robustness and what kind of mutant phenotypes might show distinctive differences. Taken together, in this chapter we provide promising avenues of investigation for future mutant screening with the goal of elucidating evolutionary functional divergence.

Finally, in **chapter 8** an overview is given of all the preceding modelling work and the most important conclusions obtained. This chapter reflects on whether the thesis objectives were reached and includes a discussion on the relevance of trichome patterning in a more general context, as well as points of critique and issues that remain unresolved thus far.

Chapter 2

Spectral methods for prediction uncertainty quantification in Systems Biology

Anna Deneer, Jaap Molenaar, Christian Fleck

2.1 Abstract

Uncertainty is ubiquitous in biological systems. These uncertainties can be the result of lack of knowledge or due to a lack of appropriate data. Additionally, the natural variability of biological systems caused by intrinsic noise, e.g. in stochastic gene expression, leads to uncertainties. With the help of numerical simulations the impact of these uncertainties on the model predictions can be assessed, i.e. the impact of the propagation of uncertainty in model parameters on the model response can be quantified. Taking this into account is crucial when the models are used for experimental design, optimization, or decision-making, as model uncertainty can have a significant effect on the accuracy of model predictions. We focus here on spectral methods to quantify prediction uncertainty based on a probabilistic framework. Such methods have a basis in, e.g., computational mathematics, engineering, physics, and fluid dynamics, and, to a lesser extent, systems biology. In this chapter, we highlight the advantages these methods can have for modelling purposes in systems biology and do so by providing a novel and intuitive scheme. By applying the scheme to an array of examples we show its power, especially in challenging situations where slow converge due to high-dimensionality, bifurcations, and spatial discontinuities play a role.

2.2 Introduction

Every mathematical model in Systems Biology is subject to uncertainty and incomplete knowledge [98–101]. This can be in the form of unknown model

structure, unknown model parameters and imperfect experimental data. Characterizing and quantifying these sources is crucial, as the uncertainty can translate into inaccuracies in the model predictions. Information about the quality of model predictions is vital when applied as support for decision-making or optimization routines such as experimental design and parameter estimation [102]. The aim of uncertainty quantification (UQ) is to determine the likeliness of certain outcomes, given that some aspects of the system under study are not (exactly) known.

Generally, uncertainty is distinguished into two classes [94, 95, 103]. The first class is so-called aleatoric uncertainty. Aleatoric uncertainty stems from the intrinsic variability found in the system under consideration, for this reason it is also referred to as statistical uncertainty. For example, in the case of parameter estimation, this uncertainty is related to the fact that parameters may essentially vary over the system components (e.g., cells) [85], so that for the system as a whole only a distribution of parameter values can be estimated, and not one precise value per parameter.

In contrast, the second class of uncertainty, termed epistemic (or systemic) uncertainty, is caused by a lack of information [94, 95, 103]. For example, in the case of parameter estimation, this may be caused by imperfect data sets that contain noisy, incoherent, or missing data points [83]. In such cases, the uncertainty could be reduced by performing extra experiments.

In biological systems both types of uncertainty are typically present [98]. In terms of modelling, both are usually dealt with by employing a probabilistic framework [104]. In the case of epistemic uncertainty, model parameters can be represented according to a probability density function (PDF) [95, 101]. The choice of the type of PDF and the corresponding distribution of parameters is usually based on previous knowledge. For example, the case of a completely unknown parameter could be described by a uniform distribution on a broad (positive) interval. In other cases a parameter could be known to follow a normal or lognormal distribution with known mean and variance, established in previously performed experiments [105].

Among the field of UQ, Monte Carlo (MC) methods are most commonly used [88, 89]. In a MC approach the parameter PDFs are sampled and model responses for each sample recorded, thus providing a distribution of model outcomes and an indication of the uncertainty therein (e.g. by analyzing the distribution moments). These methods are simple in their implementation

and are widely applicable. However, for models that have a large number of parameters or are computationally expensive, these MC procedures are often not feasible [94, 95].

As an alternative to MC, meta-modelling techniques are frequently adopted to deal with models that would otherwise be intractable. Support vector machines [106], artificial neural networks [107] and Bayesian networks [108] are examples of surrogate- or meta-modelling techniques used in Systems Biology. In this chapter we focus on stochastic spectral methods, in particular polynomial chaos expansion (PCE), an approach that is widely used in engineering systems [94, 103] and to a lesser extent in biological systems [97, 109–111] for UQ purposes. The aim of these approaches is to represent the model response as a series expansion. The advantage of this representation is that an approximation of the model response is obtained for *all* values of the uncertainty parameters. This form allows immediate evaluation of statistics of the model outcome, either analytically or through sampling of the stochastic parameters, which can be done significantly faster than through MC methods for models that are problematic and computationally expensive [112].

This advantage comes at the cost of the need to calculate expansion coefficients. For this, two classes of spectral methods are in use. In the first class the governing equations of the model are reformulated such that each variable is represented by a spectral expansion. This results in a system of differential equations for the expansion coefficients and is known as intrusive spectral projection [95]. In the second class, consisting of so-called non-intrusive spectral projection approaches and followed in this paper, the expansion coefficients are determined using the model without changing the original model equations [113]. The advantage of this non-intrusive approach is that it requires only straightforward deterministic model evaluations and does not involve any reformulation of the model. It is particularly attractive in case of very large models, since then intrusive methods would become too laborious.

In the past, PCE has been shown to converge very slowly or not at all for models involving non-smooth functions [114, 115]. This is indeed a critical challenge for biological models, which often show complex, non-linear behaviour such as bifurcations and spatial discontinuities. In this chapter we provide a scheme for non-intrusive spectral projection that may overcome these problems. It is easy to implement and we show its power through applying it to a number of biological models. The examples in this chapter each have a specific problem to be overcome.

2.3 Methods

2.3.1 Spectral expansion

Let us consider a model $\mathbf{\Omega}$ that depends on a vector of stochastic input parameters $\boldsymbol{\theta}$. The model response Y can be any chosen quantity, e.g. the concentration of one of the model components or a function thereof. The uncertainty parameters θ_l are assumed to be independently distributed, each with PDF $P_l(\theta_l)$. So,

$$Y = \mathbf{\Omega}(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \sim P(\boldsymbol{\theta}), \tag{2.1}$$

where $P(\boldsymbol{\theta})$ is the joint probability density function (PDF) for all U uncertainty parameters: $P(\boldsymbol{\theta}) = \prod_{l=1}^{U} P_l(\theta_l)$. For reasons of clarity, we restrict here the explanation to Y and θ being scalar functions. In the section Practical Aspects we show how to deal with more than one uncertainty parameter. Note that for correlated parameters the PDF would follow from the corresponding multivariate distribution.

The underlying model could be of any type, e.g., an ODE, a PDE, an algebraic, or a statistical model. This implies that Y may also depend on time and space. The challenge is to analyse the behaviour of Y as a function of $\boldsymbol{\theta}$. In cases where the numerical evaluation of the underlying model takes a considerable amount of computational time this tends to obstruct any form of comprehensive analysis. In this paper we present the use of a method that aims at making this tractable. The idea is to replace the original model by a meta-model, which is achieved by representing the output Y in terms of an expansion. This meta-model can be constructed such that it represents the underlying model to a high degree of accuracy, with the advantage of being much faster to evaluate than the original model.

This meta-model can be used to determine the distribution of the model response Y or reconstruct the function accurately at given points in the parameter space. In the spectral approach the model response is represented by

$$Y = \sum_{n=0}^{\infty} c_n \phi_n(\theta), \tag{2.2}$$

where c_n are the expansion coefficients (which can be time and/or space dependent) and ϕ_n are functions that are orthonormal with respect to the distribution of the uncertainty parameters as weight functions for an inner product,

as dealt with in S1 Appendix A. For example, suitable basis functions for uniformly distributed parameters are Legendre polynomials, while for normally distributed parameters Hermite polynomials qualify. For any practical purpose the expansion needs to be truncated to a certain degree:

$$Y^s = \sum_{n=0}^{N} c_n \phi_n(\theta), \tag{2.3}$$

where N is the truncation degree. So, for such a meta-model Y^s N expansion coefficients have to be calculated. The advantage of (2.3) is that the statistics of the model response Y can be evaluated very fast, either analytically or through sampling of the parameters θ. The main computational cost of the expansion comes from the computation of the coefficients c_n. Below, we provide an easy-to-implement scheme for the calculation of these coefficients.

We use a non-intrusive approach to the spectral expansion, i.e., we treat the model equations as a black box, not requiring any tailoring to the equations to include the parameter uncertainties. The most commonly used method for non-intrusively determining the coefficients is through Gaussian quadrature schemes [95]. Here, we propose an alternative scheme. It is applicable to any set of orthonormal functions, allowing the flexibility to tackle different modelling challenges. A key feature in the scheme is the introduction of the symmetric matrix:

$$\hat{B}_{n,m} = \int \phi_n(\theta)\, \theta\, \phi_m(\theta)\, P(\theta)\, d\theta. \tag{2.4}$$

Its eigenvalues $\lambda^{(l)}, l = 1, 2, \ldots$, are real and its eigenvectors $u^{(l)}$ orthonormal. In the appendix we show that these eigenvalues and eigenvectors can be used to derive an expression for the coefficients c_n. After substitution, (2.3) then reads as

$$Y^s(\theta) = \sum_{l=0}^{N} Y(\lambda^{(l)})\, u_1^{(l)}\, \psi_l^s(\theta), \tag{2.5}$$

where

$$\psi_l^s(\theta) \equiv \sum_{n=0}^{N} u_{n+1}^{(l)}\, \phi_n(\theta). \tag{2.6}$$

The striking point here is that this expansion requires to evaluate the model only $N+1$ times, namely for each of the eigenvalues $\lambda^{(l)}, l = 0, \ldots, N$. Note that all terms in the expansion that do not depend on the uncertainty parameter

θ can be calculated in advance, so once and for all. This saves computation time for any future application. For models that take a long time to evaluate the use of (2.5) is a very fast alternative, compared to e.g. a Monte-Carlo approach. Note that expansion (2.5) is only exact in the limit $N \to \infty$. Taking a finite value for N introduces an inaccuracy. Therefore, N must be chosen with care and it is often not obvious beforehand which value of N will give reliable results. We will showcase in the Examples underneath how one may deal with the choice of N. In the subsection Segmentation in the next section, we propose an adjusted scheme to deal with cases where a high degree might result in infeasible computational times.

2.3.2 Practical aspects

Here, we treat some specific aspects of the method presented above.

PDFs and basis functions

We already mentioned Legendre and Hermite polynomials as typical basis functions for PCE. Legendre polynomials are defined over $[-1,1]$ and are orthogonal with respect to the uniform distribution. The Hermite polynomials are defined over $\mathbb{R}$ and are orthogonal with respect to the Gaussian distribution. Both polynomials can be normalized with appropriate prefactors to ensure orthonormality. These two families of polynomials are most commonly used to represent biological parameters. The uniform distribution is typically applied in uninformed cases and the lognormal distribution in cases where there is prior information available on a parameter. In practice, sometimes other classical orthogonal polynomial families are appropriate, e.g., Laguerre polynomials for Gamma distributions. Also, non-polynomial functions may be applied, such as spherical harmonics. The approach presented here can be used for any set of orthonormal functions.

The Legendre and Hermite polynomials are defined for standard uniform $\mathcal{U}(-1,1)$ and standard normal $\mathcal{N}(0,1)$ variables, respectively. In practice, the biological parameters are often not restricted to the corresponding intervals. In these cases we have to apply an isoprobabilistic transformation. For example, to obtain a normally distributed random variable k with mean μ and variance σ, so $k \sim \mathcal{N}(\mu,\sigma)$, from $\theta \sim \mathcal{N}(0,1)$, we need the transform

$$k = \mu + \sigma\theta, \qquad \theta \sim \mathcal{N}(0,1). \tag{2.7}$$

To obtain a uniformly distributed random variable k on the interval $[a,b]$, so $k \sim \mathcal{U}(a,b)$, from $\theta \sim \mathcal{U}(-1,1)$, we need the transform

$$k = (b+a)/2 + (b-a)\theta/2, \quad \theta \sim \mathcal{U}(-1,1). \tag{2.8}$$

A lognormally distributed random variable $k \sim \text{Lognormal}(\mu,\sigma)$ is obtained from $\theta \sim \mathcal{N}(0,1)$ via the transformation

$$k = \mu \exp\left[\alpha\theta - \frac{\alpha^2}{2}\right], \qquad \theta \sim \mathcal{N}(0,1), \tag{2.9}$$

where $\alpha = \sqrt{\ln\left(1+\frac{\sigma^2}{\mu^2}\right)}$.

Multiple uncertainty parameters

Typically, biological models involve more than one random parameter, which means that the PCE basis $\{\phi_n(\boldsymbol{\theta}), n \in \mathbb{N}^M\}$ is multivariate. Extending (2.5) to the M-dimensional case is straightforward:

$$Y^s(\theta_1,\ldots,\theta_M) = \sum_{l_1=1}^{N} \ldots \sum_{l_M=1}^{N} Y(\lambda^{(l_1)},\ldots,\lambda^{(l_M)}) u_1^{(l_1)} \ldots u_1^{(l_M)} \psi_{l_1}^s(\theta_1) \ldots \psi_{l_M}^s(\theta_M). \tag{2.10}$$

Similar to MC approaches, PCE suffers from the curse of dimensionality [103, 116]. Note from (2.10) that the number of times the model has to be evaluated scales as N^M, where N is the expansion order and M the number of uncertainty parameters.

Segmentation

In cases where the model is computationally expensive, it is advantageous to keep the expansion order relatively low to ensure feasible computation times. However, models that show complex response surfaces (e.g., bifurcations) will in a straightforward approach require high order expansions to capture the complexity. This is problematic since it not only requires to evaluate the model often, but also leads to time consuming summations in (2.10). To overcome these problems, we propose a scheme that segments the parameter intervals into subintervals. Within each of these segments we then perform a separate expansion. In this approach we have to deal with a trade-off: the number of expansions is multiplied, but per expansion we have a (much)

lower order of expansion. Below we argue why the second positive aspect greatly counterbalances the first negative aspect.
To determine the segments we define a scaling function g_m with $M \in \mathbb{N}$ and $L \in \mathbb{R}$ by:

$$g_m : [-L, L] \to \mathcal{I}_m = \left[\frac{(2m-1)L}{2M+1}, \frac{(2m+1)L}{2M+1} \right], m \in \{-M, M\} \subset \mathbb{Z}, \tag{2.11}$$
$$\theta \mapsto g_m(\theta) = \frac{2mL}{2M+1} + \frac{\theta}{2M+1}.$$

This scaling function g_m divides the interval $[-L, L]$ into $2M+1$ segments $\mathcal{I}_m$ of equal length. Of course, L must be larger than or equal to any value of θ. For example, consider the case $L = 1$. Then, the whole interval is $[-1, 1]$. For a segmentation granularity of $M = 1$, this interval is divided into three subintervals: $\mathcal{I}_{-1} = [-1, -0.33]$, $\mathcal{I}_0 = [-0.33, 0.33]$ and $\mathcal{I}_{+1} = [0.33, 1]$. The expansion of Y on any subinterval $\mathcal{I}_m$ is given by:

$$Y(g_m(\theta)) = \sum_{l}^{N} Y(g_m(\lambda^{(l)})) u_1^{(l)} \psi_l(\theta). \tag{2.12}$$

Upon a variable transformation $y = g_m(\theta)$, Eq. 2.12 becomes:

$$Y(y) = \sum_{l}^{N} Y(g_m(\lambda^{(l)})) u_1^{(l)} \psi_l(g_m^{-1}(y)) : y \in \mathcal{I}_m. \tag{2.13}$$

After segmentation, the expansion on the interval $[-L, L]$ as a whole is a superposition of the expansions on the subintervals:

$$Y(y) = \sum_{m}^{M} \sum_{l}^{N} Y(g_m(\lambda^{(l)})) u_1^{(l)} \chi_m(y) \psi_l(g_m^{-1}(y)) : y \in [-L, L], \tag{2.14}$$

where $\chi_m(y)$ is an indicator function for selecting the correct segment:

$$\chi_m(y) = \begin{cases} 1 : y \in \mathcal{I}_m, \\ 0 : y \notin \mathcal{I}_m. \end{cases} \tag{2.15}$$

We can also define an index function to select $m^* \in \{-M, M\}$ for which $\chi_m(y) = 1$:

$$m^* = z(y) = \left\lfloor \frac{(2M+1)y}{2L} + \frac{1}{2} \right\rfloor \text{for} |y| \leq L. \tag{2.16}$$

Using both index functions we can finally write the segmented reconstruction as:

$$Y(y) = \sum_{l}^{N} Y(g_{z(y)}(\lambda^{(l)}))u_1^{(l)}\psi_l(g^{-1}_{z(y)}(y)) : y \in [-L, L]. \tag{2.17}$$

Segmentation allows the use of lower order polynomials while maintaining the same accuracy (assuming a sensible choice for N and M) as a non-segmented higher order expansion. The number of model solutions required now scales as $(2M+1)^K N^K$, where K is the number of uncertainty parameters, M the segmentation granularity, and N the expansion order. The reduced accuracy by using a lower order expansion is compensated for by evaluating the model more often, as a result of zooming in. Expanding up to the N-th order for p uncertainty dimensions requires solving the system N^p times. Reconstruction requires the summation of N^{2p} terms. Therefore, it is advantageous to keep N as low as possible. Normally, the reconstruction error is large for low N, but this is mitigated by segmentation. Segmentation requires to evaluate the system $(2M+1)^p N^p$ times, but due to segmentation N can be taken much smaller.

To illustrate this with an example, we take a system with 2 species of interest and 5 uncertainty parameters θ_i. The expansion order is taken as $N = 8$. This implies summing over $2 \times 8^{10} = 2,147,483,648$ terms per time point and per parameter set $(\theta_1, ..., \theta_5)$. In the case of segmented expansion, we can choose a lower N, for example $N = 3$ with a segmentation granularity of $M = 1$. The number of terms to be summed over is $2 \times 3^{10} = 118,098$. This is dramatically more efficient and stems from the fact that one only has to determine the segment in which the parameter set $(\theta_1, .., \theta_5)$ falls and choose the corresponding expansion coefficients. In the Results section we will present Example II in which segmentation indeed proves to be very beneficial.

Haar wavelet expansion

Traditional PCE methods are known to have difficulties with capturing discontinuous behaviour [114, 115]. Spectral convergence is only observed when solutions are sufficiently regular and continuous. Just like Fourier expansions, PCE suffers from Gibbs phenonema at discontinuities, resulting in slow convergence [94]. Haar wavelets have been suggested to overcome these difficulties [83, 115]. In contrast to global basis functions like the aforementioned polynomial systems, wavelet representations lead to localized decompositions, resulting in increased robustness at the cost of a slower convergence rate [83, 94]. Here, we discuss that Haar wavelets can be easily incorporated

in the framework presented above and in Example IV underneath we show how they can be applied in practice.
As mother wavelet we take

$$\psi^W(y) = \begin{cases} 1 & \text{for } 0 \le y < \frac{1}{2} \\ -1 & \text{for } \frac{1}{2} \le y < 1 \\ 0 & \text{otherwise} \end{cases}. \tag{2.18}$$

By introducing a scaling factor j and a sliding factor k, we may construct the wavelet family

$$\psi^W_{j,k}(y) = 2^{j/2}\psi^W(2^j y - k), \qquad j = 0, 1 \ldots; \quad k = 0, \ldots, 2^j - 1. \tag{2.19}$$

Given the uncertainty parameter θ with its cumulative distribution function $F(\theta)$, we define the basis functions as

$$W_{j,k}(\theta) \sim \psi^W_{j,k}(F(\theta)). \tag{2.20}$$

By concatenating the indices j and k into one index $i \equiv 2^j + k$, we may expand the meta-model Y similarly as we did in (2.2):

$$Y(\theta) = \sum_{n=0}^{\infty} c_n W_i(\theta). \tag{2.21}$$

Sensitivity analysis

In sensitivity analysis one quantifies the effects of changes in the parameters on the variability of the model response. Here, we show how our PCE approach allows for sensitivity analysis in an elegant way. In the case of local sensitivity analysis, small parameter variations around a certain point in parameter space are used to determine the effect on the model output [117]. This sensitivity is estimated via calculation of the partial derivatives of model output with respect to parameters, evaluated in that point [118]. Alternatively, global sensitivity approaches do not specify a specific point in parameter space [119]. For example, Sobol indices are a popular sensitivity measure as they provide a measure of global sensitivity and accurate information for most models [90]. Sobol indices are based on the decomposition of the variance of the output Y as a function of the contribution of the parameters (and possibly their combination), also called the ANalysis of VAriance, or ANOVA [90]. Thanks to the orthonormality of basis functions in PCE, Sobol indices

can be determined analytically from the coefficients of the PCE [120, 121] So, once these coefficients are known, one gets the Sobol indices nearly for free. Given the PCE expansion of output Y, the total variance of the model output is given by

$$\hat{D} = \sum_{i \in I - \{0\}} c_i^2, \tag{2.22}$$

where I is the multi-index set of all variables and c_i the expansions coefficients. The 0^{th} coefficient is not included as this is a constant. The partial variance is given by

$$D_{\theta_i} = \sum_{i \in I_{\theta_i}} c_i^2, \tag{2.23}$$

where I_{θ_i} is the multi-index set of parameter θ_i, i.e. where the i^{th} term in the multi-index is larger than 0. The Sobol indices are then given by

$$S_{\theta_i} = \frac{D_{\theta_i}}{\hat{D}}, \tag{2.24}$$

In this way the relative contribution of parameter θ_i to the variance of the output is easily calculated.

2.3.3 Summary of implementation

In this section we provide an overview of the steps needed to arrive at a meta-model using PCE:

1. Determine which of the model parameters may show stochastic behaviour and decide upon an appropriate PDF for such parameters.
2. Choose a truncation degree N.
3. Based on the PDF in the previous steps, calculate the appropriate basis functions $\phi_n(\theta), n = 0, 1, 2 \ldots, N$.
4. Determine the $N \times N$ matrix $\hat{B}$ as defined in 2.4. For example, for Legendre polynomials $\hat{B}$ reads as

$$\hat{B}_{n,m} = \frac{n}{\sqrt{2n+1}\sqrt{2m+1}} \delta_{nm+1} + \frac{m}{\sqrt{2n+1}\sqrt{2m+1}} \delta_{nm-1}, \tag{2.25}$$

   where $n, m = 0, 1, 2 \ldots, N$.

   For Hermitian polynomials $\hat{B}$ reads as:

$$\hat{B}_{n,m} = \sqrt{n}\delta_{nm+1} + \sqrt{m}\delta_{nm-1}. \tag{2.26}$$

5. Calculate the eigenvalues $\lambda^{(l)}, l = 1, 2, \ldots, N$ and orthonormal eigenvectors $u^{(l)}$.

6. Calculate $\psi_l^s(\theta) = \sum_{n=0}^{N} u_{n+1}^{(l)} \phi_n(\theta)$.

7. Calculate $Y(\lambda^{(l)}), l = 1, 2, \ldots, N$ by evaluating the model N times.

8. Arrive at the metamodel $Y^s(\theta) = \sum_{l=0}^{N} Y(\lambda^{(l)}) u_1^{(l)} \psi_l^s(\theta)$.

9. Eventually, apply post-processing through, e.g., sensitivity analysis.

2.4 Results

To test the performance of the present PCE approach in biological simulations, we have chosen four typical examples. Through these examples, we show how to deal with several challenges usually encountered in systems biology.

The first example has only one uncertainty parameter. Its simplicity allows comparison between the results of our approach with an exact solution.

The second example concerns a biochemical reaction network and is higher dimensional, i.e., it contains more than one uncertainty parameters. We use it to highlight the advantages of segmentation.

The third example is the glycolytic oscillator, which shows bifurcations, i.e., different dynamic behaviour for different parameter sets [122]. We use it to demonstrate the power of global sensitivity analysis, which in the PCE framework can be achieved without significant additional computational costs once the PCE coefficients have been calculated. In addition, this example allows us to show the use of mixed expansions, since the parameter PDFs follow different distributions. This leads to a combination of different families of basis functions, thus highlighting the flexibility of the PCE approach when applied to varying input uncertainties.

The last two examples have a spatial dimension. First, we consider the Schnakenberg model which is a well-known model of pattern formation and comes with challenges such as shifts from non-patterning to patterning regions [123]. In this example we demonstrate the advantage of using Haar

wavelets over polynomial basis functions for systems with bifurcations. Second, we study a model describing pattern formation in plants, more specifically patterning of the hairs found on top of leaves, so-called trichomes [40]. In this model we show how to adapt the approach such that the computational costs are reduced as much as possible by carefully choosing the quantity of interest, without changing the standard set of steps.

2.4.1 Example I. Exponential decay: comparing performance of PCE to MC and an analytical solution

For this example case, we consider the extremely simple reaction system consisting of one decaying species:

$$A \xrightarrow{k} \emptyset.$$

Its dynamics is described by $A(t) = A_0 \exp^{-kt}$, where A_0 is the initial concentration of A at $t = 0$ and k the rate of decay. We test the PCE method against: *1)* the exact, analytical solution and *2)* the classical Monte Carlo approach.

The quantification of the sources of uncertainty constitute the second step in the analysis. This entails identification of the parameters that are unknown and modelling them in a probabilistic context. In this case, we assume that k is distributed according to a lognormal distribution with known mean and variance, i.e. $k \sim \text{Lognormal}(\mu, \sigma)$, and we choose $\mu = 0.5, \sigma = 0.2$. The PDF for k is shown in Figure 2.1 and the derivation for the exact PDF for A is given in S2: Derivation of the probability density function for the exponential decay model.

Next, we determine how the uncertainty in k propagates through the model and affects concentration $A(t)$. To that end, we expand the function $A(t) = A_0 \exp^{-kt}$ in terms of Hermitian polynomials. Using (2.9), k is transformed into a standard normal variable θ. To arrive at the meta-model $Y^s(\theta)$ we truncate the expansion to a certain order N, as shown in (2.3). Choosing N is not straightforward and will involve some experimentation. In Figure 2.1 we compare results for $N = 5$ to the analytical solution for different k values. This shows that for this expansion order the reconstructed function accurately matches the analytical solution.

In Figure 2.1, we focus on the distributions of $A(t = 1)$ and $A(t = 5)$. This is achieved by sampling the PCE using a large sample set χ of reduced (i.e. standard normally distributed) variables θ, $\chi_{sim} = \{\theta_j, j = 1, \ldots, n_{sim}\}$. The trun-

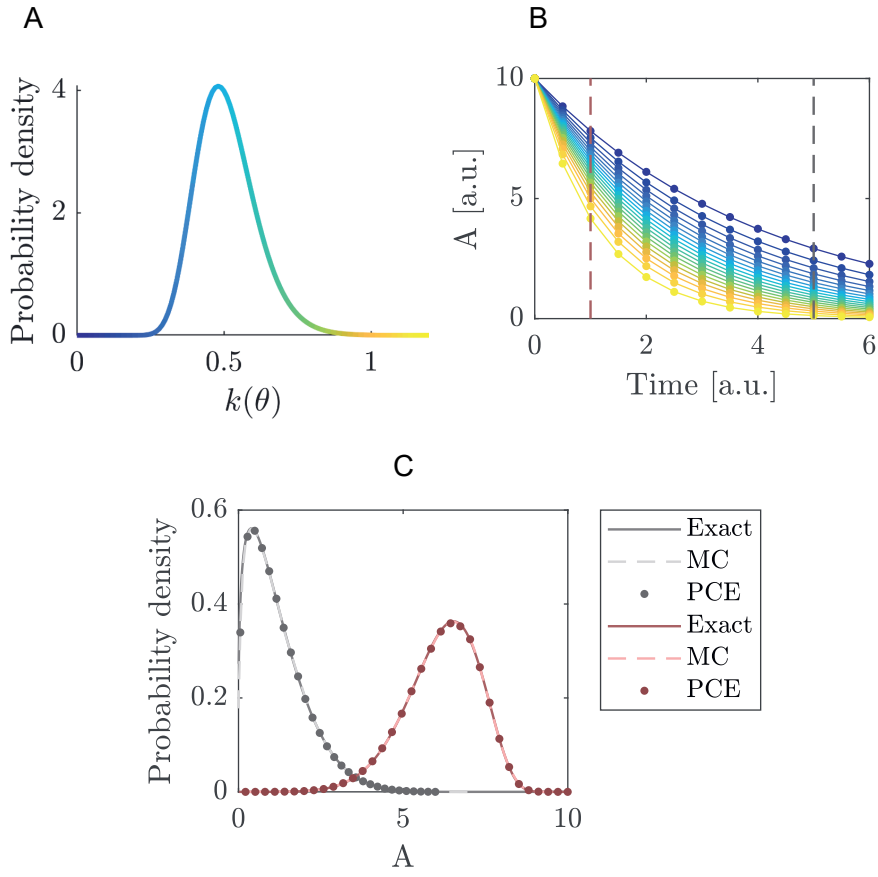


Figure 2.1: **Quantifying the uncertainty propagated by the decay rate in the exponential decay model.** A: Probability density function of the decay rate $k(\theta)$ with $\mu = 0.5$ and $\sigma = 0.2$. The colour gradient corresponds to the value of $k(\theta)$. B: The concentration of A up to $t = 6$ seconds. The solid lines indicate the analytical solution and the dots indicate the reconstruction using PCE with Hermite basis functions and an expansion order $N = 5$. The colour for each of the solutions correspond to the colour of the line in A, which indicates the value of $k(\theta)$ used for each of the depicted solutions. C: The dashed lines in B indicate a cross-section along the model response space at $t = 1$ for the red line, and at $t = 5$ for the grey line, determined through three methods. First, the exact dynamics of the model (solid lines), second through MC sampling using the exact model dynamics (dashed lines) and finally, through MC sampling of the reconstructed function as obtained through PCE (dots).

cated series is then evaluated onto this sample: $Y^s_{sim} = \{\eta_j = \sum_{n=0}^{N} c_n \phi_n(\theta_j), j = 1,\dots,n_{sim}\}$. These PDFs are obtained by kernel smoothing [124] using a sample set with $n_{sim} = 10^6$, drawn from the standard normal distribution with $\mu = 0$ and $\sigma = 1$. The kernel density estimator is given by

$$\hat{f}_Y(y) = \frac{1}{n_{sim}h} \sum_{j=1}^{n_{sim}} K\left(\frac{y-\eta_j}{h}\right), \tag{2.27}$$

with kernel function $K(t) = \frac{1}{\sqrt{2\pi}} \exp^{-t^2/2}$ and bandwidth h, which is determined by Silverman's rule of thumb [125]. Figure 2.1 shows that both MC and PCE perform well in reproducing the exact PDFs.

2.4.2 Example II. Biochemical reaction network: dealing with higher dimensions

In this example we present a simple model with multiple uncertainty parameters. The model describes the dynamics of two proteins x_1 and x_2 which bind together to form a dimer x_3. We consider the following reactions:

$$\emptyset \xrightarrow{k_1} x_1 \tag{2.28}$$

$$x_1 + x_2 \underset{k_3}{\overset{k_2}{\rightleftharpoons}} x_3 \tag{2.29}$$

$$\emptyset \xrightarrow{k_4} x_2 \tag{2.30}$$

$$x_1, x_2, x_3 \xrightarrow{k_5} \emptyset \tag{2.31}$$

In this network, the proteins x_1 and x_2 are produced at rates k_1 and k_4. Proteins x_1 and x_2 reversibly bind to form species x_3, with binding rate k_2 and unbinding rate k_3. All three proteins are degraded at the rate k_5.These interactions are visualized in a reaction scheme in Figure 2.2. The ODEs for this system are

$$\dot{x_1} = k_1 - k_2 x_1 x_2 + k_3 x_3 - k_5 x_1 \tag{2.32}$$

$$\dot{x_2} = k_4 - k_2 x_1 x_2 + k_3 x_3 - k_5 x_2 \tag{2.33}$$

$$\dot{x_3} = k_2 x_1 x_2 - k_3 x_3 - k_5 x_3 \tag{2.34}$$

We use for the parameters $k_1 - k_5$ log-normal distributions and expand the functions $x_1 - x_3$ in terms of Hermitian polynomials. In Figure 2.2 we

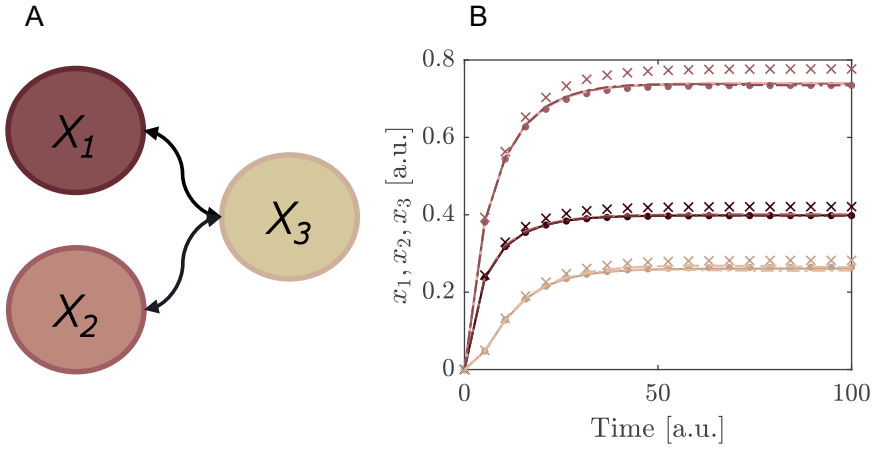


Figure 2.2: **Comparing the segmented expansion with the standard non-segmented expansion.** A: Interaction scheme of the model in Eqs.(2.32)-(2.34). Note that x_1, x_2 and x_3 all have the same degradation rate and x_1, x_2 have production rates in the model but this is not indicated in the scheme. B: Reconstruction of the system by a Hermitian expansion. For the segmented reconstruction we used $N = 2, M = 1$ (crosses) and $N = 3, M = 1$ (dots). For non-segmented expansion the expansion order was $N = 8$ (dashed lines) and $N = 9$, (dash-dotted lines). Note that these lines overlap with the true model solution (solid lines). We used two log-normal distributions with mean and standard deviation $\mu_1 = 0.1, \sigma_1 = 0.1$ for k_1, k_4, k_5 and $\mu_2 = 0.4, \sigma_2 = 0.1$ for k_2, k_3.

compare the results of the segmented expansion with the non-segmented expansion and the exact results. For this comparison we chose the degree of expansion and segmentation granularity such that the same number of model evaluations were required. We found that the subsequent summation to reconstruct the solutions for the differential equations improved by factors of 1000-30,000 when using the segmented expansion. See Table 2.1. As mentioned in the Methods section, this improvements stems from the large reduction in the number of terms to be summed over in the segmented case compared to the non-segmented expansion.

Dividing the parameter intervals into smaller sub-intervals is a relatively straightforward and simple way to circumvent huge computation times. Other, more intricate methods have been developed to tackle models with an even

Table 2.1: **Benchmarking of segmented and non-segmented expansion.**

N	M	N_λ	N_Σ	$t_\Sigma[s]$	Fold-change
2	1	7776	3.07E+03	0.03	
3	1	59049	1.77E+05	0.17	
6	0	7776	1.81E+08	31.96	1065
9	0	59049	1.05E+10	5402	31776

An overview of the number of model evaluations N_λ, the number of summation terms N_Σ and the time in seconds spent on summation (t_Σ), for different orders N of expansion, both segmented $M = 1$ and non-segmented $M = 0$. The last column highlights the speed-up factor when segmentation is used (keeping N_λ constant).

larger amount of parameters [96, 116, 126]. For example, using an adaptive algorithm that is based on classical statistical learning tools can result in a "sparse" PCE, that consists of only the significant coefficients in the expansion, thereby reducing the computational cost. This method has been tested on models of stochastic finite element analysis with up to 21 parameters [96].

2.4.3 Example III. Glycolytic oscillator: mixed input PDFs and global sensitivity

Living cells obtain energy by breaking down sugar in the biochemical process called glycolysis. In yeast cells, this glycolysis was observed to behave in an oscillatory fashion, where the concentration of various intermediates were increasing and decreasing within a period of several minutes [127]. This glycolytic oscillator can be modelled as a two-component system with a negative feedback [122]:

$$\dot{x} = -x + \alpha y + x^2 y \tag{2.35}$$

$$\dot{y} = \beta - \alpha y - x^2 y \tag{2.36}$$

where x and y are the concentrations of ADP (adenosine diphosphate) and F6P (fructose-6-phosphate) and α, β are kinetic parameters. Depending on the values of α and β the system will be in a stable limit cycle or a stable fixed point [122]. In this example, we assume α to be uniformly distributed on the interval $[0.1, 0.5]$ and β to follow a lognormal distribution with $\mu = 0.3$ and

$\sigma = 0.1$. Because both uncertainty parameters come from a different distribution, the expansion will consist of multivariate polynomials $\Psi_{N,M}$ which are tensor products of the univariate polynomials. In this case, Legendre polynomials are used to expand $\alpha(\theta_1)$ and Hermite polynomials for $\beta(\theta_2)$, where $\theta_1 \sim \mathcal{U}(-1,1)$ and $\theta_2 \sim \mathcal{N}(0,1)$. This results in a mixed polynomial for the overall expansion as exemplified with a 3rd order Legendre polynomial $\tilde{L}_3$ and a 3rd order Hermite polynomial $\tilde{H}_3$, giving $\Psi_{3,3} = \tilde{L}_3 \cdot \tilde{H}_3$ (Figure 2.3).

The distributions of the uncertainty parameters were chosen such that they include the bifurcation point from stable limit cycle to the stable fixed point (Figure 2.3). For the purpose of this example we are interested in the concentration of y only and therefore reconstruct this model response using PCE. A good approximation is obtained with a truncation degree of the PCE of $N = 10$. This value is relatively high, due to the bifurcation in the system. However, this case shows that convergence can be reached using PCE despite such challenges. Yet, the computational costs are still very tractable. In the following examples we deal with cases in which still higher order expansions are necessary due to non-smooth bifurcations.

In post-processing we may use the PC coefficients to determine the first order Sobol indices for the parameters α and β at each time point, providing a representation of the global sensitivity based on variance decomposition. The Sobol indices are readily available from the PC coefficients (see Methods). They have the advantage of being global measures of sensitivity. In Figure 2.3 we show the first order Sobol indices given in (2.23) for the first and second random variable. They indicate the contribution to the total output variance of either θ_1 or θ_2 individually. Higher-order terms would give an indication of interaction effects between θ_1 and θ_2, which are also readily available from the PCE coefficients but are not considered here for brevity.

2.4.4 Example IV. Schnakenberg model: dealing with spatial discontinuities

In this example we introduce a spatial component. We consider the Schnackenberg model, which is one of the simplest, but yet realistic two-species system that can produce periodic solutions and therefore has become a prototype for reaction diffusion systems. The Schnakenberg model consists of the

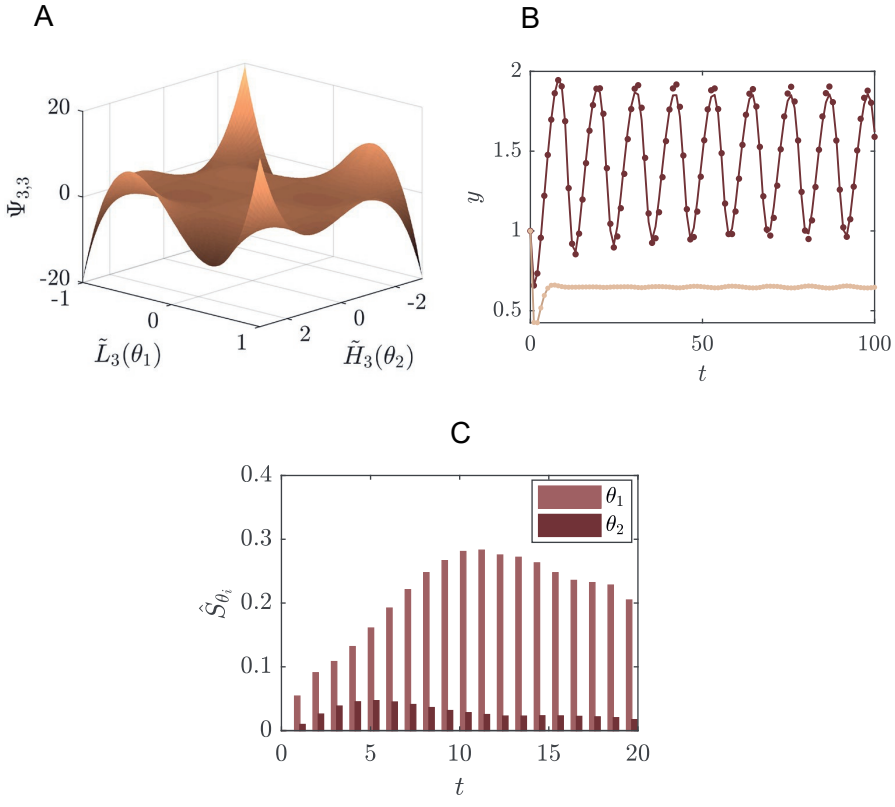


Figure 2.3: **Example of a system of glycolysis as defined in (2.35)-(2.36) where PCE is performed for the concentration of species y and for two uncertainty parameters.** A: Example of a multivariate polynomial, consisting of the tensor product between 3rd Legendre polynomial of the first random variable $\tilde{L}_3(\theta_1)$ and the 3rd Hermite polynomial of the second random variable $\tilde{H}_3(\theta_2)$. B: Solutions of concentration of fructose-6-phosphate (y) in the glycolytic oscillator model for two different points in the parameter space, obtained by solving the ODEs (lines) and reconstructing via PCE (dots). $\alpha = 0.1, \beta = 0.46$ produces an oscillation (dark coloured) whereas $\alpha = 0.5, \beta = 0.46$ gives a stable fixed point (light coloured). We used $N = 10$ as expansion order for the Legendre and Hermite polynomials. C: The first order Sobol sensitivity index $\hat{S}_{\theta_i}$ for the two random variables in the glycolytic oscillator model for the first 20 time points.

following (dimensionless) equations [123]

$$\dot{u} = \nabla^2 u + \gamma(\alpha - u + u^2 v), \tag{2.37}$$

$$\dot{v} = d\nabla^2 v + \gamma(\beta - u^2 v). \tag{2.38}$$

α, β are reaction rates, γ a scale parameter and d the ratio of diffusion constants between the species u and v. The species u is auto-catalytically produced by the $u^2 v$ term in (2.37), whereby species v is consumed. There are certain combinations of the parameters α, β, γ and d for which the system will exhibit a stable pattern [128]; this region of parameter space is called the Turing space (TS). For the purpose of this example we limit the number of uncertainty parameters to one: the parameter α, fixing the other parameters at $\beta = 1, \gamma = 5$ and $d = 20$. We assume α to be distributed as $\alpha \sim \mathcal{U}(0.001, 0.45)$ and determine the TS for a range of α (Figure 2.4) using linear stability analysis (for details see [123, 128]). To that end, the model is simulated on a 1D grid of 20 cells. We focus on the concentration of species v at steady state and consider an expansion by both Legendre polynomials and Haar wavelets.

Polynomial chaos expansion is known for being inaccurate in regions that contain discontinuities [94, 114, 115]. In this example, the lack of convergence in PCE can be seen along the boundary of the patterning space (TS) in Figure 2.4, where the expansion by Legendre polynomials is indicated with the dashed lines. For the reconstruction of concentration of v in terms of Legendre polynomials, we used a segmented expansion with $N = 18$ and a segmentation granularity of $M = 3$, leading to a total of 126 model evaluations used in the expansion. To show that Haar wavelets perform much better in such a region, we additionally do an expansion in terms of Haar wavelets. As resolution level we take $N = 6$, which means a total of $N_w = 128$ wavelets are used in the expansion. In Figure 2.4 the performances of Legendre polynomials and Haar wavelets are compared in the vicinity of $\alpha = 0.23$ (middle inset, Figure 2.4), showing that the Haar-wavelets provide an improvement in accuracy at the bifurcation point, while using the same number of model evaluations (i.e. the same amount of information and computational cost) for the expansion.

2.4.5 Example V. Trichome patterning: dealing with spatial discontinuities

As an extra example of pattern formation we consider a model that describes trichomes. Trichomes are hairs found on the epidermal layer of leaves. In *Arabidopsis Thaliana* these trichomes form a regular pattern, where each trichome

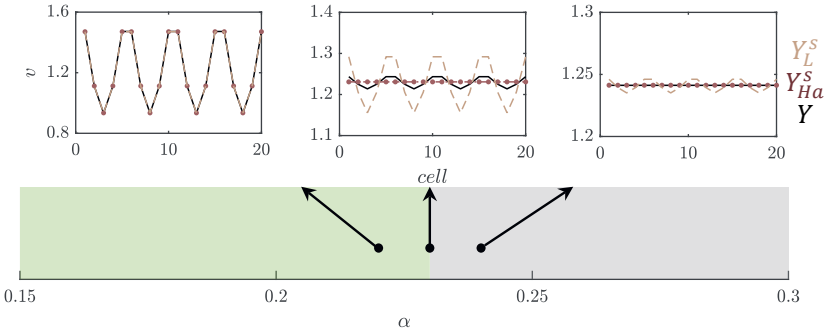


Figure 2.4: **Reconstruction of the concentration of species v of the Schnakenberg model, as defined in (2.38) at steady state, comparing Legendre and Haar wavelet expansions.** We consider the patterning of v in the Schnakenberg model for $\alpha \in [0.15, 0.3]$, as indicated on the main x-axis and assuming $\alpha \sim \mathcal{U}(0.001, 0.45)$, while fixing $\beta = 1, \gamma = 5$ and $d = 20$. The colour along this axis indicates the region in the 1D-parameter space whether a stable pattern will form (i.e., α is inside the Turing Space (TS), indicated by green colouring), or a homogeneous spatial distribution of v (grey colour). We highlight three examples of the patterns formed for different values of α, one inside the TS (left inset, $\alpha = 0.22$), one close to the boundary (middle inset, $\alpha = 0.23$) and one outside the TS (right inset, $\alpha = 0.24$). Within these examples we compare the true solution Y (solid black line) to the reconstructed function of v by PCE in terms of Legendre polynomials Y_L^s by a segmented expansion with polynomial order $N = 18$ and segmentation granularity $M = 3$ (dashed line), or Haar wavelets Y_{Ha}^s (dots), using the first 128 wavelets in the expansion, i.e. $N = 6$ resolution levels.

is separated by around three to four epidermal cells [7]. The model studied here consists of three proteins and their interactions which, taken together, can explain features of trichome patterning [31, 40]. Protein TRANSPARARANT TESTA GLABRA1 (TTG1) binds to the transcription factor GLABRA3 (GL3) which together form a trichome-promoting complex, called the activating complex (AC) [40]. Experimental data suggests that TTG1 is depleted from cells neighbouring a trichome [40]. For this reason the interaction between TTG1 and GL3 is modelled in a substrate-depletion form (Figure 2.5), where TTG1 acts as a substrate for the formation of AC [40]. After non-dimensionalisation this model consists of four parameters, none of which have been experimentally determined, highlighting the substantial amount of uncertainty within this model [35, 81]. Here, we examine the propagation of uncertainty in the parameters to the predicted pattern.

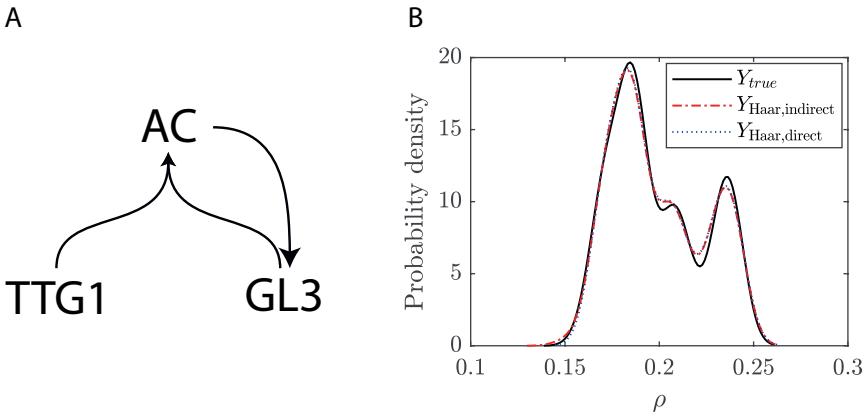


Figure 2.5: **Uncertainty quantification for the trichome system.** A: Schematic of the model. B: Probability density function of the trichome density in the Turing Space using either the indirect (dashed red line) or direct expansion (dotted blue line) method and for comparison the solution of the real model (solid black line). A resolution level of 3 (i.e., a total of 16 wavelets) has been used for the expansion.

The trichome patterning is described by the following set of coupled ODEs [40]:

$$\dot{\text{TTG1}}_j = \alpha - \lambda \text{TTG1}_j - \text{TTG1}_j \text{GL3}_j + \delta \hat{L} \text{TTG1}_j \tag{2.39}$$

$$\dot{\text{GL3}}_j = \beta \text{AC}_j^2 - \text{GL3}_j - \text{TTG1}_j \text{GL3}_j \tag{2.40}$$

$$\dot{\text{AC}}_j = \text{TTG1}_j \text{GL3}_j - \text{AC}_j \tag{2.41}$$

where α, λ, δ and β are parameters in the model and $\hat{L}$ describes the coupling between the cells. The subscript j indicates the j^{th} cell. We solve these equations for 400 cells, grouped on a hexagonal grid of 20 by 20 cells.

In this example we focus on the parameter α, the basal production for TTG1. We assume this parameter to be uniformly distributed on the interval $[0.4, 0.9]$. We are interested in the number of trichomes that are predicted by the model, therefore we consider the trichome density ρ (total number of trichomes divided by the total number of cells in the simulated tissue) as the model response of interest. The number of trichomes is determined by simulating the system until steady state is reached and counting the number

of cells for which the concentration of AC exceeds a threshold. The amount of AC is considered to be an indicator for trichome cell fate in plants, however the biological threshold for this is unknown. We set this threshold to the half-maximum of AC in the system. This leads to the following description of trichome density:

$$\rho = \frac{|\mathcal{T}|}{N} \tag{2.42}$$

$$\mathcal{T} = \left\{ j \in \mathcal{J} \;\middle|\; \mathrm{AC}_j \geq \frac{1}{2}\mathrm{AC}_{\max} \right\} \tag{2.43}$$

where $\mathcal{T}$ is the set of cells which exceed the AC threshold, $\mathcal{J}$ is the set of all cells on the grid, N is the total number of cells and $|\mathcal{T}|$ is the cardinality of $\mathcal{T}$.

The goal is to determine the uncertainty in ρ as a result of the uncertainty in α. To this end, we employ two different approaches. For both approaches we first transform α to a standard uniform variable by $\alpha = T^{-1}(\theta)$, using the transform function for a uniform variable given in (2.8). The first approach, refered to as the indirect approach, is the same as used in Example IV. To reconstruct the concentration at steady state for all cells, we expand the concentration of AC using Haar wavelets. From the result we may determine ρ. In this process we discriminate between cases where there is a pattern and where there is no pattern. For the latter, we need not solve the system as $\rho = 0$. Through linear stability analysis we determine beforehand whether a pattern will form or not, i.e., whether the chosen parameter set is in the Turing Space (TS) [123]. For a certain realisation θ we can determine ρ by

$$\rho(\theta) = \begin{cases} \frac{|\mathcal{T}^s|}{N} & \text{if } \theta \in \mathrm{TS} \\ 0 & \text{if } \theta \notin \mathrm{TS} \end{cases} \tag{2.44}$$

where $\mathcal{T}^s$ is the set of trichomes as determined from the reconstructed AC concentration profile.
Our second, direct approach is to directly reconstruct ρ as

$$\rho^s(\theta) = \sum_{l=0}^{N} Y(\lambda^{(l)})\, u_1^{(l)}\, \psi_l^s(\theta) \tag{2.45}$$

Similarly, as we did for $\rho(\theta)$ we can define $Y(\lambda^{(l)})$ as

$$Y(\lambda^{(l)}) = \begin{cases} \frac{|\mathcal{T}|}{N} & \text{if } \lambda^{(l)} \in \mathrm{TS} \\ 0 & \text{if } \lambda^{(l)} \notin \mathrm{TS} \end{cases} \tag{2.46}$$

In other words, we only solve the system and determine the trichome density if the parameter set falls within the Turing space. This lends robustness to the PCE for the non-smooth parts of the function $Y(\theta)$ and at the same time limits the amount of simulations to be performed, as the non-patterning parameter combinations need not be solved for.

We show here that there are multiple ways in which the uncertainty in the output can be captured. In this case of trichome patterning, we first tested an indirect method where the model output consists of concentration profiles from which the pattern features have to be extracted in post-processing, and secondly, we showed the pattern feature could also be expanded directly, by taking the density as model output. Comparing the indirect and direct approaches, we conclude that both have similar levels of accuracy (Figure 2.5). Note that in both cases the expansions converge to the real solution at resolution level N = 6, which means a summation of 128 wavelets. The PDF in Figure 2.5 is constructed using 10^3 samples which costs 4.7 seconds for the wavelet reconstruction as opposed to 80.9 seconds for solving the full model.

2.5 Discussion and Conclusions

Through a series of examples we have presented an efficient and widely-applicable version of spectral methods for quantification of the effect of parameter uncertainty on model outcomes. The present scheme utilizes non-intrusive spectral projection based on polynomial functions or wavelets. The orthonormal properties of those functions provide a novel scheme to determine the expansion coefficients in a computationally fast way. The scheme is similar to the Golub-Welsh algorithm known from Gaussian quadrature [129]. In fact, the points indicating the roots of the polynomials used obtained in quadrature procedures, exactly correspond to the eigenvalues obtained from the matrix which plays a central role in our approach. Quadrature methods as well as sampling methods are traditionally used to determine the coefficients of expansions of functions one is interested in [95]. The approach used here does effectively the same and provides an alternative for existing approaches, with the advantage that it is flexible and applicable to any set of orthonormal basis functions.

The method presented here requires no modification of the model equations. This is in general the main advantage of non-intrusive methods: there is no need to recast the model into a probabilistic framework. Instead, the random behaviour of parameters is accounted for through a set of determin-

istic simulations of the process for a restricted number of parameter values. These values are chosen such that they reflect the uncertainty in the parameters. To test the performance of our method we applied it to five different models: (1) a model of exponential decay, (2) a biochemical reaction network, (3) the glycolytic oscillator, (4) the Schnakenberg model, and (5) a trichome model. The latter two models describe spatial pattern formation. For each test case, the results of the PCE were compared to, if available, analytical solutions, non-PCE numerical simulations, or Monte Carlo simulations. In these comparisons we mostly focused on the accuracy of PCE. Even though the computational advantage is one of the major reasons for using PCE techniques, we do not focus on it because this aspect has already been extensively explored in various applications. For such comparisons we refer to other publications, see for example [95, 130, 131]. The logic behind the compututational advantages of PCE over MC extends to the methods we have presented here and is evident in, e.g., in the case of trichome pattern formation (Example V), for which the output of the model was obtained 17 times faster when using PCE instead of MC.

The accuracy of the reconstruction by PCE depends on the choice of expansion order and basis functions. While the latter is determined by the PDFs of the input parameters, the choice of expansion order has to be chosen by the user. For example, in Examples I and III we chose $N = 5$ and $N = 10$, respectively. These choices were based on careful observation of the convergence properties of the method. In some cases the expansion order has to be chosen prohibitively large. For such a situation we propose an alternative approach that segments the parameter interval into subintervals, essentially zooming in on these sub-intervals such that a lower expansion order can be used in each sub-interval. In Example II we have shown that this segmentation approach can greatly reduce the computational costs in summation part of the expansion, thus providing a way to circumvent the curse of dimensionality. Such adaptations are required for the more difficult high-dimensional cases and the segmentation is a relatively straight-forward method to tackle dimensionality problems. It is an alternative for so-called sparse PCE methods, that utilize only a small subset of the polynomial basis functions in order to limit the amount of model evaluations [96, 116, 126].

Convergence of the PCE may be poor in regions of the parameter space around a bifurcation [103]. In PCE, smooth polynomials are used in the expansions and they may show effects similar to the Gibbs phenomenon in Fourier expansions, i.e., the spectral basis is not suitable and leads to a slow

convergence. Since smooth functions like the Hermite and Legendre polynomials will fail to describe steep or discontinuous solutions, we explored the use of Haar wavelets. These wavelets lead to localized decompositions and this produces more robust behaviour [115]. We show in Example IV (Schnakenberg model) the advantages of using Haar wavelets over polynomials by focusing on the region in parameter space where the system transitions from spatially heterogeneous dynamics to spatially homogeneous dynamics take place. Around this bifurcation point an expansion in terms of Haar wavelets also shows slow convergence, but the accuracy of the expansion is 10 times better than when Legendre polynomials are used. For a fair comparison we kept the number of model evaluations in both approaches the same. In the vicinity of bifurcations Haar wavelets show greater robustness than the traditional polynomial basis functions. They thus provide a useful tool for biological systems which often feature such discontinuity.

In Example V (trichome pattern formation) we have highlighted the flexibility of the method: certain quantities, e.g., the scalar quantity of trichome density, can either be directly expanded or indirectly. By making use of that adaptability the number of model evaluations can be reduced while maintaining the same level of accuracy.

Overall, the approach presented here consists of a number of easy-to-implement steps and is widely applicable to a variety of systems which would normally be intractable when used in the context of uncertainty quantification.

2.6 Author Contributions

C.F. conceived the project, A.D. performed all calculations, A.D., J.M. and C.F. wrote the manuscript.

2.7 Supplementary info

2.7.1 S1: Expanding a meta model

To illustrate the derivation of (2.5) and (2.6), we consider here the case with only one uncertainty parameter θ, distributed according to PDF $P(\theta)$. The involved Hilbert space consists of functions of θ that are square-integrable with respect to the inner product

$$\langle f, g \rangle = \int f(\theta)\, g(\theta)\, P(\theta)\, d\theta \tag{2.47}$$

for any two functions $f(\theta)$ and $g(\theta)$ in this space. So, $P(\theta)$ acts as weight function in the inner product.

In this Hilbert space we make use of a set of basis functions $\phi_n(\theta), n = 0, 1, 2, \ldots$ that are orthonormal with respect to this inner product. These may be Legendre, Hermite, Jacobi, Chebyshev, Laguerre, and other polynomials, but also Bessel functions, Hankel functions, wavelets and other functions may be used. The only requirement is orthonormality.

We assume the model response $Y(\theta)$ to be in the Hilbert space. This implies that we may expand $Y(\theta)$ in terms of the basis functions:

$$Y(\theta) = \sum_{n=0}^{\infty} c_n\, \phi_n(\theta), \tag{2.48}$$

with the expansion coefficients given by

$$c_n = \langle Y, \phi_n \rangle = \int Y(\theta)\, \phi_n(\theta)\, P(\theta)\, d\theta. \tag{2.49}$$

It is convenient to associate with $Y(\theta)$ the matrix $\hat{Y}$ whose elements read as

$$\hat{Y}_{n,m} = \langle \phi_n, Y \phi_m \rangle. \tag{2.50}$$

Assuming Y to be an analytical function in θ, we may write down its Taylor expansion:

$$Y(\theta) = \sum_{k=0}^{\infty} d_k\, \theta^k. \tag{2.51}$$

For the matrix $\hat{Y}$ this implies that it is given by the expansion

$$\hat{Y} = \sum_{k=0}^{\infty} d_k\, \hat{B}(k), \tag{2.52}$$

with the elements of matrix $\hat{B}(k)$ given by

$$\hat{B}_{n,m}(k) = \langle \phi_n, \theta^k \phi_m \rangle. \tag{2.53}$$

Using the completeness relation of the basis functions

$$\sum_{n=0}^{\infty} \phi_n(\theta)\, \phi_n(\theta')\, P(\theta) = \delta(\theta - \theta'), \tag{2.54}$$

we may factorize this matrix. For $k = 0$ we have $[\hat{B}(0)]_{n,m} = \delta_{n,m}$, and for $k \geq 1$

$$\hat{B}(k) = \hat{B}^k, \tag{2.55}$$

where $\hat{B} \equiv \hat{B}(1)$. So, $\hat{Y}$ can be written as

$$\hat{Y} = \sum_{k=0}^{\infty} d_k\, \hat{B}^k. \tag{2.56}$$

with

$$\hat{B}_{n,m} = \int \phi_n(\theta)\, \theta\, \phi_m(\theta)\, P(\theta)\, d\theta. \tag{2.57}$$

$\hat{B}$ is a symmetric, $(\infty \times \infty)$ matrix. So, it has real eigenvalues $\lambda^{(l)}, l = 1, 2, \ldots$ and there exists an orthogonal basis of ∞-dimensional eigenvectors $u^{(l)}$. Note that, as usual in the literature, but in contrast with the basis in the Hilbert space introduced above, the counting here starts at $l = 1$.
The completeness relation for the $u^{(l)}$ reads as

$$\sum_{l=1}^{\infty} u_i^{(l)} u_j^{(l)} = \delta_{i,j}. \tag{2.58}$$

Using the relations obtained above, we may rewrite the expansion for Y in the

following way:

$$
\begin{aligned}
Y(\theta) &= \sum_{n=0}^{\infty} \hat{Y}_{0,n}\, \phi_n(\theta) && (2.59)\\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} d_k\, (\hat{B}^k)_{0,n}\, \phi_n(\theta) && (2.60)\\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} \sum_{l=1}^{\infty} d_k\, (\hat{B}^k)_{0,i}\, u_{i+1}^{(l)}\, u_{n+1}^{(l)}\, \phi_n(\theta) && (2.61)\\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \sum_{l=1}^{\infty} d_k\, (\lambda^{(l)})^k\, u_1^{(l)}\, u_{n+1}^{(l)}\, \phi_n(\theta) && (2.62)\\
&= \sum_{n=0}^{\infty} \sum_{l=1}^{\infty} Y(\lambda^{(l)})\, u_1^{(l)}\, u_{n+1}^{(l)}\, \phi_n(\theta) && (2.63)\\
&= \sum_{l=1}^{\infty} Y(\lambda^{(l)})\, u_1^{(l)} \psi_l(\theta), && (2.64)
\end{aligned}
$$

In the last step we used an orthogonal basis transformation from $\phi(\theta)$ to $\psi(\theta)$ with the help of the orthogonal vectors $u^{(l)}$:

$$\psi_l(\theta) \equiv \sum_{n=0}^{\infty} u_{n+1}^{(l)}\, \phi_n(\theta). \tag{2.65}$$

The important conclusion is that we may express the model response $Y(\theta)$, which originally depends on the continuous parameter θ, in terms of discrete values $Y(\lambda^{(l)}), l = 1, 2, \ldots$. This implies that Y needs to be evaluated only at the points $\theta = \lambda^{(l)}, l = 1, 2, \ldots$ in parameter space. Note that most ingredients of the formalism can be calculated in advance and once and for all.

The corresponding meta-model response Y^s we are aiming at is obtained from Y by taking into account only the lowest N basis functions, for some integer N. So,

$$Y^s(\theta) = \sum_{n=0}^{N} c_n\, \phi_n(\theta). \tag{2.66}$$

Note, that the coefficients $c_n, n \leq N$, do not change if we omit the terms with $n > N$, thanks to the orthonormality of the basis functions. For the meta-

model we have

$$Y^s(\theta) = \sum_{l=0}^{N} Y(\lambda^{(l)})\, u_1^{(l)}\, \psi_l^s(\theta). \tag{2.67}$$

We remark that omitting higher order basis functions has also consequences for the ψ_l. Instead of ψ_l we now use

$$\psi_l^s(\theta) \equiv \sum_{n=0}^{N} u_{n+1}^{(l)}\, \phi_n(\theta). \tag{2.68}$$

In practice one should for each case analyze the accuracy of Y^s in approximating Y, especially as a function of N.

2.7.2 S2: Derivation of the probability density function for the exponential decay model

For the exponential decay model an analytical expression for the probability density function (PDF) can be obtained. In order to derive the PDF we start with:

$$P(y,t) \quad = \quad \int_0^\infty \delta(y - A(t,\theta)) P(\theta) d\theta.$$

For log-normal distributed k we have for the PDF of the uncertainty parameter θ and $k(\theta)$:

$$\begin{aligned} P(\theta) &= \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2} \\ k(\theta) &= \mu e^{\alpha\theta - \alpha^2/2}, \end{aligned}$$

with

$$\alpha \quad = \quad \sqrt{\ln\left(1 + \frac{\sigma^2}{\mu^2}\right)}.$$

The function $A(t,\theta)$ is given by:

$$A(t,\theta) \quad = \quad A_0 e^{-k(\theta)t}.$$

Defing $u = y - x$ we arrive at:

$$P(y,t) = \int_{u(0)}^{u(\infty)} \delta(u) \left| \frac{\partial A}{\partial \theta} \right|_{\theta=\theta(u)}^{-1} P(\theta(u)) du.$$

Now, it holds:

$$\begin{aligned} \frac{\partial A}{\partial \theta} &= -\alpha_0 t e^{-kt} \alpha k \\ \theta(u) &= \frac{1}{\alpha} \ln \left[\frac{e^{\alpha^2/2}}{\mu t} \ln \frac{y-u}{x_0} \right], \quad u < y. \end{aligned}$$

and further:

$$\begin{aligned} u(\infty) &= y - A(t,\infty) = y - A_0 e^{-k(\infty)t} \\ k(\infty) &= \infty \quad (\alpha > 0) \\ u(\infty) &= \begin{cases} y - A_0 & : t = 0 \\ y & : t > 0 \end{cases} \\ k(-\infty) &= 0 \\ u(0) &= y - A_0 \end{aligned}$$

Using these results we arrive at:

$$\begin{aligned} P(y,t) &= \int_{u(0)}^{u(\infty)} \delta(u) \frac{e^{k(u)t}}{x_0 t \alpha k(u)} \frac{e^{-\theta^2(u)/2}}{\sqrt{(2\pi)}} du \\ P(y,t) &= \frac{e^{k(u=0)t}}{\sqrt{2\pi} A_0 t \alpha k(u=0)} e^{-\theta^2(0)/2} \cdot \theta(y)\theta(A_0 - y). \end{aligned}$$

Inserting

$$k(\theta(u=0)) = \frac{1}{t} \ln \left(\frac{A_0}{y} \right),$$

we finally arrive at:

$$P(y,t) = \frac{1}{\sqrt{2\pi}\alpha y \ln\left(\frac{A_0}{y}\right)} \exp\left[-\frac{1}{2\alpha^2} \left\{ \ln\left(\frac{e^{\alpha^2/2}}{\mu t} \ln\left(\frac{A_0}{y} \right) \right) \right\}^2 \right] \cdot \theta(y)\theta(A_0 - y).$$

Chapter 3

Identification of the trichome patterning core network using data from weak *ttg1* alleles to constrain the model space

Rachappa Balkunde*, Anna Deneer*, Hanna Bechtel, Bipei Zhang, Stefanie Herberth, Martina Pesch, Benjamin Jägle, Christian Fleck and Martin Hülskamp

* Contributed equally

Published in: Cell Reports 33, 108497 (2020).
doi.org/10.1016/j.celrep.2020.108497

3.1 Abstract

The regular distribution of trichomes on leaves in Arabidopsis is a well-understood model system for two-dimensional pattern formation. It involves more than 10 genes and is governed by two patterning principles, the activator-inhibitor (AI) and the activator depletion (AD) mechanisms, though their relative contributions are unknown. The complexity of gene interactions, protein interactions and intra- and intercellular mobility of proteins makes it very challenging to understand which aspects are relevant for pattern formation. In this study we used global mathematical methods combined with a constraining of data to identify the structure of the underlying network. To constrain the model, we performed a genetic, cell biological and biochemical study of weak ttg1 alleles. We find that the core of the trichome patterning is a combination of the AI and AD mechanisms differentiating between two pathways activating the long-range inhibitor CPC and the short-range inhibitor TRY.

3.2 Introduction

Mathematical modelling has become an integrated discipline in developmental biology aiming to integrate the knowledge about a biological system in order to arrive at either conceptual statements or predictions about the consequences of experimental manipulations. One frequent problem, however, is that the available data sets do not contain enough information for the parameters of the mathematical model to be uniquely identified. This means that the model is capable to reproduce the data with many different parameters and this may lead to the undesirable situation that no clear conceptual statements can be made and many different model predictions are possible. The way out of this situation is to use global analysis methods in combination with further constraining observations or data. It is possible, although challenging, to use mathematical modelling approaches in these situations and to arrive at conceptual statements on the developmental process despite the incomplete information about the biological system. Trichome patterning in Arabidopsis is an excellent model system for which this constraint global analysis method needs to be applied. On the one hand it presents a fairly simple developmental process that can be described as a two-dimensional pattern formation [31]. On the other hand, the underlying gene regulatory network involves many gene interactions, protein-protein interactions and intercellular transport processes for which the parameters are largely unknown ([35]).

Our understanding of trichome formation in Arabidopsis thaliana is based on the genetic identification of the key genes and a detailed molecular and cell biological analysis [35, 36, 132, 133]. Trichomes are initiated without reference to already existing positional information with a regular spacing. Although there is a remarkable degree of variability in the relative distances [32], trichomes are normally not found immediately next to each other [34]. This pattern is established early during leaf development. On young rosette leaves incipient trichomes are typically separated by 3-4 epidermal cells and their distance is increased during leaf expansion because of cell divisions and growth of the intervening cells [34]. Trichome patterning in Arabidopsis thaliana is regulated by a gene regulatory network involving trichome promoting and trichome inhibiting genes. Three genes act as the major positive regulators: TRANSPARENT TESTA GLABRA1 (TTG1) encodes a WD40 protein [47–49], GLABRA1 (GL1) encodes a R2R3 MYB related transcription factor [45], and GLABRA3 (GL3) a basic helix-loop-helix (bHLH)-like transcription factor [34, 44, 46]. In addition, MYB23 and EGL3 act in a partially redundant manner with GL1 and GL3, respectively [134–136]. TRIPTYCHON (TRY) and

CAPRICE (CPC) represent negative regulators of trichome development as the corresponding single mutants show trichome clusters and a higher trichome density, respectively [34, 50, 51]. They encode R3 single repeat MYB proteins and act partially redundant with five additional homologs [50, 53–59].

The trichome promoting and inhibiting proteins show a complex protein interaction pattern. TTG1 and GL1 both bind to GL3/EGL3 [38, 43, 44, 134, 136–139] and the binding of one of them counteracts the binding of the other to GL3 [64]. This competitive behaviour is also seen at the level of target gene regulation such that the transcriptional activation of TRY by GL3 and TTG1 is counteracted by GL1 and the activation of CPC by GL3 and GL1 is inhibited by TTG1 [64]. Also, the negative regulators bind to GL3/EGL3 thereby preventing the binding of GL1 [44, 69, 140].

Theoretical models have been developed to explain how the gene regulatory network formed by the activators and inhibitors can create a de novo trichome pattern [38, 40–42]. The molecular interaction schemes are consistent with two general principles, an activator-inhibitor (AI model) and an activator depletion model (AD model) [35]. According to the activator-inhibitor model the activators promote the expression of the inhibitors. The inhibitors move between cells and repress the activators in the neighbourhood. The activator-depletion model explains the establishment of a trichome pattern patterning by the depletion of the activator in the neighbourhood of trichome initials. It is based on the finding that TTG1 can move between cells and that TTG1 is trapped in trichome initials by binding to GL3 [40, 141]. Both principles are able to generate regular spacing patterns de novo [17]. While the experimental data support the existence of both mechanisms it is very difficult to assess their individual biological significance because the two mechanisms involve the same genes.

Both, the AI and the AD models had been primarily used to demonstrate the minimal requirements to describe trichome patterning. Most importantly, the models do not *need* the additional genes that are, however, known to act in the network. In this study, we constrain the model to unfold structural elements by analysing the TTG1 gene in greater detail. TTG1 appears to be a key component in the AI as well as in the AD model. We focus on the molecular function of weak *ttg1* alleles. While strong ttg1 alleles are devoid of trichomes, weak ttg1 alleles produce a reduced number of trichomes [48, 49, 142–144]. Remarkably, all weak *ttg1* alleles exhibit trichome clusters.

It was possible to predict this unusual behaviour of TTG1 alleles by changing several parameters of the TTG1 function in the AD model [40]. It is not known whether the role of TTG1 in the AI model also contributes to the phenotype of weak *ttg1* alleles and if so what the relative contributions of the AD and the AI mechanisms are. According to the AD model, cluster formation can be explained by a failure to remove TTG1 from the neighbourhood of trichome initials. In the AI model, cluster formation in weak *ttg1* alleles could be explained if TRY would not be expressed. We show that the interaction of mutant TTG1 proteins carrying the point mutations of the weak ttg1 alleles with GL3 protein are strongly impaired and that mutant TTG1 protein is not depleted around trichomes. We further show that TRY expression but not CPC expression is strongly reduced in ttg1 alleles. We developed a new mathematical model that combines the AI and AD models. We reveal the requirement of additional structural elements in the network and demonstrate that the reduced interaction of TTG1 to GL3 is sufficient to explain all trichome phenotypes of weak *ttg1* mutants.

3.3 Results

3.3.1 Weak *ttg1* alleles exhibit irregular spatial distribution of trichomes

Strong ttg1 alleles show a glabrous phenotype indicating that TTG1 has a positive role in trichome formation. All weak alleles exhibit a cluster phenotype [48, 49, 143] suggesting a negative function in trichome formation (Figure 3.1). Thus strong, and weak *ttg1* alleles appear to have opposite genetic functions. Interestingly, this is specific to the trichome trait as seed colour, seed coat mucilage and root hair formation phenotypes do not show the opposite phenotype in weak *ttg1* alleles*ttg1-9*, *ttg1-11* and *ttg1-12* alleles (Figure S1). To understand the molecular basis of cluster formation in the weak *ttg1* alleles, we analysed the three weak *ttg1* alleles *ttg1*-9 [49, 142], *ttg1-11* and *ttg1-12* [142, 143] in more detail using two strong *ttg1* alleles *ttg1-1* [48] and *ttg1-13* [143] as a reference. The three weak alleles have point mutations leading to amino acid exchanges at different positions [48, 49, 142, 143] (Table S1).

During the phenotypic analysis of the three weak *ttg1* alleles we noticed that trichomes appeared to be less regularly distributed compared to wild type, suggesting a randomized trichome pattern. We used TrichEratops to generate coordinates [145]. We limit the analysis to the *tt1-9* allele because *ttg1-11* and *ttg1-12* alleles produce too few trichomes on each leaf to yield sufficient data for a meaningful statistical analysis. To quantitatively com-

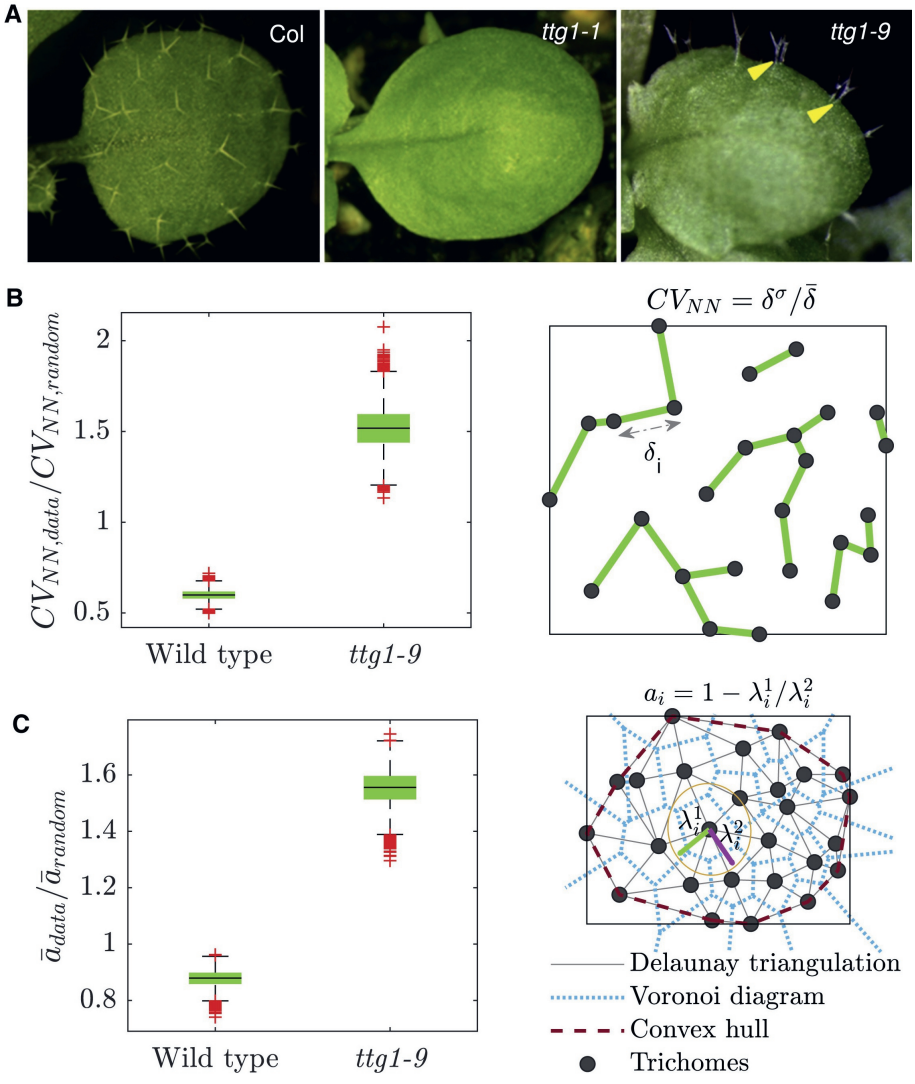


Figure 3.1: **Trichome Patterning defects in *ttg1-9* leaves** (A) Fourth rosette leaves of wild type and *ttg1-1* and *ttg1-9* mutants. Note that the strong *ttg1-1* allele has no trichomes, while the weak *ttg1-9* allele exhibits clusters (yellow arrowheads). Scale bars, 1 cm. (B) Nearest neighbour distances. δ^{σ} represents the standard deviations and $\bar{\delta}$ the mean of the nearest neighbour distances. (C) Mean anisotropy, defined as the ratio of the eigenvalues as a measure of deviation from isotropy.

pare the regularity of trichome patterns on wild-type and *ttg1-9* leaves we used different measures.

As a first measure we use the coefficient of variation of the nearest neighbour distances distribution (CV_{NN}) which is defined as the standard deviation normalized by the mean of the nearest neighbour distances (Figure 3.1B). For wild type we found a regular but not perfect pattern [32] with a CV of 0.33 ± 0.05. The *ttg1-9* allele exhibits a more irregular pattern with a mean CV of 0.84 ± 0.22. Next, we compared the leaf phenotypes to randomly generated point patterns. For the random patterns we consider the difference between the trichome densities for wild type and *ttg1-9* and simulate 10^4 instances, where for each instance the points are placed with a uniform probability across the simulated region. Using the bootstrapping method, we generate 10^4 samples based on the CV_{NN} distribution from the leaves and compare the bootstrap distributions with the random references (Figure 3.1B). In this comparison we show that the wild-type pattern exhibits significantly less variability in trichome distances than the corresponding random reference (one-tailed Mann-Whitney U-test, $p < 0.01$), whereas the *ttg1-9* allele shows a higher irregularity than the random distribution with similar density ($p < 0.01$).

Because the nearest neighbour distances give a very narrow viewpoint on the region around a trichome, we decided to employ another quantifications of the variability, which takes into account a more appropriate region of interest. Towards this end we calculated the anisotropy of the neighbourhood around the trichomes [32] (Figure 3.1C). The anisotropy from the experimental data is taken relative to the corresponding random references. Comparing between the anisotropy of bootstrapped experimental results and the random references we come to the same conclusion as for the CV_{NN}. Testing between wild-type and *ttg1-9* showed that the *ttg1-9* allele has a significantly higher relative mean anisotropy (one-tailed Mann-Whitney U-test, $p < 0.01$), confirming that *ttg1-9* spatial distributions are more irregular than wild-type.

For a hypothetical homogeneous (not depending on space) random trichome pattern the cluster probability depends inversely on the trichome density (Figure S2A). Thus, *ttg1-9* mutants should have fewer clusters than wild type. This is calculated for wild type and *ttg1-9* in Figure S2A. Thus, cluster formation in ttg1-9 mutants cannot be explained in full by a homogeneous random process, suggesting that it is due to a deterministic process. To differentiate between stochastic processes and deterministic processes that only appear random we determined the correlation dimension [146]. This gives an

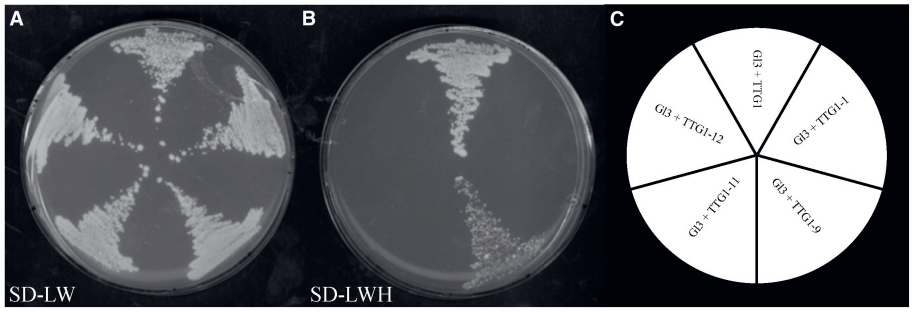
estimate of the number of parameters needed to explain the measured variability. We find for both wild type and ttg1-9 data a correlation dimension smaller than the homogeneous random patterning process (Figure S2C-F), indicating that deterministic factors control a considerable part of the variability in trichome distances. The theoretically expected probability of clusters is not dependent on the area available for trichome formation (Figure S2B) suggesting that these deterministic factors are represented by altered genetic interactions in *ttg1-9* mutants.

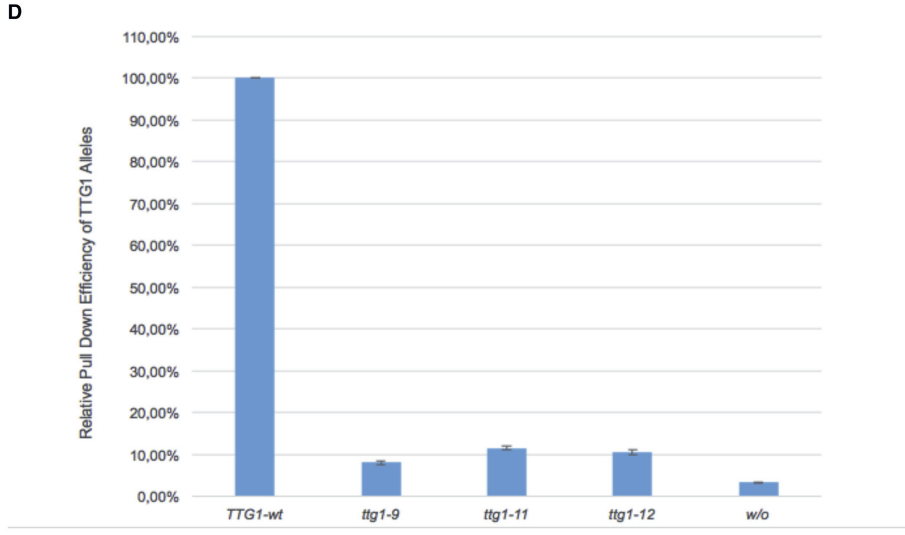
3.3.2 Amino acid exchanges in weak *ttg1* alleles reduce or abolish the interaction with GL3, EGL3 and TT8

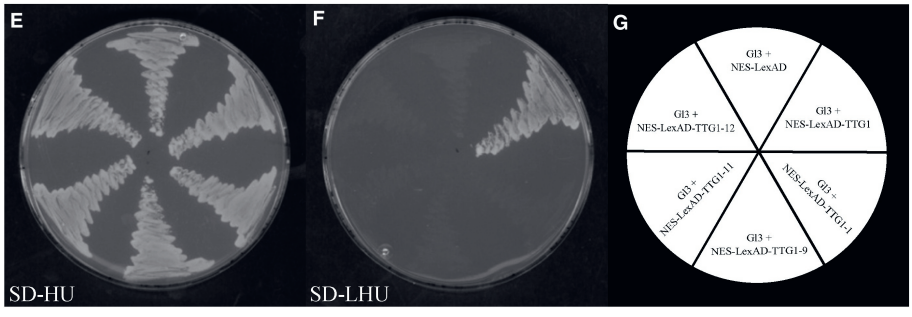
To explore the molecular function of the weak *ttg1* alleles, we studied the protein interaction of TTG1 with GL3. TTG1 is considered to regulate trichome formation by binding to GL3. It has been shown before, that the TTG1 protein lacking the 26 C-terminal amino acids (corresponding to the *ttg1-1* allele) does not interact with GL3 [44, 141] and that two newly identified weak alleles, *ttg1-23* and *ttg1-24*, show reduced interaction in yeast two hybrid assays [144]. We therefore studied the interaction of the mutant TTG1 proteins with GL3 in yeast two-hybrid interaction experiments (Figure 3.2A-C). We found no interaction between GL3 and the TTG1-1, TTG1-11 and TTG1-12 mutant proteins. The TTG1-9 mutant protein exhibited reduced binding as judged by weak colony growth. To independently test the protein interaction between GL3 and TTG1 proteins, we studied the interactions using a pulldown assay by expressing the proteins in human embryonic kidney (HEK293TN) cells and a quantification of the precipitated proteins in a luminescence-based mammalian interactome mapping (LUMIER assay (Blasche and Koegl, 2013)). In these experiments the pull down efficiencies of the three weak mutant TTG1 proteins and TTG1-1 were about tenfold lower than wild type but in all experiments clearly significantly above background (t-test, $p < 0.001$) (Figure 3.2D). The LUMIER assay appears to be more sensitive and indicate that binding of the mutant TTG1 proteins to GL3 is greatly reduced but not completely abolished. Also the binding of the mutant TTG1 proteins to EGL3 and TT8 was impaired in yeast two-hybrid assays (Table S2).

3.3.3 Nuclear targeting and spatial distribution in the epidermis of TTG1 mutant proteins is impaired

GL3 has been shown to trap TTG1 in the nucleus [40]. We reasoned that the reduced interaction of weak TTG1 mutant proteins with GL3 also affects their







(Caption on next page.)

Figure 3.2: **Protein-Protein Interactions and Nuclear Transport of Wild-Type and Mutant TTG1 Proteins.** (A–C) Protein-protein interactions between TTG1 mutant proteins and GL3. (A) Control for presence of plasmids. (B) Interaction assay on medium supplemented with 5 mM 3-aminotriazole (3-AT). Yeast growth indicates positive interactions. (C) Schematic presentation showing the positions of different combination on the plates. (D) Pull-down efficiency of TTG1 alleles by ProtA-tagged GL3 in LUMIER assays. The pull-down efficiency of the three weak TTG1 alleles is shown relative to wild type (defined as 100%). w/o is the control in which protein extract from non-transformed cells was used. (E–G) Yeast-based nuclear transportation trap (NTT) assay to test the ability of GL3 to mediate nuclear transport of different TTG1 mutant proteins. (E) Control for presence of plasmids. (F) Nuclear transport assay. Yeast growth indicates GL3-mediated nuclear transport of NES-LexAD-TTG1 proteins. (G) Schematic presentations showing the positions of different combinations on the plates.

nuclear transport. We tested this using the yeast nuclear transportation trap assay (NTT) [147]. TTG1 was fused to the transactivator LexAD (LexA DNA-binding domain and GAL4AD transactivation domain) and a nuclear export sequence from HIV Rev protein to generate NES-LexAD-TTG1. The NES sequence mediates nuclear export of TTG1. Binding to GL3 can overcome the NES driven nuclear export, which in turn leads to an activation of the LexAD-responsive LEUCINE2 gene reporter. Wild-type TTG1 was efficiently targeted to the nucleus (Figure 3.2E-G) [141]. By contrast, none of the four TTG1 mutant proteins were directed to the nucleus in these assays (Figure 3.2E-G). Thus, the weak binding of mutant TTG1 proteins to GL3 is not sufficient to mediate nuclear transport.

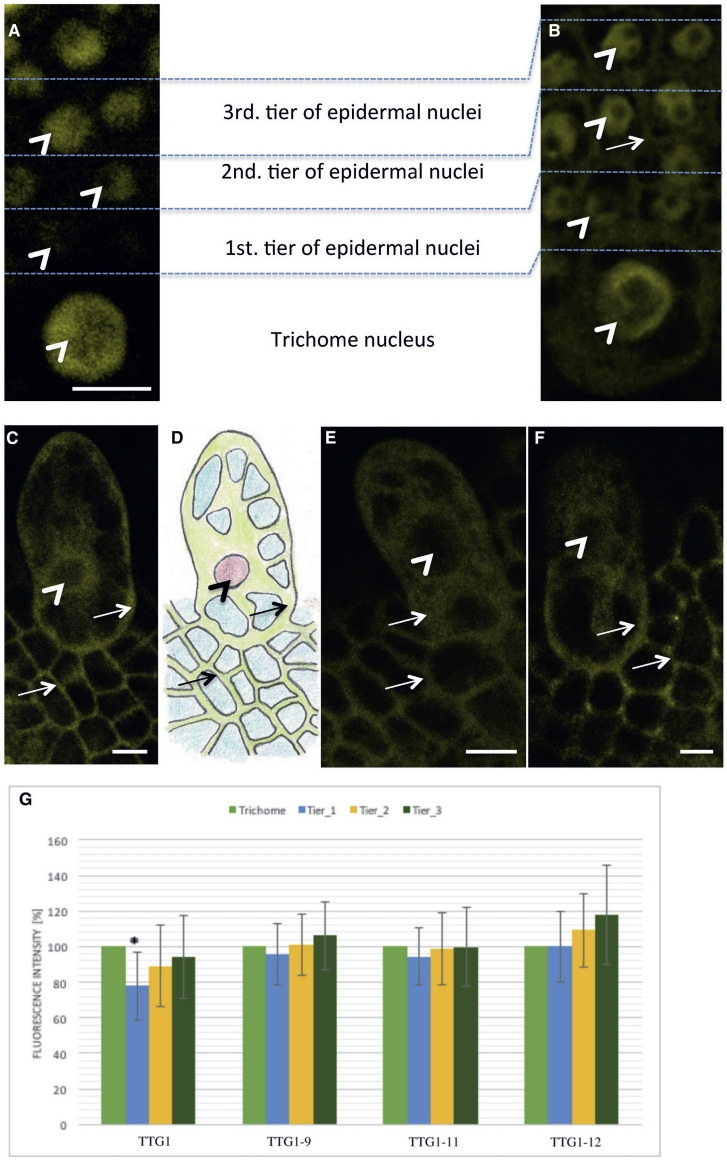
It had been previously shown that wild-type TTG1-YFP protein is localized to the nucleus when expressed in the *ttg1* mutant background [40]. This genetic situation can be considered to reflect the wild-type situation as two intact TTG1 gene copies are present. To analyse the localization of TTG1-9, TTG1-11 and TTG1-12 proteins in plants we generated transgenic lines expressing YFP tagged TTG1 proteins under the control of the TTG1 promoter. We used the wild-type background to enable the analysis in the context of normal trichome development. TTG1-YFP was localized exclusively in nuclei when expressed in the *ttg1-13* mutant background (Figure 3.3A, [40]). Wild-type TTG1-YFP expressed in wild type background showed nuclear and cytoplasmic localization (Figure 3.3B). As this line harbours four TTG1 gene copies it is conceivable that the increased gene dosage leads to a saturation of the system such that GL3 cannot efficiently target TTG1 to the nucleus any-

more (Figure 3.3B). The three mutant proteins were found predominantly in the cytoplasm, and trichome nuclei had much less fluorescence intensity than the cytoplasm (Figure 3.3C-F).

The GL3 dependent localization of TTG1 also results in a depletion of TTG1-YFP in immediate neighbour cells of a developing trichome (39% of the fluorescence found in the trichome) when expressed in the *ttg1-13* mutant background[40, 141]. The reduced binding or no binding of mutant TTG1 proteins to GL3 suggested to us that they show an altered spatial distribution in the leaf epidermis. As a reference we show TTG1-YFP expressed in the *ttg1-13* mutant with high levels of fluorescence in the nucleus of the trichome and much lower levels in the epidermal nuclei immediately next to the trichome (1st tier, Figure 3.3A, [40]). Wild-type plants expressing TTG1-YFP under the TTG1 promoter showed a similar distribution (Figure 3.3B). In the first tier of cells next to the trichome TTG1-YFP levels were significantly ($p=0.002$; Students t-test, Table S3) reduced to 78% as compared to the trichome cell, though this depletion is clearly less pronounced as compared the expression in *ttg1-13* mutants (39%, [40]). This is likely due to the difference in the gene dosage in the two experiments. It is conceivable that in this situation GL3-driven trapping cannot efficiently cope with too much TTG1 protein leading to reduced depletion. This view is supported by our modelling approach. When simulating different TTG1 levels, we found a decreased depletion for higher TTG1 levels (Figure S3B). In a next step, we analysed the depletion in the TTG1-9-YFP, TTG1-11-YFP and TTG1-12-YFP lines. All three lines showed no detectable depletion of the signal in the neighbouring epidermal cells (Figure 3.3C-F, Table S3). This was consistent with our modelling results simulating the effect of different protein amounts (Figure S3B). Our findings show that the mutant TTG1 proteins are not efficiently targeted to the nucleus and show not the characteristic depletion of TTG1 protein in the immediate neighbour cells.

3.3.4 Expression of TRY and CPC in *ttg1* mutant alleles

One possibility to explain the cluster phenotype would be that TRY is not properly expressed in the weak ttg1 alleles. Mutations in TRY result in the formation of trichome clusters in about 9% of all trichome initiation sites [148, 149]. Genetic experiments had suggested that TRY and TTG1 act in the same pathway as trans-heterozygous mutant combinations show clusters [148]. To tested the possibility that the cluster phenotype in weak ttg1 alleles is due to a reduced TRY activity by analysing the expression of TRY using



(Caption on next page.)

Figure 3.3: **Localization of Wild-Type and Mutant TTG1 Proteins in the Epidermis of *Arabidopsis thaliana*** (A) TTG1-YFP expressed in *ttg1-13* is localized in the nucleus (arrowhead). High fluorescence is found in the trichome. In 1st-tier nuclei, TTG1-YFP is barely visible, and intensity increases in the 2nd and 3rd tiers. (B) TTG1-YFP expressed in wild-type background is found in the cytoplasm (arrows) and the nuclei (arrowhead). Depletion in 1st-tier nuclei is much less pronounced as compared to (A). (C) TTG1-9-YFP shows fluorescence in the cytoplasm (arrows) and no or little fluorescence in the nucleus (arrowhead). We found no depletion of fluorescence in the cytoplasm (arrows) of the neighboring cells. (D) Color-coded cells shown in (C) to indicate the relevant compartments (red, nucleus; yellow, cytoplasm; blue, vacuole). Epidermal cells contain one large vacuole and a thin cytoplasmic layer at the cortex at this stage of leaf development. (E) TTG1-11-YFP. (F) TTG1-12-YFP. (G) Relative fluorescence intensity in the 1st, 2nd, and 3rd tier of cells as a percentage of the fluorescence measured in the trichome cell in different genetic situations. Error bars are shown; statistical analysis is shown in Table S3. Scale bars, 10 μm.

a 1.8 kb promoter fragment of TRY (pTRY:GUS) driving the β-glucoronidase (GUS) reporter gene [52]. In addition we studied the expression of the CPC gene to judge the specificity of regulation events. Here, we used a 525 bp long 5' upstream region of the CPC gene (pCPC:GUS) [52].

The pTRY:GUS marker was expressed ubiquitously in young wild-type leaves with slightly elevated levels in trichomes (Figure 3.4A). In mature leaves, pTRY:GUS was expressed only in trichomes (Figure 3.4F). In *ttg1-1*, *ttg1-9*, *ttg1-11* and *ttg1-12* mutants we found neither the initial ubiquitous expression on young leaves nor the trichome specific expression on mature leaves (Figure 3.4).

The pCPC:GUS marker was expressed in trichomes in young and mature leaves (Figure 3.5A,F). The strong *ttg1-13* allele revealed no CPC expression (Figure 3.5E,J) and the two weak *ttg1-11* and *ttg1-12* alleles showed clear expression in the few trichomes present on young leaves (Figure 3.5B,G,C,H,D,I). The expression of CPC was not obviously changed in the *ttg1-9* allele (Figure 3.5B,G).

3.3.5 Rescue of the *ttg1-9* cluster phenotype by expressing TRY under the CPC promoter

The expression analysis of TRY and CPC in weak ttg1 mutants suggests their differential regulation such that CPC is activated while TRY is not. If this

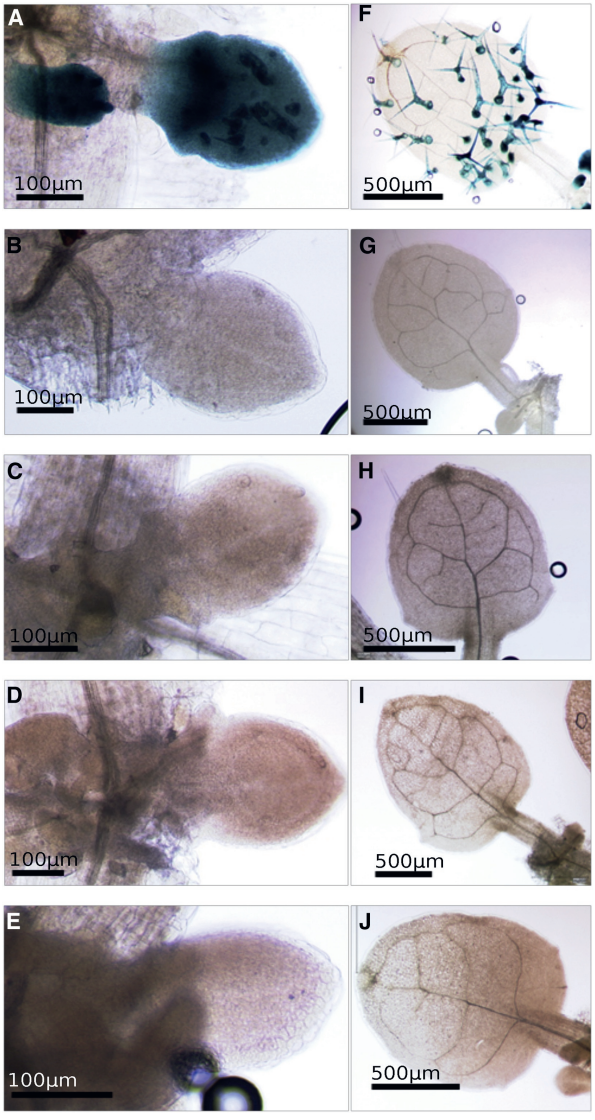


Figure 3.4: **Expression Pattern of pTRY:GUS in Different *ttg1* Alleles.** (A–J) Left column shows the expression pattern of pTRY:GUS in a young leaf (A–E). Right column presents the expression pattern of pTRY:GUS in a mature leaf (F–J). (A and F) Col-0. (B and G) *ttg1-9*. (C and H) *ttg1-11*. (D and I) *ttg1-12*. (E and J) *ttg1-13*.

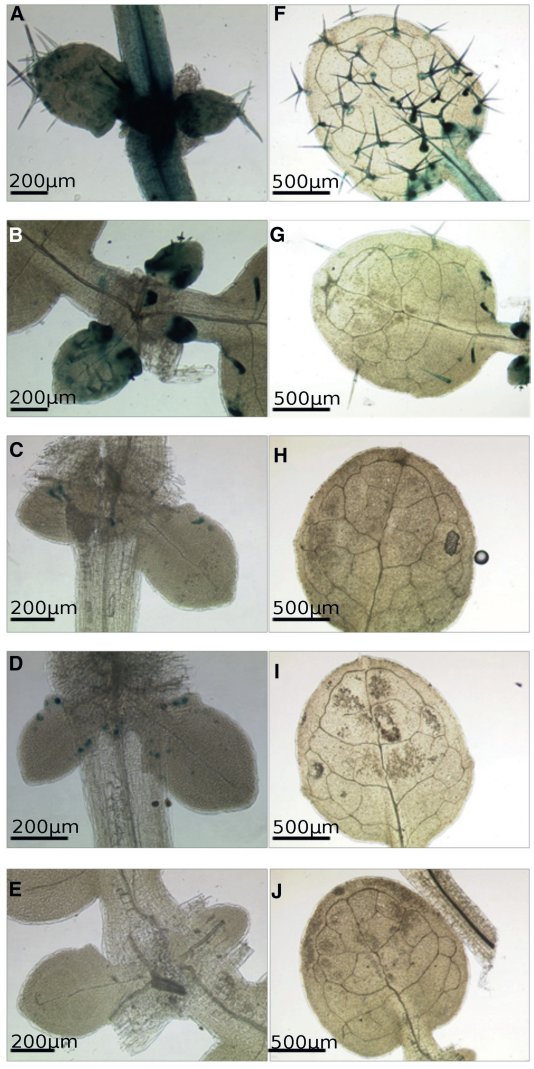


Figure 3.5: **Expression Pattern of pCPC:GUS in Different *ttg1* Alleles.** (A–J) Left column (A-E) shows the expression pattern of pCPC:GUS in a young leaf (A–E). Right column (F-J) shows the expression pattern of pCPC:GUS in a mature leaf (F–J). (A and F) Col-0. (B and G) *ttg1-9*. (C and H) *ttg1-11*. (D and I) *ttg1-12*. (E and J) *ttg1-13*.

differential regulation were the reason for the cluster phenotype one would expect that the expression of TRY under the CPC promoter could rescue the cluster phenotype in weak *ttg1* mutants. To test this, we generated transgenic *ttg1-9* lines expressing TRY under the control of 525 bp of the 5' upstream region of the CPC gene (pCPC:TRY [52]). We analysed the plants in the T1-generation to statistically cover the whole range of rescue phenotypes [52]. We observed a general reduction of the trichome number to about 38% of that observed in the single mutant *ttg1-9* line. This is consistent with the previous observation that pCPC:TRY expression in the try mutant reduces trichome number [52]. The cluster frequency was partially rescued. While *ttg1-9* plants exhibit a cluster frequency of 18%, the cluster frequency was significantly reduced to 4.6% in *ttg1-9* pCPC:TRY plants ($p < 0.01$, by Mann-Whitney U-test, Figure 3.6, Table S4). The pCPC:TRY expression in *ttg1-11* and *ttg1-12* alleles did not result in a reduced trichome number (Table S4). However, the cluster frequency was significantly reduced in pCPC:TRY *ttg1-11* and pCPC:TRY *ttg1-12* lines. In *ttg1-11* pCPC:TRY we found a cluster frequency of 2.8% while *ttg1-11* showed 7.0% clusters ($p < 0.01$ by Mann-Whitney U-test). In *ttg1-12* clusters were found with a frequency of 8.93% while in *ttg1-12* pCPC:TRY the cluster formation was significantly reduced to 3.1% (Figure 3.6). Although we never found a complete rescue these data indicate that the lack of TRY activation in weak *ttg1* alleles causes cluster formation.

3.3.6 Mathematical modelling of TTG1 dynamics in a combined AI-AD model

Weak *ttg1* alleles show three phenotypic aspects: First, reduced trichome density. Second, strongly enhanced trichome cluster formation. Third, a seemingly randomized trichome pattern. How are these three aspects theoretically related? To answer this question, we analysed hexagonal point patterns (Figure 3.7A). In a first step, we manipulated a regular hexagonal pattern by randomly perturbing the point coordinates such that the CV of the nearest neighbour distances matched the wild type pattern (Figure 3.7B). We assumed that the variation in spacing found in wild-type patterns is the effect of intrinsic fluctuations of cellular processes. These fluctuations result in small cell-to-cell differences in the young epidermal tissue, which can have a considerable effect on the final trichome pattern [32].

In a second step, we randomly removed points until the density matched the experimentally observed density of *ttg1-9* alleles (Figure 3.7C). This re-

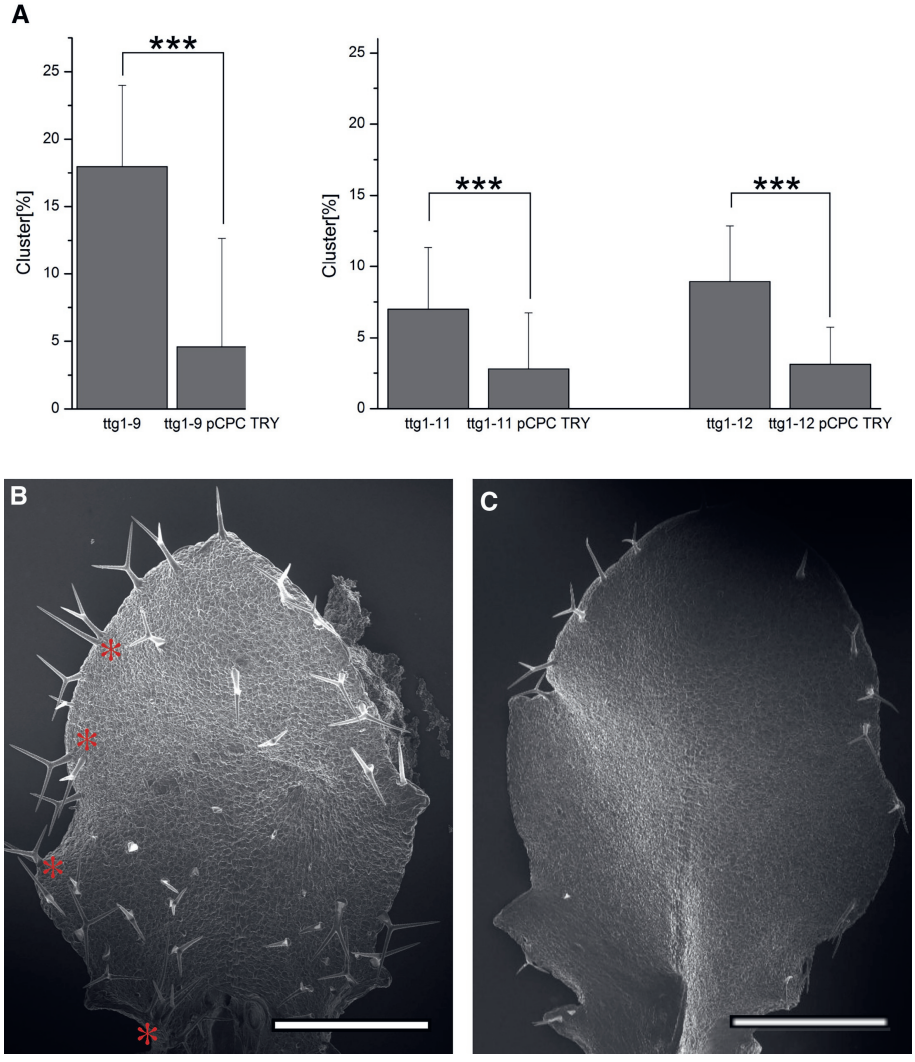
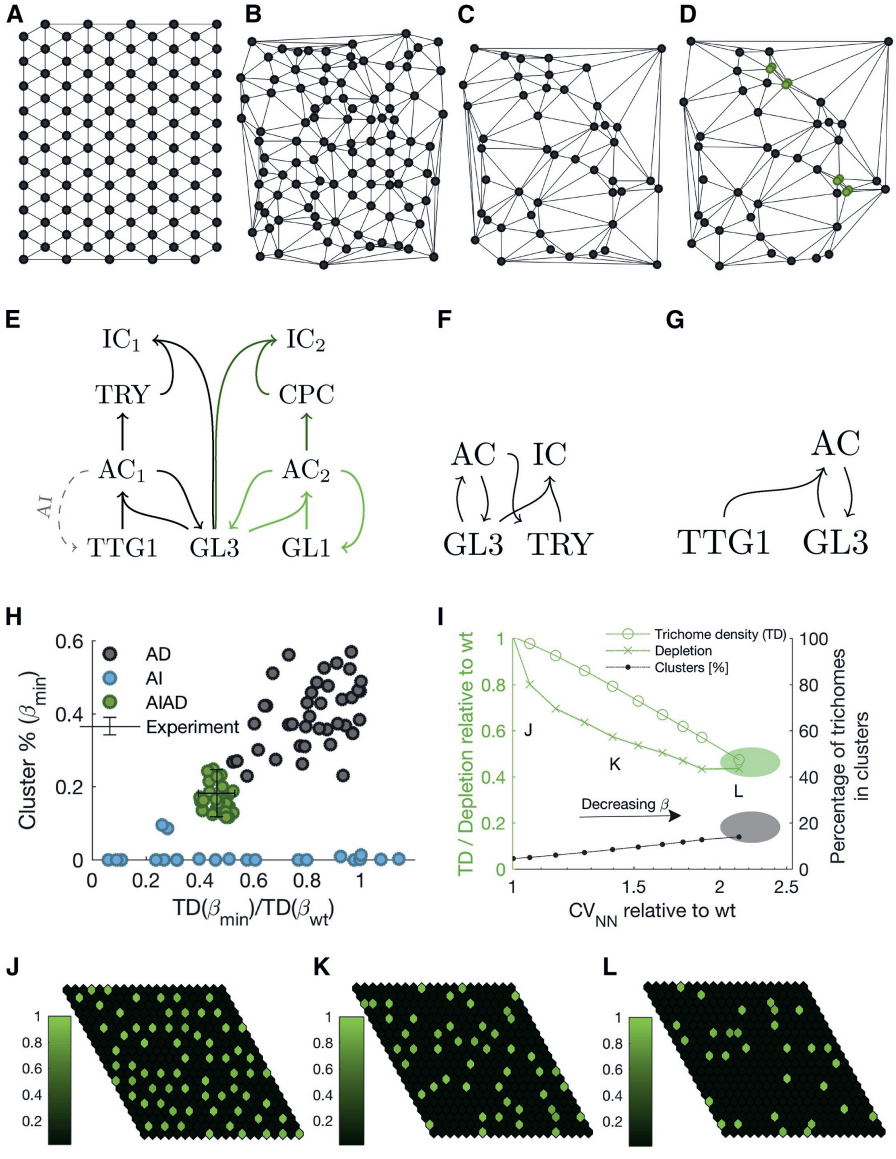


Figure 3.6: **Rescue of the Cluster Phenotype in Weak *ttg1* Mutants by pCPC:TRY.** (A) Diagrams showing the cluster frequency in *ttg1* alleles and the respective pCPC:TRY rescue lines. Three black asterisk indicate a statistically significant difference between the single mutants and the pCPC:TRY rescue lines ($p < 0.01$, Wilcoxon test). (B) Representative scanning electron microscope (SEM) image of a young *ttg1-9* plant. Red asterisks indicate the position of trichome clusters. (C) Representative SEM image of a *ttg1-9* pCPC:TRY plant. Scale bars, 400 *µm*.



(Caption on next page.)

Figure 3.7: **Analysis of Point Patterns and Development of Trichome Model Networks.** (A–D) Distribution of points on hexagonal grid. Lines between points are edges determined by Delaunay triangulation and indicate connectivity between points. Starting from a completely regular point distribution (A), point coordinates are perturbed (B), randomly removed (C), and clustered (D) (clusters indicated in green). The pattern in (D) represents a typical *ttg1-9* leaf. (E–G) Interaction scheme of the combined activator-inhibitor (AI) activator-depletion (AD) (AI-AD) model (E), the AI model (F), and the AD model (G). The dashed, gray edge in (E) indicates an additional feedback that can be used to convert the AI-AD to an AI-type network. TTG1 and GL3 form an active complex (AC) that activates TRY. In turn, TRY binds to GL3, thus forming an inactive complex (IC). The colors in (E) indicate the different additions needed to explain the *ttg1-9* phenotype. The black edges indicate the simplest form of the combination of the networks in (F) and (G). Light green indicates the edges needed for including GL1, and dark green edges are added upon CPC inclusion. For simplicity, basal production, degradation, and diffusion are not indicated. (H) Comparison of the AI (blue), AD (gray), and AI-AD (green) models. Each point is a different parameter set tested for the *ttg1-9* phenotype. The trichome densities are relative to wild type, where β_{min} is the strongest possible mutation for the TTG1-GL3 interaction parameter and β_{wt} is the value for the wild-type simulation. (I) Each point on the lines indicates a different relative change in binding strength (β) and the mean effect this has on the pattern in terms of relative trichome density (left y axis, circles), depletion (left y axis, crosses), percentage of clusters (right y axis), and coefficient of variation (CV) in nearest neighbor distances (x axis). From the leftmost point in the plot to the rightmost point, the TTG1-GL3 interaction is decreased, as indicated by the arrow. The green and gray shaded area indicates experimental data ranges for trichome density and cluster percentages on *ttg1-9* leaves, respectively. A shaded line indicates the 95% confidence interval determined by bootstrapping of a varying number of simulations; note that this area is smaller than the line width. The letters J, K, and L in the plot correspond to example simulations given below the plot. (J) Example simulation of a wild-type situation. (K) Example simulation that resembles the wild-type situation. In this simulation, the binding between TTG1 and GL3 is decreased, but not to such an extent that it replicates *ttg1-9* phenotypes. (L) An example of a *ttg1-9* simulation. In this case, the pattern shows clusters, reduced density, and irregularity, in the same relative quantities as observed experimentally.

sulted in an increased variability as measured by the nearest neighbour distribution (CV_{NN}). In a third step, clusters were introduced by moving randomly selected points towards each other, thus forming clusters (Figure 3.7D). This resulted in a CV matching the *ttg1-9* allele ($CV_{NN} = 0.85$), without introducing any additional noise on the resulting point pattern. Thus the pattern of the *ttg1-9* alleles only appears to be more random and the stochasticity underlying the trichome patterning process is not increased. Therefore, we do not take additional noise sources into account. Rather, as shown by the analysis of these point patterns, the changes of trichome density and clustering are sufficient to explain the full extent of the observed irregularity of the *ttg1-9* phenotype.

To gain mechanistic insight into the role of TTG1 in the patterning network we derived a minimal combined AI AD model (Figure 3.7E, black edges). In this model, TTG1 and GL3 form an active complex (AC). The binding of TTG1 to GL3 leads to a depletion around GL3 maxima. The AC formed by TTG1-GL3 activates both TRY and GL3. Based on the previously published models [38, 40] and new findings we used the following assumptions: (i) TTG1 and TRY are non-cell-autonomous [40, 150], GL3 is cell-autonomous [38]. (ii) TTG1 and GL3 form the AC [44], which activates its own inhibitor TRY and has a positive feedback-loop with its activator GL3 [38]. (iii) Inhibition is mediated by TRY binding to GL3 [38] or the GL3 TTG1 dimer [64]. The inhibited complex is explicitly modelled as a dimer (ID) or implicitly as a trimer (IT). (iv) Activation of GL3 and TRY by AC is modelled as activation by two AC units (i.e. a tetramer of TTG1-GL3-GL3-TTG1) [64]. This type of non-linearity is a requirement for pattern formation. (v) High concentrations of AC are considered to correspond to trichome cell-fate.

The model indicated in black edges in Figure 3.7E is the simplest combination of AI (Figure 3.7F) and AD (Figure 3.7G) patterning motifs and can be reduced to either an AI or AD network by cutting or adding one edge. In particular, adding activation of TTG1 by AC (dashed edge in Figure 3.7E) yields an AI network similar to previously published AI model (Digiuni et al., 2008) (Figure 3.7F). Removing the activation of TRY by AC results in the previously published AD model [40] (Figure 3.7G). To understand what is gained by the combination of the two motifs we studied how the two networks shown in Figure 3.7E-G perform in explaining the *ttg1-9* phenotype. As the parameters for the wild-type network are unknown we analysed 106 randomly generated parameter sets for each model using Quasi Monte-Carlo methods and selected for those able to generate a wild type pattern defined by the experimentally

observed trichome density and zero cluster frequency. Each of those parameter sets constitutes a possible wild type situation. Next, we analysed the networks by systematically reducing the binding strength of TTG1 to GL3 and analysing the changes in trichome density and cluster frequency.

We investigated the predicted phenotype for the lowest possible TTG1-GL3 binding strength, which still yields a trichome pattern, and compared the values with the experimentally observed *ttg1-9* phenotype. Typically, in the AI model the reduced binding strength almost never led to the formation of clusters (Figure 3.7H, blue solid circles). By contrast the AD model tends to generate too many clusters and higher trichome density than observed in *ttg1-9* mutants (Figure 3.7H, grey solid circles). Only the combined model was able to reproduce the experimentally observed trichome density and cluster frequency (Figure 3.7H, green solid circles). We therefore analysed the AIAD model in more detail. To understand how trichome density, cluster frequency and pattern randomness depend on the interaction strength of TTG1 to GL3, we quantified the three traits for successively decreasing interaction strength (Figure S4). As observed in the *ttg1-9* mutant, decreased interaction strength led to a reduced trichome density (0.46 ± 0.06 relative to wild-type), an increased apparent randomness ($CV_{NN,ttg1\text{-}9}/CV_{NN,wt} = 2.22 \pm 0.22$) and more clustering ($18\% \pm 7\%$).

Next, we analysed how well the combined model can explain the observed lack of depletion of TTG1-9 in the *ttg1-9* mutants for a range of reduced binding strengths. Unexpectedly, the model predicted a stronger depletion upon a reduction of the binding strength (Figure S4B). This behaviour does not depend on a specific parameter set, but rather is a generic feature of the AD motif (Figure S5). The failure of the model to reproduce the reduction of depletion can only be overcome by adding additional patterning elements not depending on TTG1. This type of TTG1-independent regulatory events are realistic given that the ttg1 mutant phenotype can be partially rescued by overexpression of GL3 or EGL3 ([44, 136]).

Following these considerations, we included GL1 into the network (Figure 3.7E, light green edges, Figure S6A). GL1 and GL3 can form an active complex, which activates TRY. The interactions included in this sub-part are based on the previously published AI model [38] (Figure 3.7F). Similar to the AI model tested in Figure 3.7H this model cannot reproduce the cluster formation, however, in contrast to the model without GL1, it is possible to find a reduced depletion of TTG1. The lack of cluster formation is likely due to the

activity of TRY, which is maintained throughout different mutation strengths by the GL3-GL1 dimer, thus consistently inhibiting neighbouring cells (Figure S6). To overcome this, we introduced the inhibitor CPC (Figure 3.7E, dark green edges). Through differential dimer formation by TTG1-GL3 and GL1-GL3 we modelled activation of the inhibitors by the different dimers (Figure 3.7G), where TRY is activated by TTG1-GL3 and CPC by GL1-GL3. This is a simplified form of the competitive binding model suggested previously [64]. Based on the *cpc* and *try* mutant phenotypes [7, 50] and the finding that CPC is more stable than TRY [151] we considered CPC to act on long-range distances and TRY to act on short-range distances. Thus, in this model, the loss of TRY through the TTG1 mutation results in a loss of short-range inhibition whereas CPC is still functioning at long ranges to maintain the trichome pattern. This model is able to predict the changes in trichome density, percentage of trichomes in clusters and the CV_{NN} for reduced TTG1-GL3 interaction strengths in the same ranges as experimentally observed (Figure 3.7I-L). Also the lower ranges of TTG1-GL3 interaction strength show a reduced amount of depletion compared to wild type. Futhermore, the difference between the inhibitor mobilities is found back in the parameter values used to simulate the *ttg1-9* phenotype (Figure S7A). Further support for this model is found in the analysis of the TTG1 knockout mutant. Our model predicts that the TTG1 knockout is glabrous and can be rescued by overexpression of GL3 (Figure S6C-D), which is in line with the experimental observations on these mutants [44, 136].

Our experimental data show that TRY expression is not seen in the *ttg1-9* mutants and indicate that the cluster phenotype can be rescued by trichome-specific expression of TRY under the CPC promoter. Consistent with this, our combined AIAD model predicts a reduced TRY expression when the interaction strength of TTG1 to GL3 is reduced (Figure S7B). We also simulated the *ttg1-9* pCPC:TRY rescue experiment by varying the parameters for the activation of TRY by the GL3-GL1 dimer. In these simulations we are able to show a partial to full rescue of cluster phenotype based on the change in GL3-GL1 dependent TRY activation (Figure S7C-D).

3.4 Discussion

Although it is well established that TTG1 plays a major role in trichome formation in *Arabidopsis thaliana*, its molecular function in trichome patterning remains elusive for at least two reasons. First, because it is a central com-

ponent in two patterning mechanisms, the activator-inhibitor and activator-depletion mechanisms, it is difficult to assess the biological significance in each of them [35]. Second, competitive complex formation of TTG1 - GL3 and GL3 – GL1 suggests that TTG1 has different roles in the activation of different MBW target genes [64]. Our analysis of weak *ttg1* alleles sheds some light on both aspects that will be discussed in the following.

3.4.1 Regulation of TRY by TTG1

The analysis of genetic interactions between TRY and TTG1 suggested a regulation of TRY by TTG1 [148]. Various try/+ ttg1/+ double-heterozygous combinations of try and different *ttg1* null-alleles exhibited clusters. Thus, the reduced activity of both genes is sufficient to cause a phenotypic effect indicating that they act in the same pathway. Recent molecular data suggest that TTG1 regulates MBW target genes - including TRY - in a differential manner in the root hair system [144]. One possible molecular explanation is the competitive complex formation of the MBW complex. It was reported that the activation of the TRY promoter by TTG1 and GL3 is counteracted by GL1 whereas the GL3 GL1 dependent activation of CPC promoter is repressed by TTG1 (Pesch et al., 2015). Consistent with the regulation scheme we observed a differential regulation of TRY and CPC in weak *ttg1* alleles. Our data can explain the observed lack of TRY activation in weak *ttg1* alleles by two related causes: First, the reduced interaction of mutant TTG1 to GL3 and second, the binding of GL1 to GL3 is not counteracted by mutant TTG1.

3.4.2 Role of TTG1 in the activator-depletion scenario

The activator-depletion model for trichome patterning is based on three experimental data sets: the interaction of TTG1 with GL3, the nuclear targeting of TTG1 by GL3 and the lack of depletion in the cells immediately next to incipient trichomes [40, 141]. We show that all three aspects are impaired in weak *ttg1* alleles. Thus, by all criteria, the activator-depletion mechanism is not operating in these alleles. This raises the question whether the randomization of the trichome pattern in weak *ttg1* alleles is caused by the lack of the activator-depletion mechanism. This is difficult to answer, as it is not clear whether or to what extent the activator-inhibitor mechanism is still operating. The mutual competition of TTG1 and GL1 for binding to GL3 [64] suggests that TTG1 GL3 and GL3 GL1 can act separately to transcriptionally activate inhibitory patterning genes. In this light one could postulate that GL3, GL1 and CPC can still function as an activator-inhibitor unit in weak *ttg1* alleles.

In fact, the results from the mathematical modelling suggest that GL3, GL1 and CPC are necessary to reproduce all observations.

3.4.3 Is the reduced interaction of TTG1 mutant protein with GL3 sufficient to explain the ttg1 trichome phenotypes?

When reducing the TTG1-GL3 interaction strength in the AIAD model we can robustly reproduce the patterning defects indicating that the proposed AIAD model is sufficient to explain the full spectrum of trichome defects in strong and weak *ttg1* mutants. A reduction of TTG1-GL3 interactions is predicted to cause reduced AC levels. This, in turn, leads to a lower activation of TRY. The reduced activation of TRY explains the increased cluster frequency.

One aspect that cannot be matched by the simplest form of the model (black edges in Figure 3.7E) is the loss of depletion in the ttg1 mutants. This stems from the characteristics of the AD sub-motif. The reduction in TTG1-GL3 interaction results in a focussing effect of TTG1, which counteracts a loss of depletion. This characteristic persists in the combined AIAD model and suggests that TTG1 independent regulation aspects are missing from the network. After the inclusion of GL1 into the model we can find reduction of depletion as a result of reduced TTG1-GL3 interaction, however, clusters are not formed in this model. Ultimately, the differential regulation of the inhibitors CPC and TRY was needed to find loss of depletion in combination with the patterning defects found for the *ttg1-9* allele. Here, two main requirements needed to be fulfilled. First, an additional motif needed to be introduced that could pattern independently of TTG1, this was achieved through the inclusion of GL1 (black and light green edges in Figure 3.7E). This second patterning mechanism increases pattern-forming robustness against the TTG1 mutation, resulting in a pattern that is not dependent on the depletion of TTG1 at its core. A model without GL1 would always need TTG1 depletion as a basis for forming a pattern (Figure S5). Second, CPC was introduced under the regulation of GL1-GL3 (complete network in Figure 3.7E). Without CPC, GL3-GL1 would still activate TRY despite changes in the TTG1 GL3 interaction strength, thereby maintaining high levels of TRY such that no clusters are formed (Figure S6B). Now, with the addition of CPC and differential dimer regulation, the reduction in TTG1-GL3 interaction leads to a reduced amount of TRY in the system, resulting in clusters. Furthermore, from the selected randomly generated parameter sets we find that TRY operates on shorter ranges than CPC (Figure S7A). This relatively short-range radius of effect ensures that clusters are formed when the reduction in TTG1-GL3 in-

teraction strength results in reduced amounts of TRY.

We combined the AI and AD spatial patterning motifs to explain a complex patterning phenotype as a result of a point mutation. To overcome the challenge of incomplete information about the system parameters we adopted a constraint global analysis method using a Quasi Monte Carlo approach. For every randomly chosen parameter set the model had firstly to reproduce the observed wild-type pattern and, subsequently, had to replicate the complex mutant phenotype. This approach revealed that neither the AI nor the AD motif is structurally capable to explain the observed data. Moreover, by seeking for a minimal model our approach disclosed that the combination of the AD and the AI motif plus an extension by an extra inhibitor is required to capture the observations. Our combined and extended AIAD model provides a link between geno- and phenotype and offers a consistent explanation how a point mutation can result into reduced trichome density and increased cluster frequency at the same time.

3.5 Methods

3.5.1 Analysis of fluorescence intensity in trichomes and surrounding tiers

Stable lines expressing TTG1 tagged with YFP and the three TTG1 mutant proteins tagged with YFP under the control of the TTG1 promoter were analysed using confocal laser scanning microscopy. The DM6000 CS Microscope was used in combination with the TCS-SP8 imaging system (Leica Microsystems, Heidelberg, Germany). Z-stacks of young trichomes and surrounding tiers were acquired with a plane thickness of around 1-1.5 μm using the 20x water immersion objective. Determination of fluorescence intensity was achieved by using the software ImageJ (Fabrice Cordelieres, Institute Curie, Orsay, France). Maximum projections of the planes displaying the trichome and the surrounding tiers of cells were created, and fluorescence intensity (mean grey value) was measured in manually placed ROIs (region of interest). For each trichome three elliptical ROIs were chosen (Figure S3A). In the epidermal cells three polygonal ROIs were selected in each tier in the cytoplasmic regions (Figure S3A). The mean fluorescence for the trichome and each of the three tiers were calculated. Trichome fluorescence intensity was set to 100% and for each tier the fluorescence percentage compared to the trichome intensity was calculated. Subsequently for each YFP-tagged TTG1 allele the mean percentage of the 1st, 2nd and 3rd tier was calculated as well as the standard

deviations (TTG1 n=12, TTG1-9 n=19, TTG1-11 n=28, TTG1-12 n=15). The data were tested for normal distribution using the Kolmogorov-Smirnov test ($\alpha = 0.05$) followed by a one-sample t-test ($p = 0.002$). All calculations were performed using Microsoft Excel 365.

3.5.2 Light Microscopy

To observe the root phenotypes of wild type (Col-0), *ttg1-1*, *ttg1-9*, *ttg1-11* and *ttg1-12*, seeds were sterilized and sown on ½ MS plates with 1% sugar. After 2 days of stratification the plates were transferred into a growth chamber (22 C, light/dark cycle of 16/8 h, humidity of 60%) and placed vertically. 2-3 days later the seedlings were transferred to a microscopic slide wrapped with parafilm on both ends so that the seedlings all had the cotyledons on one long side of the slide. Liquid 1/10 MS with 1% sugar was added to the roots and the plants, except for the leaves, were covered with a cover slide. The slides were placed vertically into a glass box filled with some 1/10 MS with 1% sugar and stored in the growth chamber. After one day the roots were examined using the DM5000 B and the DFC360 FX imaging system (Leica Microsystems, Heidelberg, Germany). Seed colour was examined using a LEICA MZ16 F and documentation was performed by using the LEICA DFC420 C imaging system. To analyse seed coat mucilage production by seed epidermal cells, a ruthenium red staining was performed. Seeds were hydrated in water for 5 min under gentle shaking. After removing the water, ruthenium red solution (0.1 mg/ml) was added. After 5 min of incubation the seeds were washed twice with water [152] and examined using a LEICA MZ16 F and documentation was performed by using the LEICA DFC420 C imaging system. Trichome coordinates of Col-0 and *ttg1-9* were extracted using TrichEratops [145]. GUS analysis was performed as described previously [153]. Light microscopy was done either using a Leica MZ16F binocular microscope or a Leica DMRE microscope equipped with a high resolution KY-F70 3-CCD JVC camera and DISKUS software. Confocal laser scanning microscopy was performed on Leica TCS-SP2 imaging system (Leica Microsystems) equipped with LCS software. 40x water immersion objective was used to obtain the z-stack images. z-stack images were merged to obtain the image in one plane. YFP fluorescence quantification was done using the histogram quantification tool in LCS software.

3.5.3 Yeast two hybrid and nuclear transport assays

For protein-protein interaction in yeast, pAS2 and pC-ACT2 plasmids (Clontech) were used fort translational fusion of proteins either with GAL4 activation or DNA binding domain respectively. TTG1, TTG1-13, TTG1-9, TTG1-11 and TTG1-12 were cloned as a fusion to DNA binding domain in the pAS2 vector and GL3, EGL3 and TT8 were cloned as a fusion to the GAL4 activation domain by LR clonase reaction. Saccharomyces cerevisiae strain AH109 was transformed as described previously for the interaction assay [154]. Yeast Nuclear Transportation Trap (NTT) assay was performed in yeast strain EGY48 as described before [147]. NTT constructs pNH2 (NES-LexAD), pNS (modified pNH2) [147], pNS-TTG1, pNS-TTG1-13 and pVT-U-GL3 [141] have been described before. pNS-TTG1-9, pNS-TTG1-11 and pNS-TTG1-12 were created by cloning SalI and XhoI fragments from corresponding entry vectors into the SalI site in pNS vector to obtain pNS-TTG1-9, pNS-TTG1-11 and pNS-TTG1-12. Yeast harbouring pNS and pVT-U plasmids grow on synthetic dropout media lacking histidine and uracil respectively. Transport of protein of interest as a translational fusion to NES-LexAD into the nucleus results in activation of rhw leucine reporter gene under GAL4 promoter and is read out by growth of yeast cells on the synthetic dropout media lacking amino acid leucine.

3.5.4 LUMIER

Two destination vectors were used for LR reactions. pcDNA3-Rluc-GW and pTREXdest30 (Invitrogen) enable the N-terminal fusion of Renilla reniformis and Staphylococcus aureus proteins, respectively [60]. GL3 was N terminally fused to the *S. aureus* ProtA sequence in pTREX-dest30-ntPrA. As a negative control, the vector pTREX-dest30-ntPrA was recombined with pENTR1A-w/ o-ccdB. *R. reniformis* Luciferase-TTG1-wt, Luciferase-TTG1-9, Luciferase-TTG1-11 and Luciferase-TTG1-12 were created as N terminal fusions in pcDNA3-Rluc-GW. pENTR1A-w/o-ccdB was used as a negative control. For LUMIER assays, each protein was transiently expressed in HEK293TN cells (BioCat/SBI; LV900A-1). Transfection and pull down assay were done as described before [60, 64] three times independently each as two technical replicas.

3.5.5 Constructs

All entry clones were in pENTRY1A/pENTR4. TTG1-YFPpEN [40], TTG1ΔC26-YFPpEN where carboxy terminal 26 amino acids are deleted [141] and TTG1pEN [64] have been described previously. TTG1-9-YFPpEN, TTG1-11-YFPpEN for

ttg-11 and TTG1-12-YFPpEN were generated by site directed mutagenesis using TTG1-YFPpEN as a template. Entry clones of all the ttg1 allelic versions without YFP fusion were also generated by site directed mutagenesis using TTG1pEN as a template. For plant transformation all the TTG1-YFP versions were cloned into pAMPAT-pTTG1-GW binary vector [40] by LR clonase reaction to express them under native TTG1 promoter.

3.5.6 Generation of a random pattern

We compare the data from leaves to a suitable random reference. There are two requirements the random patterns should fulfil. First, the points should be independent and identically distributed, i.e. we assume complete spatial randomness. Seconds, the amount and variation in density of points per representation should reflect the observations for wild-type and *ttg1-9* phenotypes. For these reasons we simulate a homogeneous Poisson point process [155]. We generate 104 random representations for both wild-type and *ttg1-9*. For each of these representations the points density is sampled from a Poisson distribution using the mean trichome density determined from either the wild-type or *ttg1-9* leaves as a parameter for the distribution. The points for each representation are uniformly and independently placed within the unit circle.

3.5.7 Quantification of the regularity of patterns

To quantify the regularity of the trichome patterns we focus on two measures. The first measure is the coefficient of variation of nearest neighbour distances. For each leaf (wild-type and ttg1-9) we calculate the Euclidian distances δ_i for each trichome t_i to its nearest neighbour t_n by

$$\delta_i = \sqrt{(x_i - x_n)^2 + (y_i - y_n)^2} \tag{3.1}$$

where (x_i, y_i) are the coordinates for the trichome t_i and (x_n, y_n) the coordinates for its nearest neighbour t_n. Next, we use the coefficient of variation (ratio of the standard deviation of the distances δ^σ to the mean $\bar{\delta}$) of the nearest neighbour distances which is defined by

$$\mathrm{CV_{NN}} = \frac{\delta^\sigma}{\bar{\delta}} \quad \text{with} \quad \bar{\delta} = \frac{1}{|\mathcal{J}_k|}\sum_{i=1}^{|\mathcal{J}_k|}\delta_i \quad \text{and} \quad \delta^\sigma = \sqrt{\frac{1}{|\mathcal{J}_k|-1}\sum_{i=1}^{|\mathcal{J}_k|}(\delta_i - \bar{\delta})^2} \tag{3.2}$$

where $\mathcal{J}_k$ is the set of trichome coordinates for a certain leaf k and $|\mathcal{J}_k|$ its cardinality. The $\mathrm{CV_{NN}}$ reported for wild-type and *ttg1-9* is the mean of $N = 9$

and $N = 11$ leaves respectively.

In addition to the CV_{NN} we use the mean anisotropy as a measure of pattern regularity. The anisotropy has been shown to be a suitable measure of the local trichome environment [32]. Following the procedure from Greese et al. [32], we determine the ratio of eigenvalues of the inertia tensor for each trichome. We report the average of these ratios for each leaf and averaging again over all leaves (or computer-generated patterns).

3.5.8 Theoretical probability of finding clusters in random patterns

For a homogeneous Poisson process the probability to find n points per unit area is given by the Poisson distribution [156]

$$P(n) = \frac{m^n e^{-m}}{n!} \tag{3.3}$$

where m is the mean number of points per area. Because we consider a homogeneous process the probability is independent of the specific location of area of interest. If ρ is the mean density of the distribution, then $m = \pi r^2 \rho$, which upon substitution into (3.3) gives

$$P(n) = \frac{\pi r^2 \rho^n e^{-\pi r^2 \rho}}{n!}. \tag{3.4}$$

The probability of the chosen area πr^2 containing no points is

$$P_0(r) = P(n = 0) = e^{-\pi r^2 \rho} \tag{3.5}$$

which is the probability that the area within a distance $r \geq 0$ contains no points. Consequently, the probability of finding at least one point within a radius r is given by

$$P_{\geq 1}(r) = 1 - e^{-\pi r^2 \rho}. \tag{3.6}$$

The relation given in (3.6) is visualized in Figure S2A for densities corresponding to wild-type and *ttg1-9*. For a random pattern of lower density, the probability of finding a cluster is lower than for a pattern of higher density.

3.5.9 Correlation dimension of experimental data and simulated patterns

The correlation dimension can be used to evaluate the number of factors that are involved in the variability of a process [157, 158]. In this case we are

interested in the application of the correlation dimension in spatial patterns. For the trichome and randomly generated patterns we use the Grassberger-Procaccia algorithm to extract the correlation dimension from the data [159]. The spatial coordinates are given in two dimensions; therefore, the maximum correlation dimension is two. This maximum dimension is found back in the analysis of the random data, showing that there is no underlying deterministic factor in determining the pattern (Figure S2C-F). For both the wild-type and the *ttg1-9* data we find correlation dimensions < 2, indicating that there is some inherent deterministic mechanism operating in the patterning of both phenotypes.

3.5.10 Reproducing trichome pattern irregularity by perturbing hexagonal patterns

To reflect the noise level of nearest neighbour distances of trichomes seen in wild-type ($CV_{NN} = 0.33$), we apply a random perturbation to the coordinates of a completely regular hexagonal pattern consisting of N = 100 points. Considering that wild-type phenotypes do not show any clusters of trichomes, we include an exclusion zone around every point. This exclusion zone is achieved by defining a region around a certain point with a radius z. The algorithm for generating the perturbed pattern is the following:

1. Generate $u_1, u_2, \ldots, u_N \sim \mathcal{U}(0,1)$ independently.
2. Set $R_1 \leftarrow r\sqrt{u_1}$, $R_2 \leftarrow r\sqrt{u_2}, \ldots, R_N \leftarrow r\sqrt{u_N}$.
3. Generate $u_{N+1}, u_{N+2}, \ldots, u_{2N} \sim \mathcal{U}(0,1)$ independently.
4. Set $\Theta_1 \leftarrow 2\pi u_{N+1}$, $\Theta_2 \leftarrow 2\pi u_{N+2}, \ldots, \Theta_N \leftarrow 2\pi u_{2N}$.
5. Select points (x_i, y_i) by a random permutation on the integers $I = \{1, \ldots, N\}$.
6. Set $x_1 \leftarrow x_1 + R_1 \cos\Theta_1$, $x_2 \leftarrow x_2 + R_2 \cos\Theta_2, \ldots, x_N \leftarrow x_N + R_N \cos\Theta_N$.
7. Set $y_1 \leftarrow y_1 + R_1 \sin\Theta_1$, $y_2 \leftarrow y_2 + R_2 \sin\Theta_2, \ldots, y_N \leftarrow y_N + R_N \sin\Theta_N$.
8. For steps 6 and 7 check if the nearest neighbour distance from point $i \in I$ is larger than $z = 0.4$, else point (x_i, y_i) is not shifted.

This exclusion zone in step 8 of $z = 0.4$ is arbitrarily chosen, but sufficient to reproduce a CV_{NN} corresponding to measurements of wild-type leaves.

Next, we generated a vector of points by drawing from ${}^nP_r$, with $n = 100$ and $r = 54$. The coordinates corresponding to the integers in this vector are

removed from the grid, yielding a density relative to the wild-type grid of 0.46. Finally, we induce clustering by first selecting 8 points out of the 46 remaining after sparsening by random permutation. Next, the selected points are split into two groups of 4 points. The coordinates of the second group are set to the coordinates of the first group plus a value smaller than the minimal nearest neighbor distance found in the previous step. This set of points is then counted as clusters, giving a density of 8 out of 46 (18%), comparable to experimental observations. The pattern showing clusters and a reduced density has a CV_{NN} similar to the *ttg1-9* leaves ($CV_{NN} = 0.85$). Edges between points are determined by MATLAB's built-in Delaunay Triangulation function.

3.5.11 Trichome patterning model

The cells are modelled on a hexagonal grid of N_x by N_y cells with periodic boundary conditions based on a modelling framework described before [38, 40]. Transport of species χ between cell j at coordinates (x, y), where $1 \leq x \leq N_x$ and $1 \leq y \leq N_y$ and its neighbour is modelled by the coupling equation

$$\begin{aligned}\hat{L}[\chi]_{x,y} =& [\chi]_{y-1,x} + [\chi]_{y+1,x} + [\chi]_{y,x-1} + [\chi]_{y,x+1} \\ &+ [\chi]_{y+1,x-1} + [\chi]_{y-1,x+1} - 6[\chi]_{y,x}.\end{aligned} \tag{3.7}$$

Based on the network presented in Figure 3.7E, we used the following system of dimensionless coupled ordinary differential equations (ODEs) to describe the change over time of TTG1, GL1, GL3, TRY, CPC and the active complex between TTG1-GL3 (AC1) and GL1-GL3 (AC2):

$$\partial_t[TTG1]_j = k_1 - [TTG1]_j(k_2 + k_3[GL3]_j) + k_2k_4\hat{L}[TTG1]_j \tag{3.8}$$

$$\partial_t[GL1]_j = k_5 + k_6[AC2]_j - [GL1]_j(k_7 + k_8[GL3]_j) \tag{3.9}$$

$$\begin{aligned}\partial_t[GL3]_j =& k_9 + \frac{k_{10}k_{11}[AC1]_j^2}{k_{11} + [AC1]_j^2} + \frac{k_{12}k_{13}[AC2]_j^2}{k_{13} + [AC2]_j^2} - \\ & [GL3]_j(k_{14} + k_3[TTG1]_j + k_8[GL1]_j \\ & + k_{15}[TRY]_j + k_{16}[CPC]_j)\end{aligned} \tag{3.10}$$

$$\begin{aligned}\partial_t[TRY]_j =& k_{17}[AC1]_j^2 + k_{25}[AC2]_j^2 - [TRY]_j(k_{18} + k_{15}[GL3]_j) + \\ & k_{18}k_{19}\hat{L}[TRY]_j\end{aligned} \tag{3.11}$$

$$\begin{aligned}\partial_t[CPC]_j =& k_{20}[AC2]_j^2 - [CPC]_j(k_{21} + k_{16}[GL3]_j) + \\ & k_{21}k_{22}\hat{L}[CPC]_j\end{aligned} \tag{3.12}$$

$$\partial_t[AC1]_j = k_3[GL3]_j[TTG1]_j - k_{23}[AC1]_j \tag{3.13}$$

$$\partial_t[AC2]_j = k_8[GL3]_j[GL1]_j - k_{24}[AC2]_j \tag{3.14}$$

The first version of the model consisted only of (3.8), (3.10), (3.11) and (3.13) and a smaller amount of the relevant parameters, giving the network of black edges only in Figure 3.7E. This was extended later by including (3.9) and (3.14); this is the combination of black and light green edges in Figure 3.7E. Finally, CPC was included through (3.12), resulting in the complete network. Note that parameter k_{25} is generally set to 0 except in the pCPC:TRY rescue simulation where $k_{25} > 0$ (Figure S7D).

3.5.12 Pattern quantification

Similar to the hexagonal point pattern analysis, we determined the coefficient of variation of the nearest neighbour distances (CV_{NN}), cluster density and trichome density for the model simulations. In order to calculate any of the quantities we first identified the trichome cells. Towards that end, the steady state concentrations of the active complex $[AC]_{(x,y)}$ are normalized by the maximum observation $[AC]_{max}$; cells that contain more than half-maximum of AC are classified as trichomes. In the model with both the GL3-TTG1 (AC1) and GL3-GL1 (AC2) active complex the sum of both complexes is used, i.e. $[AC]_{(x,y)} = [AC1]_{(x,y)} + [AC2]_{(x,y)}$. Using these criteria we can identify the set of grid-coordinates at which trichomes can be found, formally defined as $\mathcal{T} = \{(x,y)|[AC]_{(x,y)} \geq \frac{1}{2}[AC]_{max}\}$. The number of elements in this set, i.e. its cardinality $|\mathcal{T}|$, equals the number of trichomes on the grid. By dividing the cardinality by the total number of cells (i.e. grid size) we obtain the trichome density. Next, we determined the cluster frequency by identifying the elements with coordinates that are next to each other on the grid and by dividing that number by $|\mathcal{T}|$. Finally, we determined the nearest neighbour distances. For this we considered the trichome cells as a point pattern on a hexagonal grid. With the set of coordinates of trichomes on this grid we used MATLABs knnsearch with default settings. From this distribution of distances we determine the CV_{NN}, as described above.

3.5.13 Parameter search

As all parameters for the system are unknown, we used Latin Hypercube sampling to study system dynamics at different points in the parameter space. Within the parameter space there exists a sub-space where Turing patterning can occur, called the Turing Space. In this domain a diffusion-driven insta-

bility (Turing instability) can occur resulting in an inhomogeneous patterning state [3]. To test this for our randomly generated sets we use linear stability analysis where the stability of a uniform steady state is verified by determining whether effects of small perturbations to the ODE system decay over time. Turing instability was tested by the following criteria: starting from a uniform steady state (i) the steady state in the absence of diffusion is stable and (ii) the steady state in the presence of diffusion is unstable [123]. For criterion i this means that all eigenvalues of the Jacobian evaluated at steady state must be negative. To perform the same test for criterion ii we decoupled the system by Fourier transformation and analysed the eigenvalues [38, 40], where the real part of at least one of the eigenvalues must be positive.

3.5.14 Parameter criteria

To test in the mathematical model whether the patterning defects can be explained by a reduced binding of TTG1 to GL3 (β) this parameter is varied while other factors are kept constant.

In a random parameter search, the parameter sets are tested for (i) Turing instability, (ii) increase in cluster densities, (iii) relative increase in CV (iv) a relative decrease in trichome densities. Every quantity is fitted to the following experimentally determined ranges: trichome density of 0.46 ± 0.06 relative to wild type, CV of 2.22 ± 0.22 relative to wild type and a cluster density of 0.18 ± 0.07.

After an initial search of 10^6 randomly generated parameter sets we found 4 sets that fulfilled these criteria. These sets were then used to define a local search area to speed up the parameter search. With this approach we found 40 parameter sets that were used to study the pattern development for decreasing TTG1-GL3 interaction strengths. For parameter sets that match the criteria, the quantitative factors, e.g. cluster density, were averaged across multiple simulations with randomly perturbed initial conditions until convergence. Convergence is defined as $\sigma/\sqrt{N} < \epsilon\theta(k)$ where σ is the standard deviation of a pattern quantity, N the number of simulation repeats, ϵ a measure of accuracy, which we set to 0.05, and $\theta(k)$ the mean of a pattern quantification (e.g. trichome density) for parameter set k.

3.5.15 Depletion in the activator-depletion model

For the simplest combination of the AI and AD model networks, we found that loss of depletion in the ttg1-9 mutant simulations was not observed, in fact we found a focussing effect where more TTG1 was recruited to trichome cells. To facilitate the analysis we focused on the behaviour of the AD submotif. As expected, reduced binding strength resulted in higher amount of free TTG1 (Figure S5B), which in turn results in a higher effective TTG1 mobility (Figure S5A). The number of peaks (i.e. trichomes) was decreased and the peak height (i.e. AC levels in trichomes) was increased (Figure S5B). As a consequence, also TTG1 was more strongly focussed in the trichomes.

We further sought to gain insight into the underlying reason of the focussing effect. A reduction of the interaction rate between TTG1 and GL3 results in an enhanced number of unstable Fourier modes (Figure S5C) for the initial Turing instability (see section on 'Dispersion relation' below for methods). Due to this, only the major peaks of the stochastic initial perturbation grow into stable large peaks. Because TTG1 is less bound in complexes in non-peak cells, more TTG1 can be recruited to the developing peaks. This results in low trichome density with enhanced peak height and strong depletion of TTG1 around peak cells. This is contrasted by the wild-type situation where only very few Fourier modes are unstable (Figure S5B). From the stochastic initial perturbation only the high frequency modes are selected while most Fourier modes are damped. This means that the AD motif acts as a spatial high-pass filter.

3.5.16 Dispersion relation

When the binding strength between TTG1 and GL3 is decreased less AC is formed. This observation is confirmed by the model and is one of the possible reasons why the trichome density is decreased for the mutants. To determine the effect of the change in binding strength on the pattern formation, the dispersion relation is studied for different binding strengths on a 1-dimensional grid of N cells. The Fourier modes for a 1D grid are given by

$$k^2 = 4sin^2\left(\frac{\pi n}{N}\right) \tag{3.15}$$

where $1 \leq n \leq N$, and

$$\mathbf{A}_{p,q} = \mathbf{J}_0 - \mathbf{D}k^2 \tag{3.16}$$

where $\mathbf{J}_0$ is the Jacobian evaluated at steady state and $\mathbf{D}$ is the diffusion matrix. The eigenvalues for the matrix $\mathbf{A}_{p,q}$ are determined for different wavenumbers and different values for the binding strength parameter. Given that a pattern is only formed when $\mathrm{Re}(\lambda) > 0$ the dispersion relation shows which wavenumbers correspond to unstable modes and are thus a possible component of the pattern.

3.5.17 Quantification and Statistical Analysis

For the statistical analysis of the difference of the mean intensity in trichomes and the three neighbouring epidermal cell tiers the data were first tested for normal distribution using the Kolmogorov-Smirnov test ($\alpha = 0.05$) followed by a one sample t-test ($p = 0.002$). All calculations were performed using Microsoft Excel 365. For testing the statistical difference in pattern irregularity for the mean CV_{NN} and anisotropy Mann-Whitney U-test was used ($\alpha = 0.01$). The single-tailed test was used to test whether the irregularity measures were greater for random patterns than wild-type and whether *ttg1-9* patterns were more irregular than wild-type. In comparing *ttg1-9* to random patterns we tested whether *ttg1-9* was more irregular than the random reference patterns.

3.6 Author contributions

R.B., H.B., B.Z, S.H. B.J. and M.P. performed the experiments and analysed the data; A.D. and C.F. did the mathematical modelling and analysis. M.H. supervised the experiments. M.H., A.D., C.F. wrote the manuscript.

3.7 Supplementary Information

Supplementary information for this chapter is available on `https://doi.org/10.1016/j.celrep.2020.108497`.

Chapter 4

Quantitative identification of the MBW-complex binding behaviour in trichome patterning through a ratiometric approach

Anna Deneer*, Bipei Zhang*, Christian Fleck and Martin Hülskamp
* Contributed equally

4.1 Abstract

The proteins in the MYB-bHLH-WDR (MBW) complex are well conserved in higher plant species and are involved in a wide range of developmental and metabolic processes. Here, we focus on their role in trichome patterning and aim to quantify the interactions between the R2R3MYB protein GLABRA1, the WD40 protein TRANSPARANT TESTA GLABRA1, the bHLH protein GLABRA3 and both the R3MYB proteins TRIPTYCHON and CAPRICE. These proteins form the core of the network regulating trichome patterning and their binding behaviour is what drives downstream processes like cell differentiation. We quantified the dissociation constants for each of the interactions with the unifying bHLH protein through a ratiometric approach and show that the R2R3MYB and WD40 have similar binding strenghts while exhibiting negative cooperativity. Furthermore, we find that the R3MYB proteins show weaker binding and suggest that their mode of competitive binding goes beyond a displacement mechanism. Finally, through mathematical modelling, we predict an increased robustness in patterning as a consequence of these binding properties, indicating a relevant mechanism for trichome patterning.

4.2 Introduction

The MBW complex consisting of MYB and bHLH transcription factors associated with a WD40 repeat protein drives multiple traits in a range of plant species, among which metabolic pathways and cell differentiation [43, 48, 65–68]. Mutation and duplication events in the genes encoding the MBW proteins have given rise to a wide array of developmental regulatory mechanisms that provide the flexibility needed to generate the different epidermal cell types found in plants [65, 66, 160, 161]. In *Arabidopsis thaliana* the MBW complex is involved in anthocyanin biosynthesis, seed coat mucilage, seed coat pigmentation, and trichome and root hair patterning [65, 68].

An excellent system in which to study epidermal cell differentiation and patterning is that of trichome formation [7, 9, 35]. In this system the WD40 protein is encoded by TRANSPARENT TESTA GLABRA1 (TTG1) [43, 162, 163] which associates with bHLH proteins, among which GLABRA3 (GL3) [44, 136, 139, 160]. The bHLH proteins are known to homodimerize or heterodimerize, possibly leading to higher order complexes [44, 69, 136, 164–168]. The third component of the complex is a MYB protein [45, 134, 136], which in the case of trichome differentiation can be classified into two groups based on their function discovered in mutant phenotype studies [31, 34, 38, 40, 50]. In the first group, the activators, are the R2R3MYB transcription factors, among which is GLABRA1 (GL1) [45, 134, 169, 170]. The second group, the inhibitors, consists of R3MYB transcription factors [50, 52, 56, 148, 171]. Among the inhibitors are TRIPTYCHON (TRY) and CAPRICE (CPC) [34, 50, 51, 56, 139].

The interaction between the MBW components and the balance of their quantities are what drives the pattern formation process of trichomes [61, 62, 64, 172–176]. Gene regulation in developing trichomes is controlled by both MBW quantity and composition [64]. The interactions between the WD40, MYB and bHLH proteins have been studied in different settings [38, 43, 44, 60, 136]. In yeast two-hybrid experiments TTG1 and GL1 were both found to interact with GL3 but TTG1 and GL1 do not interact directly [44, 64], leading to the assumption that GL1, GL3 and TTG1 form a trimeric complex together, capable of activating trichome differentiation events [44]. It was later shown in yeast three-hybrid experiments and pulldown assays that GL1 and TTG1 counteract each other's binding to GL3 [64], which led to a new model of differential dimer formation of GL3-GL1 and GL3-TTG1 dimers. Furthermore, these different dimers show a different extent of effect on the activity of the

target promotors of TRY and CPC [64], indicating that the differential complex formation could be a mechanism of control, depending on the ratio of the dimers or higher order complexes. An additional mechanism of interaction is that between the different MYB proteins. The R3MYB proteins compete with R2R3MYB for binding to GL3 [64]. Thus, there are two mechanistically different types of competition: 1) in the competition between WD40 and R2R3MYB a possible allosteric regulation counteracts the other component from binding to the bHLH protein, 2) between R2R3MYB and R3MYB proteins, where repression of the complex activity is achieved by removal of R2R3MYB by R3MYB.

From our current understanding the ratios of the MBW proteins and the resulting composition of the MBW complex regulates the activity of different promoters [61, 62], which ultimately control the spatial and temporal distribution of specialized cell types [65, 66]. Here we define the complex interactions through a quantitative analysis of binding experiments. We first aim to identify and compare the dissociation constants of the WD40 and MYB proteins to the bHLH protein. The dissociation constants are quantified through LUMIER assays, which show that TTG1 and GL1 have similar binding properties and bind stronger than TRY and CPC to GL3. Additionally, the dissociation constants of TRY and CPC lie within very similar ranges of each other. Next, effects of competition between TTG1 and GL1 as well as between GL1 and the inhibitors TRY and CPC are analysed through competition experiments. For TTG1, GL1 and GL3 we find that GL3 homodimerization plays a significant role, together with a mechanism of negative cooperativity between TTG1 and GL1. In the case of the competition between GL1 and TRY/CPC we find strong competitive effects in the form of competitive displacement, in combination with the binding of the inhibitors to GL1 which further increases the competitive effect. Taken together, we find two very different mechanisms of complex formation, both having different effects on the complexes formed at equilibrium which provides insight into how both mechanisms can control different aspects of trichome differentiation events.

4.3 Results

4.3.1 Dissociation constants of TTG1-GL3 and GL1-GL3 dimers are in a similar range

To determine the relative amounts of the different complexes that can be formed by TTG1, GL3 and GL1, we aimed to determine the dissociation con-

stants of the TTG1-GL3 and GL1-GL3 dimers. This requires a quantitative analysis of the protein concentrations at different ratios. As we have not been able to produce soluble TTG1, GL3 or GL1 proteins in sufficient amounts for a biochemical analysis, during the last years we developed an alternative approach using LUMIER pulldown assays (LUminescence-based Mammalian IntERactome [177]). This approach has proven to be very successful for the analysis of interactions between TTG1, GL3 and GL1 proteins [64]. ProtA-tagged GL3 and Renilla-tagged GL1 or TTG1 were expressed separately in human HEK293TM cells (Figure 4.1A), raw extracts were mixed and subjected to pulldown assays. Because we could not determine the absolute protein concentrations of GL1, GL3 and/or TTG1, we use a ratiometric approach. Towards this end, we added a HA-tag to all proteins and quantified the relative protein amounts of TTG1, GL3 and GL1 in parallel to the LUMIER assays on Western blots using the HA-antibody (Figure 4.1B and Supplementary figure 4.6).

Based on these ratios we normalized the Renilla-tagged protein amounts with the GL3 levels. The ratio of GL1 or TTG1 to GL3 was varied by dilution series of the extracts. The amount of GL1 and TTG1 precipitated with ProtA-tagged GL3 was analysed by measuring the Renilla luminescence. To enable a comparison between different experiments, we normalized every measurement by the maximum intensity. Each experiment included two technical replicas and was repeated several times (biological replicas) with different ranges of the ratios between GL1 or TTG1 and GL3. The data of the biological replicas were combined to estimate the dissociation constants (K_D) for GL3-GL1 and GL3-TTG1. The K_D was calculated by fitting a model of reversible binding to the data with a non-linear least squares approach (Figure 4.1C). Because of the normalization methods that we used, these K_D are dimensionless and relative to the total GL3 concentration. To indicate this, we refer to them as relative K_D, denoted by $\bar{K}_D$. The best fits resulted in a $\bar{K}_D = 1$ for TTG1-GL3 and $\bar{K}_D = 0.5$ for GL1-GL3. The TTG1-GL3 dimer has a slightly higher $\bar{K}_D$ than GL1-GL3. However, the confidence interval for both $\bar{K}_D$ estimates are very close. When taking into account that the method to determine the relative protein amounts by western blot analysis also introduces small errors we consider the dissociation constants of the protein dimers to be in a similar range.

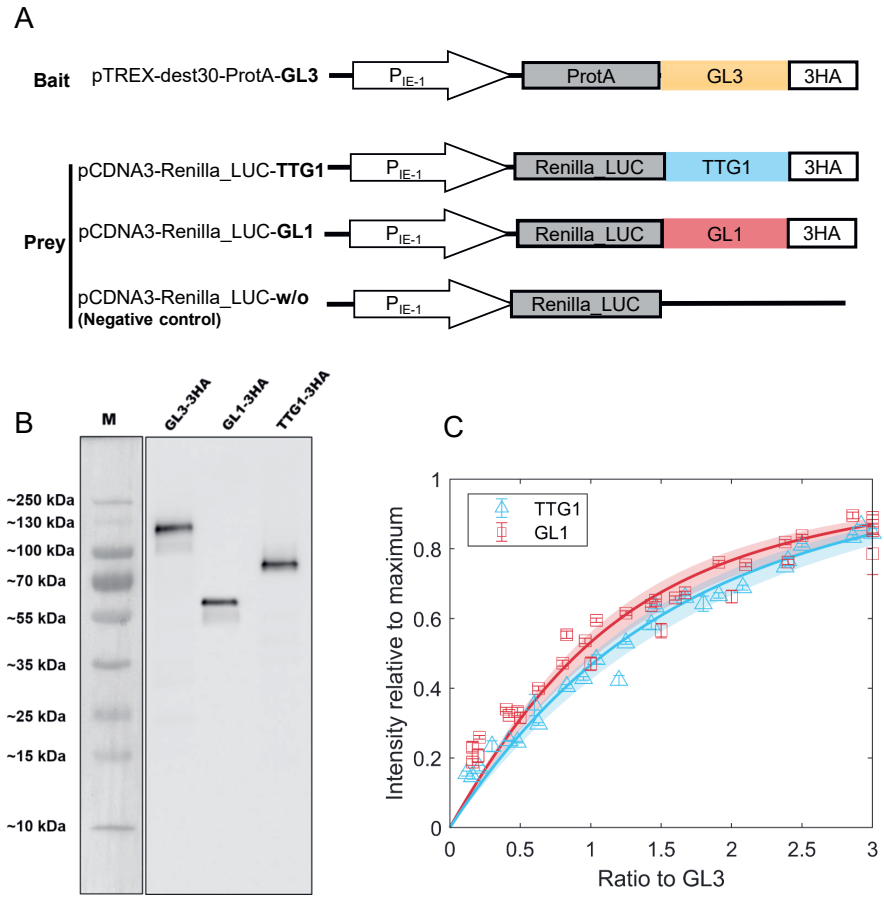


Figure 4.1: **LUMIER binding assays to estimate dissociation constants.** A: Schematic representation of constructs used in LUMIER. B: Western Blot used to determine the ratio of GL1 and TTG1 to GL1 using peak intensity. C: Model fit to LUMIER data of GL1 (red) and TTG1 (blue) binding to GL3. The shaded region indicates the 95% confidence interval on the dissociation constant estimate. Error bars indicate 3 technical replicates. Cloud of points stems from three biological replicates which are not averaged into single points due to variation in expression levels.

4.3.2 Negative cooperativity between TTG1 and GL1 for binding to GL3

To quantify to what extent the binding of TTG1 to GL3 has an effect on subsequent binding of GL1 to GL3 and vice versa, we performed two assays. In the

first approach, we added different amounts of TTG1-Renilla to GL3-ProtA in presence of GL1-YFP. As the Western blot analysis was done in parallel to the LUMIER assay, we could not calculate the GL1-YFP and GL3-ProtA ratios before starting the LUMIER analysis. The ratio of the GL1-YFP and GL3-ProtA varied between 1:1 to 1:2. To describe the mechanism underlying the competitive binding of GL1 and TTG1 to GL3 we extended the model by including the simultaneous binding of TTG1 and GL1 to a single GL3 unit (Figure 4.2A). We used the individual dissociation constants determined in the non-competitive experiments and introduced the cooperativity parameter α which describes the change in the $\bar{K}_D$ of GL1-GL3 and TTG1-GL3 upon additional binding of TTG1 or GL1, respectively. To estimate the cooperativity parameter, we fixed the $\bar{K}_D$ of TTG1-GL3 and GL1-GL3 to the values found from the non-competitive data. Fitting our data with the model revealed a cooperativity parameter of 0.2, indicating strong negative cooperativity (Figure 4.2B).

In the second approach, we quantified the GL1 GL3 interaction by adding different amounts of GL1-Renilla to GL3-ProtA in the presence of TTG1-YFP. In all three experiments we found a slightly S-shaped response curve (Figure 4.2C), suggesting a highly non-linear behaviour. Using the χ^2 score for the estimate of α we found that the lower limit of the 95% confidence interval is unidentifiable [178], as seen in the χ^2 profile for α (Figure 4.2C), which means that after a certain lower point in α, the fit of the model to the data is not improved upon further reduction of α. Given the poor fit of the model to the data, we asked the question what could result in the differing shape of the curve of the data and the model. One possible explanation is that GL1 shows weak homo-dimerization ([38, 179]; Supplementary Table 1). When including the GL1 dimerization in our model we found the S-shaped response curve (Figure 4.2D).

4.3.3 Formation of higher order complexes of GL3

The homodimerization of GL3 could lead to higher order complexes [43, 69, 136, 160]. In a first step we determine the $\bar{K}_D$ for the GL3 homodimer (Figure 4.4B). We find an estimate of $\bar{K}_D = 0.5$, indicating that the GL3-GL3 dissociation constant is very similar to TTG1-GL3 and GL1-GL3.

In a next step we assessed whether GL3 dimerization is changed by additional binding of GL1 or TTG1. Towards this end, we performed LUMIER assays using GL3-ProtA and GL3-Renilla at different concentrations in the presence of TTG1-YFP or GL1-YFP (Figure 4.4B). These experiments revealed

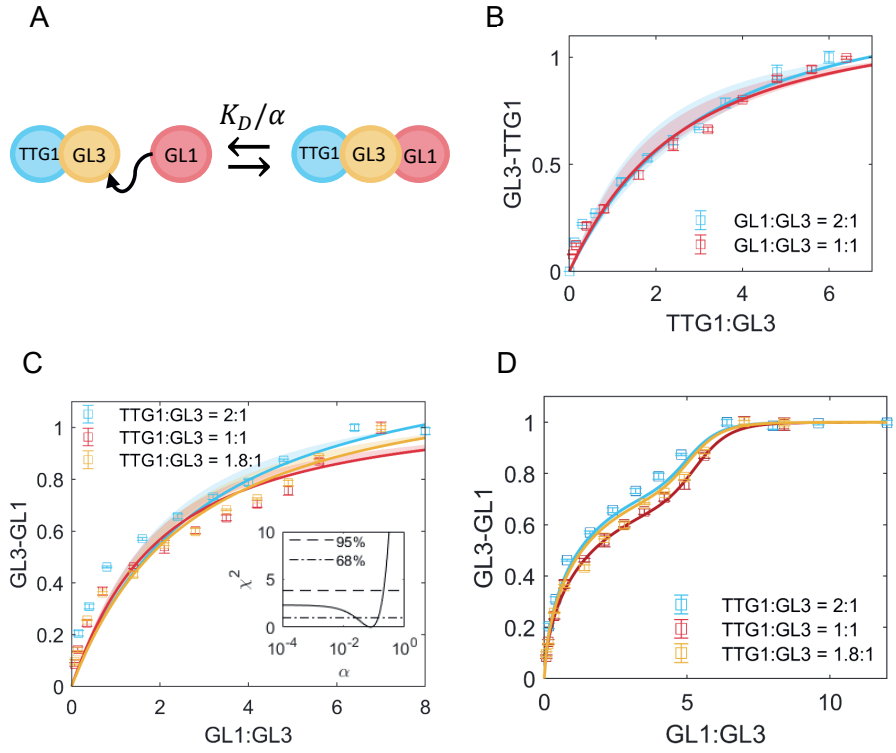


Figure 4.2: **Competition between TTG1 and GL1 for binding to GL3..** A: Schematic representation of cooperativity between GL1 and TTG1. The $\tilde{K}_D$ of GL1 is adjusted by the cooperativity parameter α as a result of TTG1 bound to GL3. B: Measurement of GL3-TTG1 binding in presence of GL1 at 2:1 (blue) and 1:1 (red) ratio. The model in A is used to fit the data and estimate $\alpha = 0.2$. Shaded regions indicate the 95% confidence interval. C: Measurement of GL3-GL1 binding in presence of TTG1 at 2:1 (blue), 1:1 (red) and 1.8:1 (yellow) ratio. The shaded region indicates 68% confidence interval as determined by the χ^2 profile of the estimate for α in the inset. The dashed line indicates the 95% confidence interval and the dash-dotted line indicates 68% confidence. D: Fit to the data in C with a function that allows higher order complexes in GL1.

that GL1 and TTG1 do not change the dimerization behaviour.

To enable the modelling of higher order complex formation mediated by

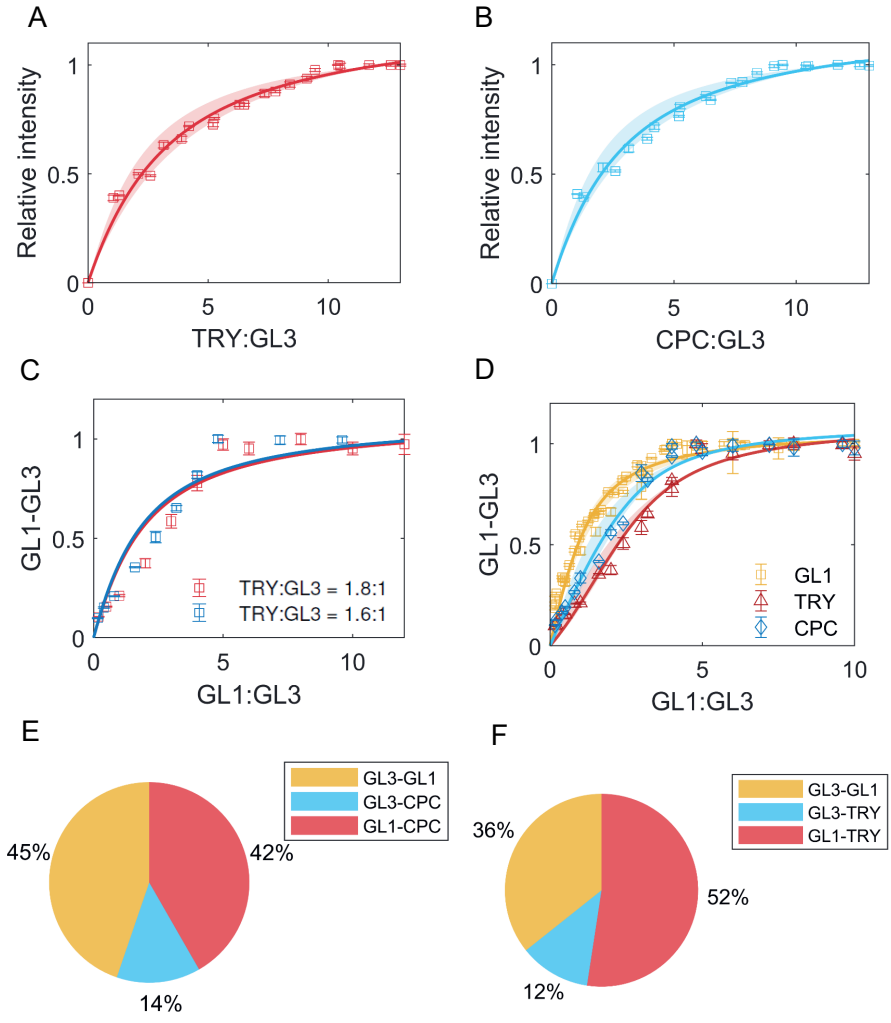


Figure 4.3: **Inhibitor binding.** A: LUMIER data and model fit for TRY-GL3 binding. The shaded region indicates the confidence interval. B: CPC-GL3 data and model fit. C: Competition of GL1 with TRY, for different ratios indicated in legend. D: Competition of GL1 with TRY (red) and CPC (blue), using a model that includes binding between the inhibitors and GL1. The yellow data indicates the GL1-GL3 binding in absence of TRY/CPC. E and F show the prediction of percentages of complexes formed using the estimates from the data for GL1 competition with TRY and CPC, respectively.

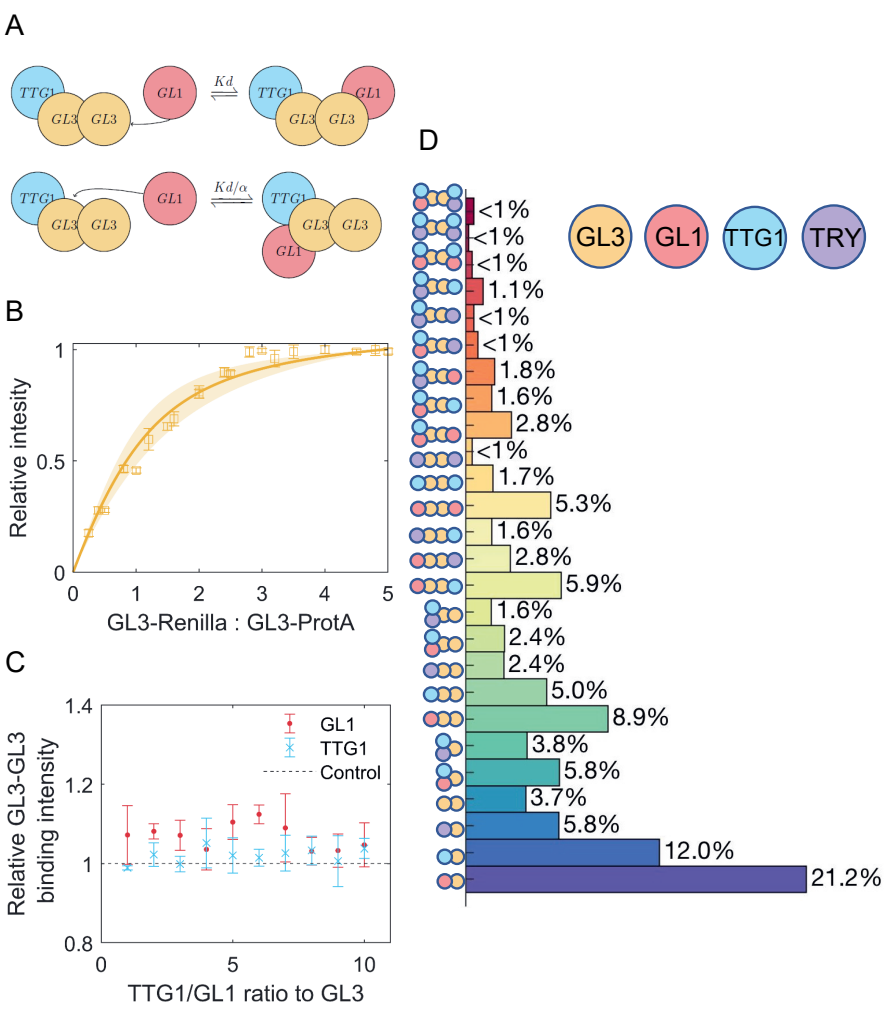


Figure 4.4: **Competition between TTG1 and GL1 for binding to GL3, including GL3 homodimerization.** A: Schematic representation of cooperativity between GL1 and TTG1. The $\bar{K}_D$ of GL1 is adjusted by the cooperativity parameter α as a result of TTG1 bound to GL3. B: Estimate of GL3-GL3 $\bar{K}_D$ using a least squares fit, the shaded region indicates the 95% confidence interval. C: Measurement of GL3-GL3 binding under different amounts of TTG1 and GL1. Amount of GL3-GL3 binding is given relative to the control where TTG1 or GL1 is absent. D: Prediction of the percentage of different possible complexes assuming equimolar amounts of GL1, TTG1, GL3 and TRY.

GL3 dimerization, we limited the analysis on the competitive complex formation data obtained in the first set of TTG1-Renilla experiments. Note that for simplicity, we did not consider the potential higher order formation possibly mediated by the weak GL1 dimerization. The extended model takes into consideration competitive complex formation for the same GL3 protein, treats GL3 dimerization to be independent of binding to GL1 and/or TTG1 and assumes that GL1 or TTG1 binding to one of the GL3 molecules of the GL3 homodimer have no effect on binding sites of the other GL3 molecule (Figure 4.4C). The latter is plausible as GL3 dimerization is not affected by GL1 or TTG1. Modelling of these events revealed a cooperativity parameter of 0.4. Thus, GL3 dimerization resulted in a cooperativity parameter twice as high as calculated without GL3 homodimerization. This suggests that GL3 homodimerization reduces the predicted negative effect of GL1 on TTG1 binding and vice versa.

To determine which of the models (competitive or cooperative with or without GL3 homodimerization) explains the data most accurately we determined the root mean square error (RMSE) for each fit. Note that each of these models consists of the same amount of parameters to estimate. Although there was only a small difference, we found the lowest RMSE for the cooperative model with GL3 homodimerization (Supplementary Table 2).

4.3.4 Formation of the inhibitor complex

The binding of GL1 to GL3 is thought to occur at the same binding site as the inhibitors TRY and CPC [139]. To quantify this competition, we first determined the $\bar{K}_D$ of the two inhibitors with GL3 using GL3-ProtA in combination with TRY-Renilla or CPC-Renilla. For TRY and GL3 we found the best fit for $\bar{K}_D = 2.7$ (Figure 4.3A), and for CPC-GL3 we found $\bar{K}_D = 2.3$ (Figure 4.3B). As the confidence intervals overlap, we consider the $\bar{K}_D$ to be in the same range. Note, however, that the $\bar{K}_D$ of the inhibitors are approximately 2-fold higher than TTG1, GL1 and GL3 suggesting that the interaction between the activators is stronger than the interaction of the inhibitors with GL3.

In a next step we quantified the effect of the inhibitors on the binding between GL3 and GL1 experimentally. In these experiments, the binding between GL3-ProtA and GL1-Renilla is measured in the presence of a fixed amount of TRY-YFP or CPC-YFP. To find the best model to describe the data, we initially used the $\bar{K}_D$ values for GL1-GL3 and TRY/CPC-GL3 obtained in the non-competitive experiments in a competitive, single binding site model.

This did not describe the data very well (Figure 4.3C). We therefore extended the model by testing whether the binding of the inhibitors to GL1 might improve the fit (Figure 4.3D) [38], thereby introducing one parameter to estimate when fixing the estimates of binding to GL3 from the non-competitive data. This led to a clear reduction of the RMSE (Supplementary Table 2), indicating that the binding of the inhibitors to GL1 could be a potential explanation for the difference between the binding behaviour predicted by the $\bar{K}_D$ obtained from pair wise interaction studies and three component assays.

4.3.5 Estimating the relative amounts of different MBW complexes

The complex interaction behaviour of the MBW proteins and the inhibitors raised the question about the relative ratios of the different multimers and how these change with different ratios of the protein amounts. To estimate this, we used the $\bar{K}_D$ values to calculate the ratios of the multimers. In a first step, we calculated the relative amounts of MBW complexes without the inhibitor. About 50% of the complexes would be expected to contain only one GL3 bound to GL1, TTG1 or both. Among the GL3 dimers containing complexes only a small fraction below 1% of all complexes are expected that contain six proteins (Supplementary Figures).

In a next step, we calculated the relative amounts of the multimers in the presence of TRY. When all proteins are present in equimolar amounts one would expect that 19% of the complexes with one GL3 protein contain TRY and are therefore expected to be inactive (Figure 3E). Among the GL3 dimer containing complexes about 23% of the complexes contain TRY suggesting that dimerization renders TRY inhibition slightly more effective.

Given that binding of TRY to GL3 is considered to represent the relevant biochemical mechanism of repression [34, 50, 51, 56], it was surprising that only about 20% of the complexes contain TRY when all are present in equimolar amounts. We use a range of a relative changes between 0.25 to 4 of individual proteins with respect to GL3. This revealed some notable observations. First, the formation of GL1 GL3 TTG1 trimers is fairly low (about 5%) for a wide range of combinations and increases to more than 20% only, if the relative amounts of GL1 and TTG1 are both four-fold higher than GL3. Second, the amount of TRY containing inhibited complexes is between 15 and 30% for a wide range of concentrations. It requires low GL1 concentrations in combination with four-fold higher TRY levels to predict more than 50% inhibited complexes.

The largest complex included in this model consists of six sub-units. As all of these proteins are involved in the patterning of trichomes, a logical follow-up question is what the effect is of forming such a high order complex on the patterning capabilities of the trichome system. To make a prediction on this we study a simple model, namely the activator-inhibitor model [17]. This model has been used as a basic framework for trichome patterning to study competitive complex formation [38], as well as other patterning-specific properties of the trichome network [39–42]. Using this activator-inhibitor model, we derive the necessary conditions for pattern formation and determine how these conditions are affected by changes in the order of the complex. From the mathematical analysis we find that the formation of higher order complexes provides the system with an advantage in terms of robustness, i.e. it becomes easier to find conditions under which a stable pattern can be formed.

4.3.6 A patterning model predicts increased robustness through multimer formation

The essential principles of trichome pattern formation have been formulated as mathematical models [38, 42, 180]. At the basis of these models are the activator-inhibitor and substrate depletion systems [17]. Here, we have analysed the activator-inhibitor model and adapted it to include the formation of higher-order complexes (> 2 sub-units) in the activator terms. We derived the conditions upon which the model would form patterns (see methods, Section 4.5.10) and determine how the size of the Turing space (in two dimensions) varies for complexes with $n = 2$, 4 or 6 sub-units. From this we find that a higher value for n, i.e. a higher order complex, leads to an increased size of the Turing space. This indicates that the number of possibilities of generating a pattern is increased. More specifically, patterning is more robust to changes in the values of a and b (production- and degradation-rate of the activator, respectively). Additionally, there is a shift in the patterning region towards higher values of a. This indicates that more activator is required to satisfy the conditions under high n.

4.4 Discussion

At the basis of trichome patterning are the interactions between the proteins that are classified as activators and inhibitors [7, 35]. Of all the proteins involved, GL3 plays a central role in that it is a binding partner to the other activators [44, 61, 62, 136, 139, 160], among which are TTG1 and GL1, as well

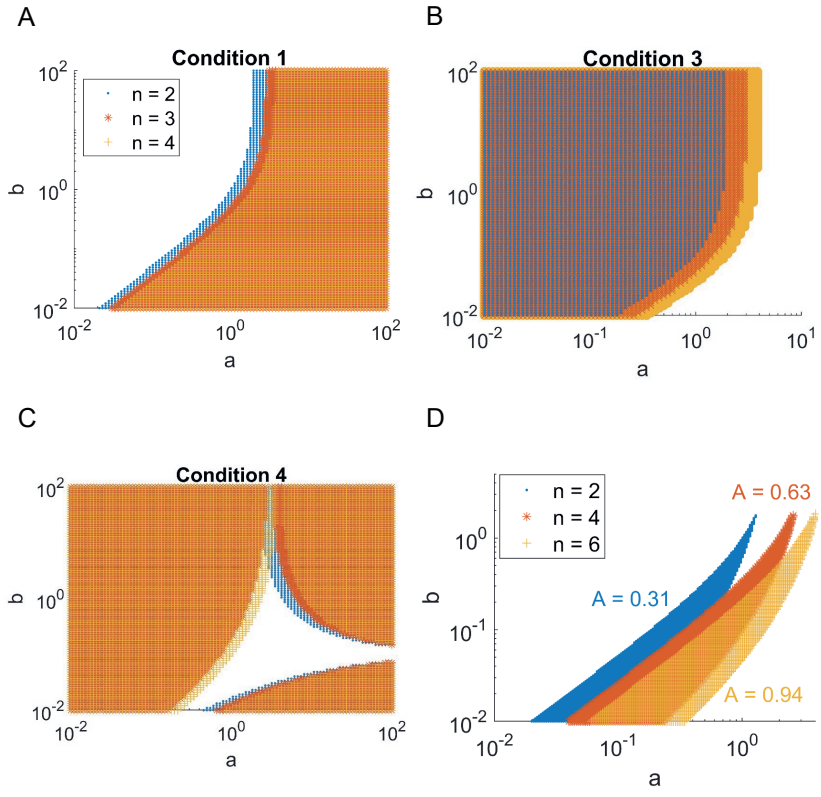


Figure 4.5: **Combinations of a and b in the activator-inhibitor model in (4.37) - (4.38) for which the patterning conditions are met.** A,B and C: Overview of the conditions derived in (4.46) - (4.49). Values of a and b for which condition (4.46), (4.48) and (4.49) are met are indicated by a marker. The different colours indicate the complex order n as shown in the legend of A. D: Regions for which all conditions are met, for complex orders $n = 2$, 4, and 6. The size of the area A of each of the regions is indicated.

as the inhibitors, among which are CPC and TRY [34, 50, 51, 56]. Here, we quantified these interactions and estimated relative dissociation constants. By testing different modes of binding and competition, we determined the most likely scenarios given the data and predicted in what amount certain complexes are present when assuming a certain ratio of individual proteins.

4.4.1 Negative cooperativity between activators suggests higher order complex compositions

It is commonly assumed that a trimeric complex of GL3, TTG1 and GL1 drives the activation of downstream promoters that are involved in trichome patterning [38, 43, 44, 61, 136]. From previously published results it was shown that there is a form of competition between GL1 and TTG1 in binding to GL3 [38, 44, 64], which led to the assumption that a differential dimer formation of GL3-GL1 and GL3-TTG1 is a more likely scenario. The results found in this paper are in line with that hypothesis, where we found a negative cooperativity parameter between GL1 and TTG1.

Additionally, we found that the individual dissociation constants for both GL1 and TTG1 lie in similar ranges, indicating that their relative amounts play an important role in determining the final composition of complexes and thus the efficiency with which certain downstream targets are activated [172–176]. Finally, there's an added layer of complexity in the form of homodimerization of GL3 [43, 69, 136, 160], which according to our binding models negates the effect of negative cooperativity between TTG1 and GL1, leading to higher-order complexes where both are found bound to either sub-unit of the homodimerized GL3. Taken together, the presented scenarios indicate a versatility in binding behaviours and a range of possible complexes in which the individual ratios of TTG1 and GL1 to GL3 play a crucial role in the final composition of complexes. This versatility could be translated into a fine-tuning mechanism where certain complex compositions lead to more or less efficient activation of targeted promoters.

Furthermore, from a mathematical analysis using an activator-inhibitor patterning model we predict that the formation of higher-order complexes leads to increased robustness in terms of forming stable patterns, suggesting an additional function for the complex formation scenarios presented here. The observations made for this simple activator-inhibitor system do not necessarily extend to the models formulated for the trichome system. This analysis can only indicate a tendency in the changes in patterning space as a result of higher order complex formation. Its strengths lie in its simplicity and thus the ease of interpretation. For any of the more extensive and intricate trichome models, such an analysis would become intractable. As the activator-inhibitor system underlies the trichome patterning models [7, 35, 38, 40, 42], it serves as an appropriate simplification for the purposes of this exploration, giving an idea of what could be expected in terms of patterning properties as

a result of higher-order complex formation as observed in the trichome patterning proteins. In the case of a trichome-patterning-specific model, taking into account all the possible complexes for the trichome proteins would lead to an overly-convoluted model from which it would be problematic to draw a similar prediction.

4.4.2 Inhibitors binding is weaker than activators and show a complexity beyond competitive binding

The estimated dissociation constants of TRY and CPC are 2-fold higher than found for GL1 and TTG1, indicating weaker binding to GL3. As the displacement of GL1 by TRY and CPC is expected to be the leading mechanism by which TRY and CPC exert their inhibitory function [38, 50, 56, 139], this would suggest that in order to achieve inhibition in non-trichome cells, the inhibitors would have to be present in higher amounts than the activators in order to compensate for the weaker binding strength. This is in line with the concentration profiles predicted by the patterning models based on activator-inhibitor principles assumed for trichome patterning, where in the trichome-peak the amount of activators is much higher than inhibitors and vice versa in epidermal cells [38–41]. While this difference in binding strength could play a role in achieving this difference, it is not the only mechanism in the patterning models. More specifically, these models include feedback-loops to achieve this distribution of high activator in trichome cells and high inhibitor in non-trichome cells [38–41]. One additional assumption the patterning models often rely on is a higher mobility for inhibitors than the activators [17, 123], which a lower binding constant could also lead to as less proteins are caught in a cell-autonomous complex [59].

Finally, we observed that the binding curves for TRY and CPC were not accurately reproduced by a simple, single-binding site competitive binding model. From our results, we found that the binding of TRY and CPC to GL1 would resolve this inaccuracy and so the model predicts that this could be a potential mechanism which is as of yet unexplored.

4.5 Methods

4.5.1 LUMIER (LUminescence-based Mammalian IntERactome)

Staphylococcus aureus protein A or *Renilla reniformis* luciferase (Rluc) was fused to the N-terminus of each protein while the third protein was fused to YFP

at the N-terminus using the backbone of pTREXdest30 and three constructs were transiently expressed in HEK293TM cells (BioCat/SBI: LV900A-1). Transfection and cell harvesting were done as described before [60, 64]. After 48 hours cells were washed three times with PBS, lysed in $750\mu l$-$1000\mu l$ lysis buffer. Extracts were normalized with respect to the YFP signal and Rluc signal (TECAN) then combined after 1 hour lysis. The total volume was kept constant by adding untransfected cell lysate. Each combination was prepared in triplicate. Protein-immunoprecipitation and luminescence measurements were done as described previously [60] using untransfected cells or cells expressing Luciferase-protein as controls. Cells solely expressing YFP-protein was also performed to exclude any nonspecific interference signal. The percentage of Rluc on the beads compared with the lysate was calculated by dividing the Rluc activity on the beads by the Rluc activity in the same amount of lysate used in the pull-down assay (input).

4.5.2 Western Blot

Western blot experiments were performed as described in Molecular Cloning [181]. Materials were used as follows: PVDF membrane (Roth), Super Signal West Femto Maximum Sensitivity Substrate (Termo Scientific), Mini Trans-Blot Cells (BioRad) for wet western blotting, Mini Protean Cells (BioRad) for SDS gel electrophoresis, and Prestained Protein Ladder (Fermentas). Protein lysate was extracted from HEK cell and detected with Anti-HA-Peroxidase (5 mU/ml 1:2500 roth). Each lane is $40x$ dilution of original lysate by lysis buffer. Relative density of each band is analysed by ImageJ (1.48v, National Institutes of Health, USA).

4.5.3 Single site, reversible binding model

In the single binding experiments the binding between the protA-tagged protein and the Renilla-tagged protein is measured. For fitting the data we use a simple, single-site binding model under the assumptions of mass balance and equilibrium. Let x stand for protA-tagged protein, y for Renilla-tagged protein, α for the association rate and β for the dissociation rate, we get

$$x + y \underset{\mu}{\overset{\gamma}{\rightleftharpoons}} c \tag{4.1}$$

$$\dot{x} = -\gamma xy + \mu c \tag{4.2}$$

$$\dot{y} = -\gamma xy + \mu c \tag{4.3}$$

$$\dot{c} = \gamma xy - \mu c \tag{4.4}$$

$$x_0 = x + c \tag{4.5}$$
$$y_0 = y + c \tag{4.6}$$

where x_0, y_0 is total amount of protein (the sum of free and bound protein). Upon substitution of (4.5) and (4.6) into $\dot{x} = 0$ we get the following expression for c at steady state

$$c = \frac{1}{2x_0}\left[\left(1 + \frac{y_0}{x_0} + \frac{\mu}{\gamma x_0}\right) \pm \sqrt{\left(1 + \frac{y_0}{x_0} + \frac{\mu}{\gamma x_0}\right)^2 - 4\frac{y_0}{x_0}}\right]. \tag{4.7}$$

Note that in (4.7) the term $\frac{y_0}{x_0}$ indicates the ratio between the Renilla-tagged protein concentration y_0 and the protA-tagged protein concentration x_0, and that the dissociation constant is given by $K_D = \frac{\mu}{\gamma}$. So, given the normalization by the total amount of protA-tagged protein x_0, we get a normalized K_D, namely $\bar{K}_D = (\frac{\mu}{\gamma x_0})$. Finally, to allow comparison between different experiments, we normalize c by c_{max} which is the amount of complex at the point of saturation.

4.5.4 Modelling higher order complexes in GL1

Given that the single site, reversible binding model showed a poor fit to the data in Figure 4.2C, we have used a model that allows homodimerization in GL1 such that higher order complexes than the trimer in Figure 4.2A can be formed. Towards this end, we introduce a hill-function [117] that describes binding in the following form:

$$s = \frac{ax + bx^n + cxz + dx^m z}{1 + ax + bx^n + cxz + dx^m z} \tag{4.8}$$

where n and m are the hill coefficients that indicate the order of protein x (in this case GL1), as a homodimer or as part of the complex with GL3 and TTG1, respectively; a, b, c and d are the coefficients for each binding term and z is the competitor protein (in this case TTG1). For Figure 4.2D the best fit is found for $n = 1$ and $m = 10$, indicating that a high form of non-linearity is part of the competitive complex formation which is more precisely defined in the following models.

4.5.5 Competitive displacement model

A protein binding experiment with competition for a single binding site is described by the following equations:

$$x + y \overset{K_{Dy}}{\rightleftharpoons} xy, \tag{4.9}$$

$$x + z \overset{K_{Dz}}{\rightleftharpoons} xz, \tag{4.10}$$

where conservation of mass requires that

$$x_0 = x + xy + xz, \tag{4.11}$$

$$y_0 = y + xy, \tag{4.12}$$

$$z_0 = z + xz. \tag{4.13}$$

Assuming that the x protein is tagged with protA again and y with Renilla, we get the following expression used to fit to the data [182]

$$xy = \frac{\bar{y}_0\left[2\sqrt{a^2 - 3b}\cos(\theta/3) - a\right]}{3\bar{K}_{Dy} + \left[2\sqrt{a^2 - 3b}\cos(\theta/3) - a\right]} \tag{4.14}$$

where the bar notation indicates the normalization by x_0 and

$$a = \bar{K}_{Dy} + \bar{K}_{Dz} + \bar{y}_0\bar{z}_0 - 1, \tag{4.15}$$

$$b = \bar{K}_{Dz}(\bar{y}_0 - 1) + \bar{K}_{Dy}(\bar{z}_0 - 1) + \bar{K}_{Dy}\bar{K}_{Dz}, \tag{4.16}$$

$$c = -\bar{K}_{Dy}\bar{K}_{Dz}, \tag{4.17}$$

$$\theta = \arccos\frac{-2a^3 + 9ab - 27c}{2\sqrt{(a^2 - 3b)^3}}. \tag{4.18}$$

K_{Dy}, K_{Dz} are the same as the dissociation constants determined with the single binding model in (4.7). In the LUMIER data and the model (4.14) is normalized by the amount of xy measured at saturation levels.

4.5.6 Extension of the competitive displacement model

As the model in Eq. (4.14) does not seem to reflect the experimental data very well, a different mechanism is explored. In this extension the binding between

the inhibitor and GL1 is included. The set of reactions are:

$$x + y \overset{K_{D1}}{\rightleftharpoons} xy, \tag{4.19}$$

$$x + z \overset{K_{D2}}{\rightleftharpoons} xz, \tag{4.20}$$

$$y + z \overset{K_{D3}}{\rightleftharpoons} yz, \tag{4.21}$$

$$x_0 = x + xy + xz, \tag{4.22}$$

$$y_0 = y + xy + yz, \tag{4.23}$$

$$z_0 = z + xz + yz. \tag{4.24}$$

In this case it is not straightforward to derive an expression for the amount of complex like the one given in (4.7) and (4.14), therefore the equations are solved numerically to find the protein concentrations at steady state. Again the protein amounts and K_D are normalized by x_0, i.e. the total amount of protA-tagged protein (GL3).

4.5.7 Cooperative binding model

For the competition experiment with GL3, TTG1 and GL1 we first use the model in (4.14) to try to fit the K_D in the presence of the third protein. Because this does not give a good fit, the model is extended to include higher order complexes with the possibility of a GL1-GL3-TTG1 complex. Given that the K_D for GL1 and TTG1 for binding to GL3 is very similar, we introduce a parameter α as a cooperativity parameter [cite] that indicates the change in the K_D when GL1 or TTG1 is already bound to GL3, giving the following model:

$$x + y \overset{K_{D1}}{\rightleftharpoons} xy, \tag{4.25}$$

$$x + z \overset{K_{D2}}{\rightleftharpoons} xz, \tag{4.26}$$

$$xz + y \overset{K_{D1}/\alpha}{\rightleftharpoons} xyz, \tag{4.27}$$

$$xy + z \overset{K_{D2}/\alpha}{\rightleftharpoons} xyz, \tag{4.28}$$

$$x_0 = x + xy + xz + xyz, \tag{4.29}$$

$$y_0 = y + xy + xyz, \tag{4.30}$$

$$z_0 = z + xz + xyz. \tag{4.31}$$

Note that the binding of y to x is always indicated by K_{D1} as seen in (4.25) and (4.27), but in the case when z is already bound to x, the K_{D1} is adjusted by α as shown in (4.27). Here the equations are solved numerically and α is estimated by a least squares fit. The interpretation of α can describe three different scenarios depending on its value: *1)* $\alpha < 1$, the K_D increases, i.e. negative cooperativity, *2)* $\alpha = 1$ the K_D is unchanged, i.e. no cooperativity but independent binding, *3)* $\alpha > 1$, the K_D decreases, i.e. positive cooperativity. In this case the signal is modelled as $xy + xyz$ and is normalized by the signal at saturation.

4.5.8 Extension of the cooperative binding model

To explore whether the homodimerization of GL3 plays an important role in the competition experiments with GL3, TTG1 and GL1 we extend the model from (4.25) - (4.28) to include GL3 homodimerization. As a result of this extension the model now consists of 46 reversible reactions, where the highest order complex is a hexamer. We describe these reactions with a set of ordinary differential equations (ODEs) which, in chemical reaction network theory is commonly written as [117]

$$\dot{c} = S\vec{v}(c,k) \tag{4.32}$$

where S is the stoichiometric matrix and $\vec{v}(c,k)$ the vector of reaction rates, which are of the form of mass action kinetics:

$$v_j = k_j \prod_{i=1}^{N} c_i^{\beta_{ij}}, \quad j = 1,\ldots,R \tag{4.33}$$

where N is the number of species, R is the number of reactions and β_{ij} is the molecularity of the reactant species i in reaction j. In this model we have $N = 22$ species and $R = 92$ reactions. Since we are dealing with a closed system at equilibrium we solve for $\dot{c} = Sv(c,k) = 0$ to find the solutions of species concentrations at steady state. Cooperativity is only included when GL1 and TTG1 bind to the same sub-unit of GL3, for all other binding events we assume independent binding and use the K_D estimates from the non-competitive data, including the GL3-GL3 binding rate. In the least squares fit to the data we have one estimable parameter, namely α the cooperativity parameter.

4.5.9 Parameter estimation and identifiability

In order to get an estimate for the K_D the model output is fitted to the data via a least-squares approach [178]. The agreement between data and model is

described by the weighted sum of squared residuals

$$\chi^2(\theta) = \sum_{i=1}^{n} \frac{(y_i - f(x_i, \theta))^2}{\sigma_i} \tag{4.34}$$

where y_i indicates the i^{th} data point, $f(x_i, \theta)$ the point as predicted by the model with parameters θ and σ_i is the corresponding measurement error. The most optimal parameters $\hat{\theta}$ can then be estimated numerically by

$$\hat{\theta} = \arg\min\left[\chi^2(\theta)\right]. \tag{4.35}$$

Confidence intervals for $\hat{\theta}$ can be derived by assuming a threshold in $\chi^2(\theta)$, defined by the region [178]

$$\{\theta | \chi^2(\theta) - \chi^2(\hat{\theta}) < \Delta_\alpha\} \quad \text{with} \quad \Delta_\alpha = \chi^2(\alpha, df) \tag{4.36}$$

the threshold Δ_α is the α quantile of the χ^2-distribution and with df the degrees of freedom, in this case $df = 1$, represents the confidence interval with confidence level α. This leads to a confidence interval for parameter θ_i with lower bound σ_i^- and upper bound σ_i^+. θ_i is identifiable if the interval $[\sigma_i^-, \sigma_i^+]$ of its estimate $\hat{\theta}_i$ is finite [178].

4.5.10 Multimers in the context of pattern formation: A mathematical analysis

We analyse the activator-inhibitor model as the typical reaction-diffusion scheme [3, 17, 123]

$$\partial_t u = f(u, v) + \nabla^2 = a - bu + \frac{u^n}{v} + \nabla^2 u \tag{4.37}$$

$$\partial_t v = g(u, v) + d\nabla^2 = u^n - v + d\nabla^2 v \tag{4.38}$$

where a, b, n and d are constants. This is a dimensionless version of the activator (u) - inhibitor (v) system. In the classical version of this system $n = 2$, but here we vary n to simulate higher-order complexes in the activator u. To determine what the effect is of increasingly higher order complexes on the pattern formation capabilities of the model, we derive the necessary conditions imposed on the model parameters by linear stability analysis [3, 123]. Instead of using any of the existing trichome models [35, 38, 40, 183, 184], we use

this simpler version as this allows an analytic approach in determining the conditions for the generation of spatial patterns. These conditions are[123]

$$f_u + g_v < 0, \quad f_u g_v - f_v g_u > 0$$
$$d f_u + g_v > 0, \quad (d f_u + g_v)^2 - 4d(f_u g_v - f_v g_u) > 0 \tag{4.39}$$
$$\tag{4.40}$$

where f_u, f_v, g_u and g_v are the partial derivatives of the reaction kinetics $f(u,v)$ and $g(u,v)$, evaluated at the uniform steady state. In the case of the activator-inhibitor system, this uniform steady state is

$$u_0 = \frac{a}{b}, \quad v_0 = \left(\frac{a}{b}\right)^n. \tag{4.41}$$

And at steady state the partial derivatives are

$$f_u = \frac{b(n-a)}{a} \tag{4.42}$$
$$f_v = \frac{-1}{\left(\frac{a}{b}\right)^n} \tag{4.43}$$
$$g_u = n\left(\frac{a}{b}\right)^{n-1} \tag{4.44}$$
$$g_v = -1 \tag{4.45}$$

Taken together, this leads to the following conditions:

$$f_u + g_v < 0 \quad \Rightarrow \quad \frac{b(n-a)}{a} < 1, \tag{4.46}$$
$$f_u g_v - f_v g_u > 0 \quad \Rightarrow \quad b > 0, \tag{4.47}$$
$$d f_u + g_v > 0 \quad \Rightarrow \quad \frac{d\,b(n-a)}{a} > 1 \tag{4.48}$$
$$(d f_u + g_v)^2 - 4d(f_u g_v - f_v g_u) > 0$$
$$\Rightarrow \quad \left(\frac{d\,b(n-a)}{a} - 1\right)^2 > 4db \tag{4.49}$$

Upon fixing $d = 10$ and varying a and b across a wide range of values, the shape of the Turing space (the region in parameter-space where spatial patterns are generated) can be visualized as in Figure 4.5D, which is obtained by combining all the sub-conditions (Figures 4.5A to 4.5C) into one region. Note that the second condition, $b > 0$, will always be satisfied (a negative degradation rate b of the activator is not biologically relevant).

4.6 Supplementary information

4.6.1 Supplementary figure 1

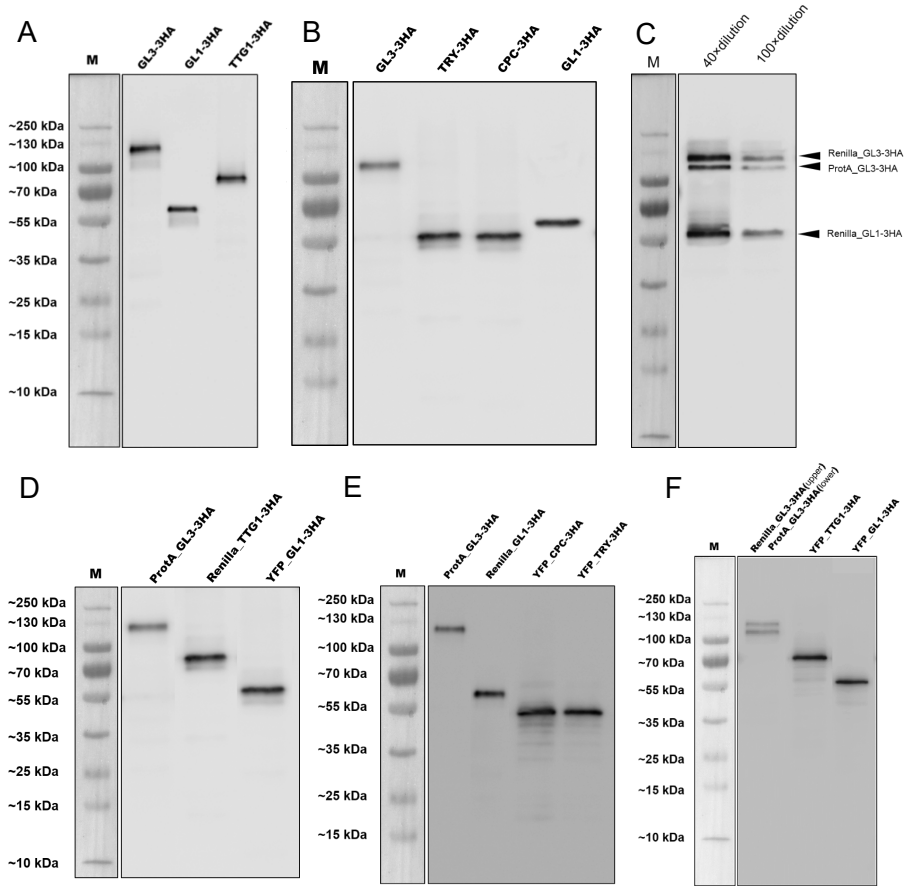


Figure 4.6: **Western blot analysis of proteins expressed in HEK cells.**

4.6.2 Supplementary table 1

Table 4.1: LUMIER experiment of GL1 homodimerization

ProtA_GL1	Renilla_GL1	Input (mean)	Pull down	Interaction
50 uL	50 uL	1396	268	-
50 uL	100 uL	2107	403	slightly
50 uL	200 uL	4022	695	+

4.6.3 Supplementary table 2

Table 4.2: Root mean square error (RMSE) of different models and datasets.

Model	Dataset	RMSE
Cooperativity, trimer	TTG1-GL3 measured, GL1 fixed	0.0588
Cooperativity, hexamer	TTG1-GL3 measured, GL1 fixed	0.0476
Competition with inhibitor	GL1-GL3 measured, TRY fixed	0.1025
Competition with inhibitor	GL1-GL3 measured, CPC fixed	0.0761
Competition with inhibitor, GL1-Inhibitor binding	GL1-GL3 measured, TRY fixed	0.0584
Competition with inhibitor, GL1-Inhibitor binding	GL1-GL3 measured, CPC fixed	0.0545

Chapter 5

Unravelling the genotype-phenotype relationship in a trichome double mutant through mathematical modelling

Anna Deneer, Jessica Pietsch, Christian Fleck and Martin Hülskamp

5.1 Abstract

The hairs on top of the leaves of *Arabidopsis Thaliana* form a pattern that is self-regulated. This system presents a unique opportunity in that it has proven to be easily genetically manipulable, leading to a diverse set of genotypes and corresponding phenotypes. Genetic analysis has revealed a complex and interconnected network which is impossible to explain intuitively, therefore, mathematical models are essential in elucidating this complexity. Both this level of complexity as well as the principles underlying the pattern formation are characteristic for many biological systems, making this an excellent model system. In this study, we focus on three unexplained genotype-phenotype connections and use a modelling approach to analyse the underlying mechanisms and how they are linked. In particular, we study knockouts of the inhibitors TRIPTYCHON and CAPRICE and show how the double knockout is not simply a combination of the single mutants as previously assumed, but rather the result of unexpected mechanisms that do not directly involve the inhibitors themselves.

5.2 Introduction

The pattern of hair cells of the leaf epidermis in *Arabidopsis* is regulated by an underlying mechanism that is capable of forming the pattern *de novo* [8, 35, 81], i.e. without any pre-existing information like cell lineage or signals

from underlying cell layers [10]. This mechanism robustly produces a pattern of hairs on leaves that maintains a consistent spacing between any pair of hair-cells, interspersed on average by the same amount of non-hair cells, with a certain degree of irregularity [32, 33]. The type of patterns formed by these so-called trichomes are reminiscent of the well-known Turing spot-patterns [3, 4]. In fact, the involved patterning genes have been found to fall within the typical roles of 'activator', 'inhibitor' and 'substrate' as represented in extensions of Turing's two component model, called the activator-inhibitor and substrate-depletion model [5, 17]. The number of genes involved in trichome patterning is extensive [34, 37] and a model capturing the entire complexity would be unhelpfully intractable. Therefore, all existing models of trichome patterning contain elements of activator-inhibitor and substrate-depletion principles [38–41, 183], composing a minimal set of components (Figure 5.1) such that certain genotypes and resulting phenotypes can be analysed through mathematical modelling.

What has been leading the development of these models is the study of mutant backgrounds in *Arabidopsis*, where the model is challenged to reproduce phenotypical, molecular and biochemical observations. In the case of trichomes, there is a wide variety of mutants available [34, 50, 136, 139], many with observable patterning defects that reveal specific aspects of the underlying machinery, thus giving insight into principles of development and how a network based on Turing's model might operate in a natural system.

Among the numerous genes found to be involved in trichome patterning, a core set has been identified (Figure 5.1), consisting of the positive regulators: TRANSPARANT TESTA GLABRA1 (TTG1) encoding for a WD40-protein [47–49], the partially redundant genes GLABRA1 (GL1) and MYB23 encoding for R2R3 MYB-type transcription factors [45] and GLABRA3 (GL3) encoding for a bHLH protein [34, 44, 46], redundantly with ENHANCER OF GLABRA3 (EGL3) [134–136]. These positive regulators fall under the category 'activators' in the context of the aforementioned patterning models [38], where TTG1 shows features that fall under 'substrate' [40, 141]. These activators interact with each other to form active complexes (ACs) that regulate other parts of the network [64]. The negative regulators, or 'inhibitors' in the models, consist of R3 MYB transcription factors, encoded by the genes TRIPTYCHON (TRY), CAPRICE (CPC) [34, 50–52], ENHANCER OF TRY AND CPC 1, 2 and 3 (ETC1, ETC2, ETC3) and TRICHOMELESS1 and 2 (TCL1, TCL2) [50, 53–59]. All of these genes together are considered to be the core network of trichome patterning and previously published models all involve

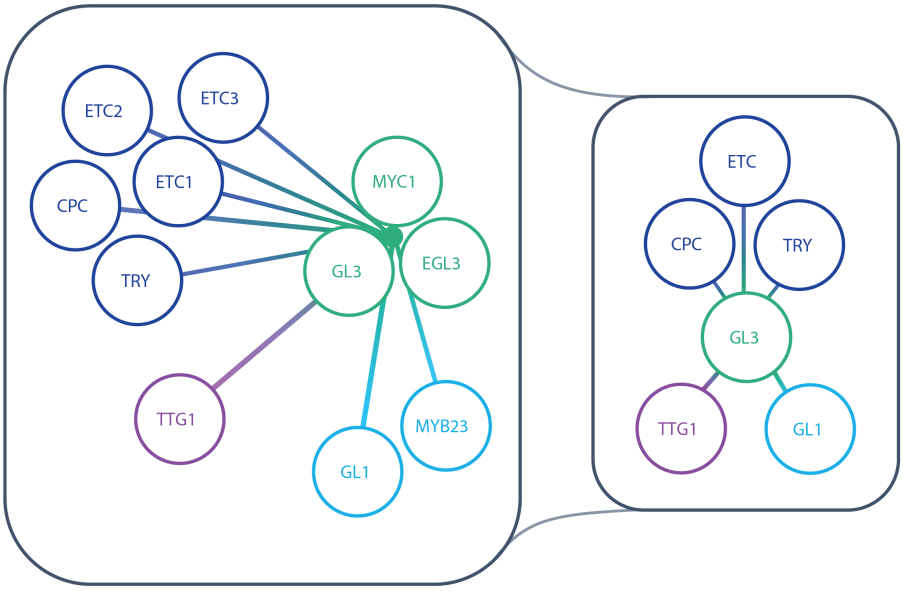


Figure 5.1: **Schematic representation of the trichome model.** Left: Network of the core components in trichome patterning and their interactions. Right: Network reduced to components involved in inhibitor mutant phenotypes.

different subsets of these genes [38–41, 183].

In this study, we focussed on mutations involving two inhibitors of trichome patterning: TRY and CPC [34, 50]. The phenotypes of the single knockout mutants of TRY and CPC each have different characteristics, where the *try* mutant shows a lower overall density and formation of small clusters, the *cpc* phenotype has a strongly increased density [50] (Figure 5.2). The double mutant *trycpc* at first glance appears a combination of those effects: a multitude of large clusters [50]. These observations suggest a difference in inhibitor behaviour, namely that TRY is hypothesised to inhibit trichome formation on a short spatial scale, whereas CPC is assumed to operate on long distance ranges [50, 59]. Yet, these simple assumptions alone are not enough to reproduce the observations of all three mutant phenotypes using any of the existing models. The phenotypes seen for the single mutants are not simply additive as assumed, so instead of some linear combination of effects, the underlying system shows a complexity that is not immediately and intuitively obvious. To

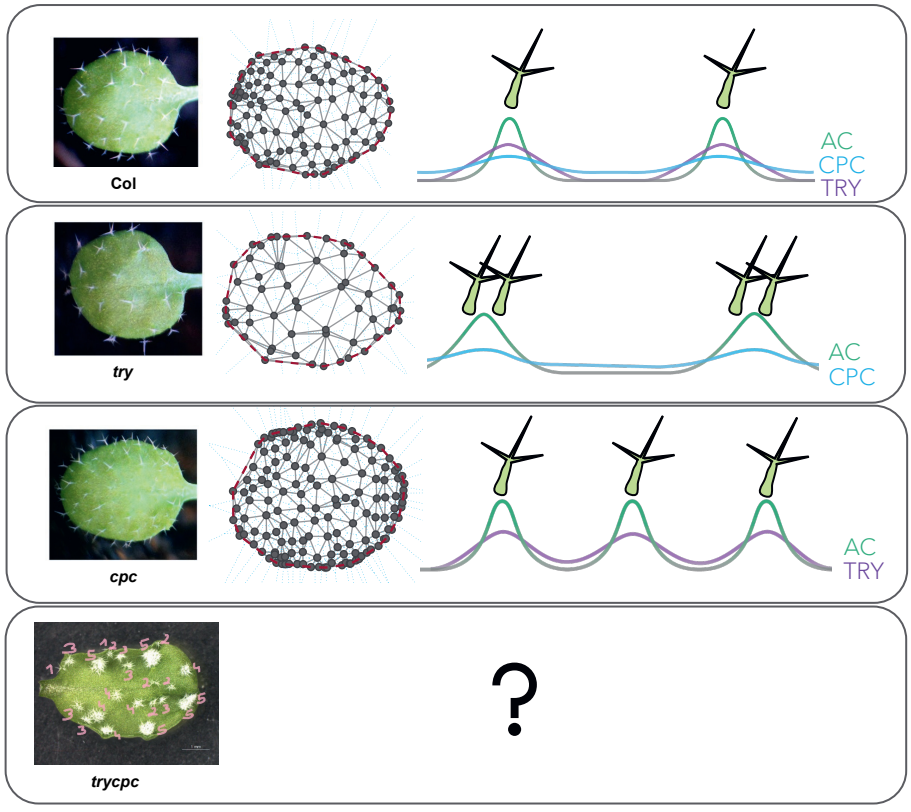


Figure 5.2: **Wild-type (Col), inhibitor mutant phenotypes and postulated mechanisms.** Left: Phenotype. Middle: Trichome coordinates and their neighbouring connections determined through Delaunay Triangulation. Right: Postulated protein profiles of active complex (AC), TRIPTYCHON (TRY) and CAPRICE (CPC), under the assumption that TRY acts as a local inhibitor and CPC as a long-range inhibitor. This leads to broad peaks in CPC (blue profile) and sharper, more localized peaks for TRY (purple profile). As a result, in the mutants *try* and *cpc*, the absence of TRY leads to broader peaks of AC and thus clusters at a larger distance; and the absence of CPC leads to sharp but more frequent peaks of AC, and thus a higher density of trichomes.

gain a mechanistic understanding of these phenotypes, a mathematical model developed and supported by experimental evidence is essential.

To determine what elements are still missing from the explanation behind the inhibitor mutant phenotypes and to learn more about what causes these kind of phenotypes, we extend upon a previously published model [183]. We show that it is not only the difference in mobility rates of inhibitors, but rather a combination of effects that involve all other components in the patterning network. Specifically, the model predicts three requirements. First, the role of cell differentiation and its temporal effect is crucial in cluster formation seen in the double mutant phenotype. Second, unlike assumptions underlying previous models, we assume mobility of activators, shown to be necessary to reproduce mutant patterns in the mathematical model and experimentally shown to be occurring for the positive regulators GL1 and TTG1. Finally, there is a dependency on the redundancy of the inhibitors through the robustness and versatility it infers on the system.

5.3 Results

5.3.1 Reproducing inhibitor mutants requires a temporal perturbation, activator mobility and inhibitor redundancy

In developing the model, we started from a previously published version that was established with the aim of reproducing multiple aspects of a mutant phenotype in one of its components, namely TTG1 [183]. As this model already contained the inhibitors TRY and CPC, we initially tested whether it could readily reproduce the *try*, *cpc* and *trycpc* phenotypes. Towards this end, we quantified aspects of the patterns and focussed on reproducing two of those quantities: trichome densities in the mutants relative to wild-type and the observed cluster densities [50]. After a parameter search covering over two million random parameter sets, we could not find a single set that was able to reproduce all three mutant phenotypes. While this does not constitute an exhaustive search of the 25-dimensional parameter space (which would be computational infeasible), it is indicative of a lack of robustness in meeting the particular criteria the model was subjected to, making it an implausible representation of the biological mechanism underlying these mutants. Some of these parameter sets could match certain individual phenotypes (e.g. the increased trichome density in *cpc*), however, we noticed a consistent failure in matching the double mutant, specifically the strong clustering behaviour in combination with a density similar to wild-type.

This prompted us to reconsider the model and investigate which elements were currently lacking. The initial results suggested that a mere lack of TRY

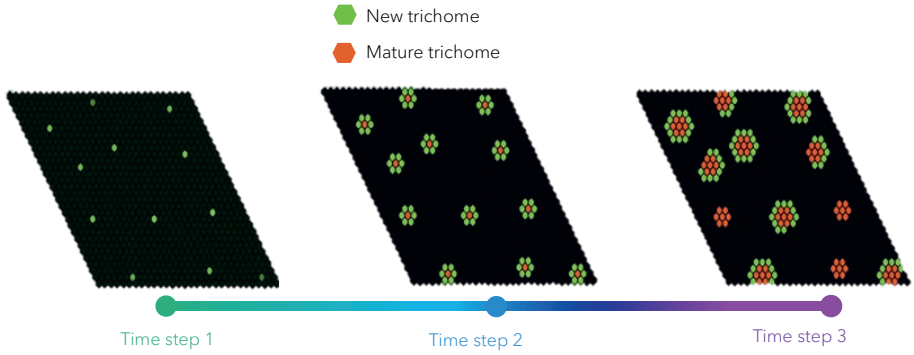


Figure 5.3: **Quasi steady state simulation of *trycpc*.** From left to right, every grid shows an intermediate steady state, where the green hexagons are high in active complex (AC) at that moment in time and thus differentiate towards trichome cell-fate, meaning that protein production is down-regulated in those selected cells. In the next time step, the orange hexagons indicate cells that have differentiated in the preceding steady state, which was used as initial conditions to reach the next quasi steady state, leading to a new set of differentiating cells and to clusters of increasing size at each quasi steady state.

and CPC in the system was insufficient to produce the nests of trichomes seen in the double mutant, we therefore aimed to define the mechanism behind the emergence of these large clusters more precisely and considered the importance of temporal differences between the trichomes within individual clusters. As reported in earlier findings [50], the clusters in *trycpc* show a certain developmental procedure where the initial location is determined by a single trichome, followed by a ring-like expansion of younger trichomes around the initiation site. This clear distinction in the timing of emergence led us to introduce such a temporal mechanism into the model (Figure 5.3).

In our simulations, we assume a time-scale separation between pattern formation and cell differentiation, given that processes underlying pattern formation like protein binding occur on much faster time scales than e.g., cell division. In this framework, we decided to introduce a mechanism that simulates the effects of cell differentiation through quasi steady state assumptions. This means that there are multiple quasi steady states within a single simulation (Figure 5.3), where the model is repeatedly run until steady state is reached, followed by cell differentiation of trichome cells. Cell differenti-

ation may lead to a perturbation in the network in one of two ways: one is that due to endoreduplication [185] the production of both activators and inhibitors is increased; another is that upon differentiation the patterning genes are downregulated [186] as they are no longer required. According to the model, the increased production would never lead to cluster formation in the ring-like structure as seen in the double mutant, as this only sharpens the peaks formed in the initial time steps. Only when we simulate the downregulation by strongly reducing or setting the relevant parameters to 0, do we find the ring-like structure as seen in the double mutant. As there is no experimental evidence available to verify which of the two scenarios are operating in the trichome system *in vivo*, we rely on the simulation results and assume downregulation upon differentiation in all of the performed simulations. More generally, a mechanism involving histone modification upon cell differentiation has been shown to reduce gene expression in plants [186, 187], which would be in agreement with the perturbation as we simulate it, yet this is a mechanism which is not specific to trichome cells in the way that it is treated in the model.

On top of this timing effect, we found that mobility of activators was another requirement for cluster formation of the type seen in the double mutant. The round shape of the clusters seen in *trycpc* mutants is indicative of the role of diffusion, which prompted us to consider this aspect in the model. Note that previous iterations of trichome patterning models were able to form clusters without this assumption of mobility, yet these were never of the size and shape as seen in the *trycpc* phenotype. Mobility of TTG1 has been shown previously and was also included in previous models, but here we additionally introduce the mobility of GL1 and GL3, limited to the free proteins and not to the complexes that are formed between the activators, assuming that these complexes would fall outside the size exclusion limit of the plasmodesmata [188] and are mostly present in the cell nucleus [60].

Finally, the model is necessarily extended with the inclusion of a third inhibitor. This third inhibitor represents the redundancy in the inhibitors ETC1, ETC2 and ETC3 (indicated only by ETC for the sake of brevity) [54, 56, 57, 59]. Note that this does not include the inhibitors TCL1 and TCL2 as these are not regulated by the complexes consisting of either GL1 or TTG1 bound to GL3 [53], unlike the other inhibitors in the system [52, 64]. In the model this overarching ETC fulfils the role of sole inhibitor in the *trycpc* background and as such plays a crucial role in the phenotype. For example, the diffusion rate of ETC has a strong effect on the number of clusters formed and

their size. Furthermore, the introduction of ETC and its regulation by the active complexes in itself constitutes an activator-inhibitor type of sub-network, capable of forming trichome-like patterns independently [38]. Without this sub-motif, the patterns formed in the double mutant simulation would be solely dependent on the activator-substrate type of patterns formed by the interactions between TTG1 and GL3 [40, 183], which would not be sufficient to reproduce the mutant phenotype, showing that the combination of these two motifs and thus the introduction of a third inhibitor are essential.

The resulting model consists of 32 parameters and in order to constrain this space beyond parameter sets that fulfil the phenotype-criteria, we make use of qPCR data on different mutant backgrounds and attempt to reproduce this with the model. This set of qPCR experiments does not only consist of the inhibitor mutants, but also activator-knockouts, and therefore serves as a more general criterium for the model to fulfil. Considering that the model is on the protein-level and the qPCR data on the mRNA-level, we search only for a qualitative agreement, where we classify the expression levels measured by qPCR as either down-regulated, unchanged or up-regulated with respect to wild-type levels. With the aim of keeping the model as simple as possible, we do not extend it by including mRNA-level processes. The parameter sets that result in matching phenotypes as well as matching the qPCR data (Figure 5.4) are considered to be the most accurate representations of the trichome network and are analysed for their mechanistic behaviour in further detail below.

5.3.2 The inhibitors filter pattern frequencies

Having identified a network and parameter sets capable of reproducing aspects of the *try*, *cpc* and *trycpc* phenotypes, we study the behaviour of this network in closer detail. Specifically, we explore the observed opposing differences in trichome densities between the single mutants.

We investigated the finding that *try* has a lower density than *cpc* and turn towards a one-dimensional model of patterning in order to gain an understanding from a mathematical perspective. We determine the dispersion relation of the 1D-system for wild-type and compare this to *try* and *cpc* as a means to predict the number of trichomes that will be formed on the simulated grid (Figure 5.5A). From this we find that the loss of TRY shifts the number of unstable modes in the opposite direction of *cpc*, such that *try* consists of more unstable modes than *cpc*. This suggests that CPC functions as

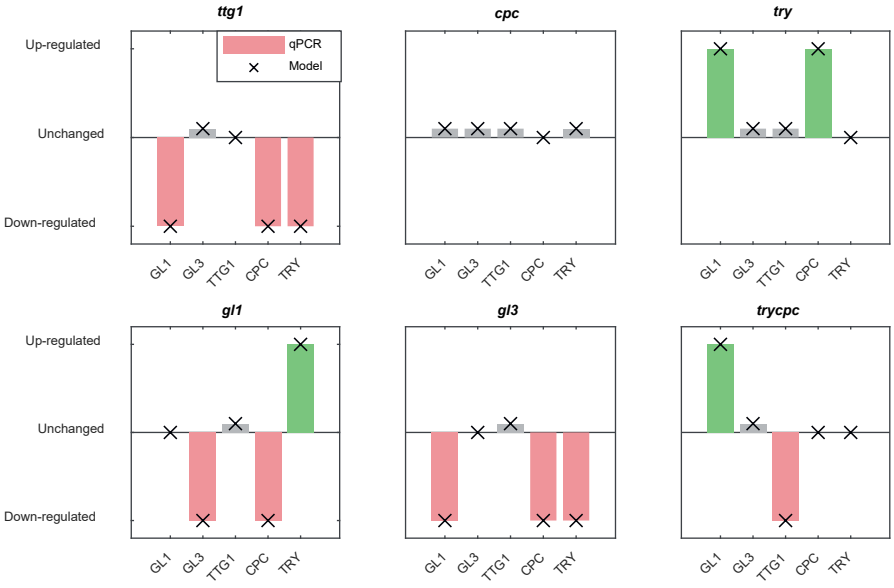


Figure 5.4: **Qualitative fit of model to qPCR data for various mutant backgrounds.**

a filter for high-densities where in its presence, the only allowable modes are relatively high-frequency modes, thus leading to a high density pattern. As a result of lacking this filtering, the *try* background shows a wider range of unstable wavenumbers. This leads to a superposition of these unstable modes after which a lower density pattern is formed. The effect of this superposition is visible in the change in the number of peaks found in the final pattern (Figure 5.5B), thus providing an explanation for the lower trichome density found in the mutant. In *cpc* we see the opposite effect: only relatively high-frequency modes are unstable, forming a high-frequency pattern, i.e. high trichome density (Figure 5.5C).

5.3.3 Stochasticity is required for realistic cluster size distribution

We aim for a more realistic representation of the clusters in *trycpc* and note that there is a discrepancy between the clusters in terms of sizes and distribution thereof. In our model, there is no stochasticity involved in the formation of clusters, thus forming ring-like structures consistent in a size that is

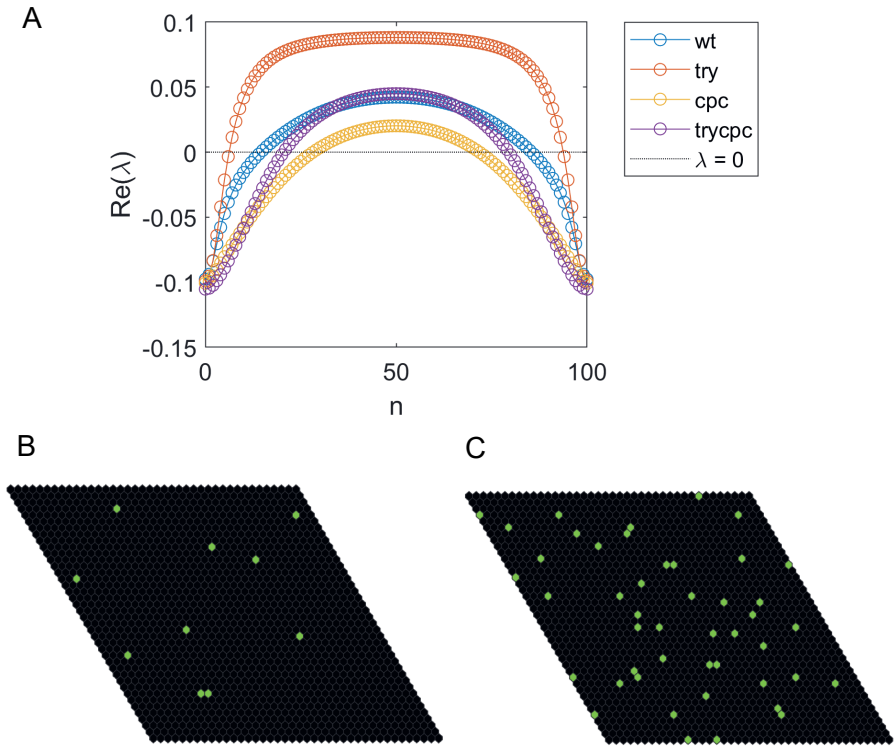


Figure 5.5: **Single mutant inhibitors and dispersion relation.** A: Dispersion relation for a 1-dimensional patterning system and for the different mutant backgrounds. B: Simulation of *try* mutant. C: Simulation of *cpc* mutant.

determined by the number of quasi steady states that are simulated. To capture a degree of stochasticity as is seen in reality, we introduce a probability of differentiation that is based on the concentration of activator complexes inside the cell, as these complexes are considered to bind to promoters of downstream targets that initiate trichome fate [61, 62, 64, 172]. While this detail of probabilistic differentiation introduces the challenge of additional unknown parameters, we aim to fine-tune these to match experimental observations. The probability parameters and the number of quasi steady states simulated is chosen such that the cluster sizes reflect those seen in plants [50] (Figure 5.6B). Furthermore, we qualitatively verify that the relative difference

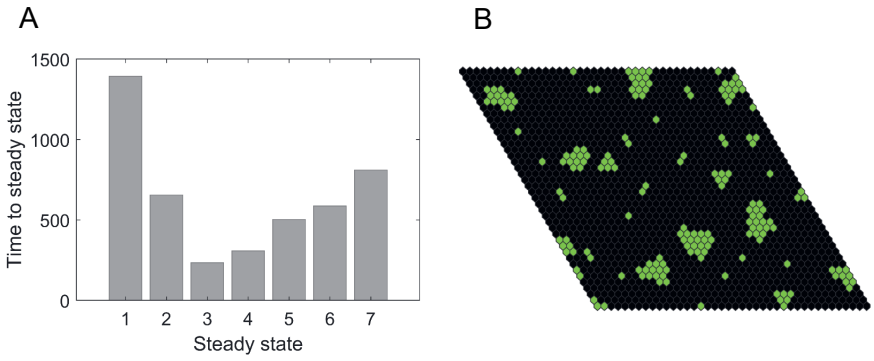


Figure 5.6: **Stochastic cell differentiation and cluster size distribution.** A: Time required to reach steady state for each of the simulated steady states, showing that the first trichomes take the longest time to initiate, followed by shorter times for the trichomes that form the next layers of the cluster. B: Simulation of *trycpc* mutant with stochastic cell differentiation, where differentiation occurs with a probability that depends on the concentration of active complex, as given in (5.18). The pattern shown here follows from the seven quasi steady states shown in A.

in duration before each steady state is reached (Figure 5.6A) reflects the age difference seen in trichomes of developing leaves, where the first trichome is the oldest and the slowest to initiate, followed by trichomes around it appearing in shorter time spans [50].

5.4 Discussion

The *cpc* mutant shows many trichomes and the *try* mutant shows clusters; the double mutant *trycpc* shows many clusters [50]. Although the *trycpc* phenotype at first glance appears a combination of the *try* and *cpc* single mutant phenotypes, upon closer inspection this turns out to be a misleadingly simple assumption. The actual underlying mechanism shows a complexity that is typical for biological systems and intuitively hard to grasp. To aid the exploration of these non-linear mechanisms, we developed a mathematical model based on previous explorations and extended by new observations. The resulting model includes new components and mechanisms which, taken together, give a possible explanation behind the observed phenotypes that were previously not fully understood.

5.4.1 Single mutant phenotypes are non-additive

The explanation behind both single mutant phenotypes relies on their assumed differences in range of effect, where TRY is assumed to be locally inhibiting trichome fate and CPC on a longer scale [50, 59]. We find support for these assumptions both in studying the dispersion relation and in the parameter sets that follow from matching the model with the phenotype and qPCR criteria. However, we know from simulations with a smaller model including only TTG1, GL3, GL1, TRY and CPC that this difference in behaviour is not sufficient to reproduce the double mutant phenotype. We resolved this by introducing a temporal perturbation that mimicked a presumed effect of cell differentiation [186] as well as introducing a third inhibitor [54]. This shows that there is a level of complexity behind the double mutation that goes beyond a mere lack of two inhibitors and that extends to the interconnectivity of parts of the network, both in the form of a mechanism (cell differentiation and activator mobility) as well as additional components (redundancy in the inhibitors). This goes to show that a network such as the one identified for trichomes does not follow linear logic the way it is intuitively often expected, but exhibits a complexity that is much more convoluted. This level of complexity is typical for biological systems but not often so readily accessible as in the case of trichomes. As these phenotypes show similarities to other mutations in plants (e.g. the too many mouths mutant of stomata [189]), the results found here could lead to promising avenues of investigation for these kind of patterning defects.

5.4.2 The model predicts a strong dependency on down-regulation of genes upon cell differentiation

The simulation of cell differentiation was a crucial component in capturing the double mutant phenotype. The idea to include it followed from the observation of differences in age of the trichomes within a cluster [50]. When modelling this effect, we considered two opposite scenarios: up-regulation of patterning genes as a result of endoreduplication [185] or down-regulation of genes that are no longer needed after differentiation and after down-stream targets have been activated [186, 187]. From our simulations we found that only the latter scenario resulted in the ring-like structure of the clusters in combination with the observed difference in initiation time. Note that this addition was not needed to reproduce the single mutant observations and that, in addition, the mobility of activators was required for the most realistic cluster shape. The postulated mechanism behind this down-regulation upon cell

differentiation is that this lowers the production of activators and inhibitors in the trichome peak which in non-differentiated cells shows the highest concentration levels for both the activators and inhibitors. Typically, the levels in the epidermal cells are drastically lower, both in the model and experimental data. After the perturbation, this peak in the prospective-trichome cell is no longer being maintained and so the activators and inhibitors spread to the neighbouring cells, where in the case of the double mutant the level of activators exceeds that of the inhibitor, thus forming a ring-like cluster. In the opposite scenario of endoreduplication, the increased activation of genes would lead to a higher peak, increasing both the amount of activator and inhibitor such that it only sharpens the initial pattern. For this reason, the down-regulation upon differentiation is predicted to be the prevailing mechanism.

5.4.3 Uncertainty in the model

There are some aspects that are not considered in the model but that could play a role or provide an alternative explanation for the patterns, e.g. cell division [185] or intracellular transport [60]. The mechanisms and components that are part of the final model were chosen with the purpose of arriving at a hypothesis that could explain certain aspects of the phenotypes using the least amount and only the simplest of assumptions, most of which followed from biological evidence. As a result of the chosen simplifications, we arrive at a hypothesis about what is at the basis of the patterning mechanism, but cannot claim that it is the full picture . These simplifications were also necessary in order to deal with the large amount of uncertainty surrounding the model. As mentioned earlier, the model consists of 32 parameters, all of which are unknown. We have data from literature to fix certain relative differences between parameters and we fine-tune the parameters involving cell differentiation based on observations, but this still leaves a large amount of uncertainty. We have used a global sampling strategy in order to identify relevant parameter sets and developed the model in incremental steps until the parameter search yielded sets that matched all criteria, ultimately leading to the model presented here.

5.5 Methods

5.5.1 Deriving the mathematical model

The model proposed here is is adapted from a previously published trichome patterning model [183] and extended by the inclusion of ETC1, 2 and 3 into a single term, denoted by ETC [54]. Additionally, we included saturating terms in the feedback loops involving the activating complex between TTG1 and GL3 (AC1) and between GL1 and GL3 (AC2). To reduce the number of parameters and to arrive at a dimensionless system, we rescale the original system of equations by the binding constant between TTG1 and GL3 (β) and the degradation rate of AC1 (λ), i.e. all concentrations were rescaled by the factor β/λ and time is rescaled to $\tau = t\lambda$. This results in the following system of coupled ordinary differential equations:

$$\partial_t[TTG1]_j = \theta_1 - [TTG1]_j(\theta_2 + [GL3]_j) + \theta_2\theta_3\hat{L}[TTG1]_j \tag{5.1}$$

$$\partial_t[GL1]_j = \theta_4 + \frac{\theta_5[AC2]_j}{1+[AC2]_j/\theta_6} - [GL1]_j(\theta_7 + \theta_8[GL3]_j) + \theta_7\theta_9\hat{L}[GL1]_j \tag{5.2}$$

$$\begin{aligned}\partial_t[GL3]_j = {} & \theta_{10} + \frac{\theta_{11}[AC1]_j^2}{1+[AC1]_j^2/\theta_{12}} + \frac{\theta_{13}[AC2]_j^2}{1+[AC2]_j^2/\theta_{12}} \\ & - [GL3]_j(\theta_{14} + [TTG1]_j + \theta_8[GL1]_j + \theta_{15}[TRY]_j \qquad (5.3)\\ & + \theta_{16}[CPC]_j + \theta_{17}[ETC]_j) + \theta_{14}\theta_{18}\hat{L}[GL3]_j \qquad (5.4)\end{aligned}$$

$$\begin{aligned}\partial_t[TRY]_j = {} & \frac{\theta_{19}[AC1]_j^2}{1+[AC1]_j^2/\theta_{20}} + \frac{\theta_{19}/\delta_1[AC2]_j^2}{1+[AC2]_j^2/\theta_{20}} \\ & - [TRY]_j(\theta_{21} + \theta_{15}[GL3]_j) + \theta_{21}\theta_{22}\hat{L}[TRY]_j \qquad (5.5)\end{aligned}$$

$$\begin{aligned}\partial_t[CPC]_j = {} & \frac{\theta_{23}/\delta_2[AC1]_j^2}{1+[AC1]_j^2/\theta_{24}} + \frac{\theta_{23}[AC2]_j^2}{1+[AC2]_j^2/\theta_{24}} \\ & - [CPC]_j(\theta_{25} + \theta_{16}[GL3]_j) + \theta_{25}\theta_{26}\hat{L}[CPC]_j \qquad (5.6)\end{aligned}$$

$$\begin{aligned}\partial_t[ETC]_j = {} & \frac{\theta_{27}[AC1]_j^2}{1+[AC1]_j^2/\theta_{28}} + \frac{\theta_{29}[AC2]_j^2}{1+[AC2]_j^2/\theta_{28}} \\ & - [ETC]_j(\theta_{30} - \theta_{17}[GL3]_j) + \theta_{30}\theta_{31}\hat{L}[ETC]_j \qquad (5.7)\end{aligned}$$

$$\partial_t[AC1]_j = [GL3]_j[TTG1]_j - [AC1]_j \tag{5.8}$$

$$\partial_t [AC2]_j = \theta_8 [GL3]_j [GL1]_j - \theta_{32} [AC2]_j \tag{5.9}$$

Note that the saturation parameters $\theta_{12}, \theta_{20}, \theta_{24}$ and θ_{28} reoccur within the same species for terms invovling activation by AC1 and AC2, which can be interpreted as a limited number of binding sites of the respective promoter. Furthermore, for TRY and CPC we assumed differential regulation by AC1 and AC2, where the activation of TRY by AC2 is reduced by a factor δ_1 and the activation of CPC by AC1 is reduced by a factor δ_2. The values for these reductions are taken from literature [64], where GUS promoter activation assays have shown a difference in activation of TRY and CPC for the different combinations of TTG1, GL1 and GL3. According to these assays, $\delta_1 = 2.5$ and $\delta_2 = 7.2$, which were thus fixed in the model.

The patterning model is simulated on a hexagonal grid and coupling between cells is described by

$$\begin{aligned} \hat{L}[\chi]_{x,y} = &[\chi]_{y-1,x} + [\chi]_{y+1,x} + [\chi]_{y,x-1} + [\chi]_{y,x+1} \\ &+ [\chi]_{y+1,x-1} + [\chi]_{y-1,x+1} - 6[\chi]_{y,x} \end{aligned} \tag{5.10}$$

where χ is the molecular species under consideration and (x, y) the coordinates of the cell indicated by the index j in (5.1) - (5.9).

5.5.2 Simulating cell differentiation

Upon closer inspection of the clusters found in the *trycpc* mutant, there appeared to be a time-separation in the initiation of trichomes. Specifically, a more mature trichome was found in the middle of the cluster and was surrounded by younger, smaller trichomes. To include a similar form of time-separation between the trichomes formed in the model simulations, we ran the model to multiple steady states within a single simulation, where each intermediate steady state was perturbed by local parameter changes. These quasi steady states would simulate the different differentiation stages seen on the leaf. The perturbations meant to simulate cell differentiation are applied only to cells that are designated as trichomes, which were identified based on a threshold-mechanism where a certain amount of AC1 and AC2 is expected to lead to a change in cell fate. In our initial surveys, this threshold was taken to be the half-maximum of the sum of AC1 and AC2 found at steady state, i.e. the set of cells with index j that are designated as trichomes is given by

$$\mathcal{T} = \left\{ j \in \mathcal{J} \;\middle|\; \mathrm{AC}_j \geq \frac{1}{2} \mathrm{AC}_{\max} \right\}. \tag{5.11}$$

where $\mathcal{J}$ is the set of all cells on the grid. Later, we refined this to a probabilistic threshold, described in more detail in section 5.5.4.

The perturbations are of a localized nature, meaning that only cells $j \in \mathcal{T}$ undergo changes to the parameters while all $j \notin \mathcal{T}$ remain unchanged. We consider two contrasting cases with regards to these parameter changes: 1) due to endoreduplication the production and activation parameters are increased in trichome cells and 2) after acquiring the cell fate, the patterning proteins are no longer required and production and activation terms are turned off in trichome cells. The parameters which are changed in both these cases are θ_1, θ_4, θ_5, θ_{10}, θ_{11}, θ_{13}, θ_{19}, θ_{23}, θ_{27} and θ_{29}. Note that parameters like protein binding, transport and degradation remain unchanged. In scenario 1 these parameters are increased 4-fold to reflect a 4-fold increase in DNA-content; and in scenario 2 the aforementioned parameters are set to 0. After this perturbation, the system is shifted away from the steady state and so we rerun the simulation until the next steady state is reached. The number of times this is repeated will influence the size of the clusters found in the double mutant.

5.5.3 Parameter search

The model described in (5.1) - (5.9) consists of 32 parameters, all of which are unknown. To deal with this large amount of uncertainty, we employ a global parameter search involving a non-linear constrained optimization routine. This search has to satisfy multiple objectives and constraints, where the objectives are formulated in a way that represents the *try*, *cpc*, *trycpc* mutant phenotypes in terms of relative trichome density and clustering, as well as matching the up- and down-regulation pattern of the qPCR data. The contraints are defined to ensure that a parameter set stays within the viable patterning space by making use of linear stability anaylsis. Towards this end, we make use of a genetic algorithm (GA) that uses a controlled, elitist algorithm, which means that it generally favours individuals with a higher fitness value while at the same time allowing for lower fitness scores if these individuals help increase the diversity of the population [190].

The optimization problem considered here is a multi-objective problem, which in its general form is formulated as

$$\min_{\boldsymbol{\theta}} (f_1(\boldsymbol{\theta}), f_2(\boldsymbol{\theta}), \ldots, f_k(\boldsymbol{\theta})) \tag{5.12}$$

where the f_i are the individual objective functions, $k \geq 2$ the number of objectives and $\boldsymbol{\theta}$ the parameter vector. The optimization problem for the inhibitor mutants considered here consists of four objectives. First, we minimize the distance between the qPCR expression levels and the model predictions through a sum-of-squares cost function

$$f_1(\boldsymbol{\theta}) = \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{(\bar{y}_i^j(\boldsymbol{\theta}) - y_i^j)^2}{{y_i^j}^2} \tag{5.13}$$

where $\bar{y}_i^j(\boldsymbol{\theta})$ is the expression level predicted by the model of the i-th gene out of N total genes for the j-th mutant out of M total mutants, and y_i^j is the corresponding datapoint. For details on how the qPCR data is used in combination with the model we refer to section 5.5.5.

The second objective is aimed at capturing the appropriate trichome densities (relative to wild-type) for each of the three mutants:

$$f_2(\boldsymbol{\theta}) = \sum_{i=1}^{3} (\bar{\rho}_i(\boldsymbol{\theta}) - \rho_i)^2 \tag{5.14}$$

where $\bar{\rho}_i(\boldsymbol{\theta})$ is the trichome density predicted by the model for the i-th inhibitor mutant and ρ_i is the experimentally observed trichome density for the *try*, *cpc* and *trycpc* phenotypes. Note that for the *trycpc* phenotype trichome density is counted as the number of clusters, as individual trichome numbers are indiscernible in the mutant leaves due to being too tightly packed.

In the third objective we aim to optimize the cluster densities in wild-type and mutants to the experimentally observed densities by

$$f_3(\boldsymbol{\theta}) = \sum_{i=1}^{4} (\bar{\nu}_i(\boldsymbol{\theta}) - \nu_i)^2 \tag{5.15}$$

where $\bar{\nu}_i(\boldsymbol{\theta})$ is the cluster density predicted by the model for the i-th genotype and ν_i is the experimentally observed cluster density for wildtype, *try*, *cpc* and *trycpc*.

Finally, the fourth objective describes the clustering behaviour of the *trycpc* phenotype given deterministic cluster formation:

$$f_4(\boldsymbol{\theta}) = 1 - r(\boldsymbol{\theta}). \tag{5.16}$$

In this objective, $0 \leq r \leq 1$ is a measure of cluster packing, where $r = 1$ means that all cells surrounding a trichome are differentiated into trichomes upon the next steady state and $r = 0$ means none of the neighbouring cells differentiate upon the next steady state.

Taken together, these four objectives guide the GA to approximate the Pareto front [190]. Furthermore, the optimization is subjected the non-linear constraint

$$c(\boldsymbol{\theta}) = \mathrm{Re}(\lambda_{\max}) \tag{5.17}$$

where $\lambda_{\max}$ is the largest eigenvalue of the Jacobian. By determining whether the real part of the largest eigenvalue is positive (i.e. $c(\boldsymbol{\theta}) \geq 0$), we learn that the parameter set $\boldsymbol{\theta}$ can form a pattern [123], which constrains the allowable range of parameters. Note that this range is also constrained by the choice of bounds on the parameters θ. In this case, we set the interval for the each of the parameters in $\boldsymbol{\theta}$ to $[0.01, 100]$, to allow a range of multiple orders of magnitude.

The GA requires an initial population to start the search from. We generated this initial population through a quasi Monte Carlo approach where we generated 10 million random parameter sets using a Sobol sequence [90] and quantified the mutant patterns for each of these sets. The top 200 sets that performed best on all four of the objectives described above were used as initial population for the GA.

5.5.4 Probabilistic cell differentiation

Identifying trichomes by a threshold mechanism as described in (5.11) is somewhat arbitrary in the choice of threshold level and is unlikely to occur in such a deterministic fashion in real cells. While this is a useful simplification for the parameter search described above, for simulations aimed to more accurately reproduce the *tryCPC* phenotype we decided to introduce a probability of differentiation that is dependent on the amount of active complex in the cells, which is assumed to drive the expression of down-stream targets responsible for deciding cell fate.

The probability of differentiation of cell j is given by

$$p_j(x_j) = \frac{1}{2}\left(1 + \mathrm{erf}\frac{x_j - \mu}{\sqrt{2}\sigma}\right) \tag{5.18}$$

where x_j is the sum of AC1 and AC2 in cell j, and σ and μ determine the width and centring of the error function [191]. The parameters σ and μ are open parameters in this problem and are chosen to be $\mu = \frac{1}{2}\mathrm{AC}_{\max}$ and $\sigma = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(x_j - \mu_{\mathrm{AC}})^2}$ where N is the number of cells on the grid and μ_{AC} is the mean of the sum of AC1 and AC2 across the grid, i.e. σ is the standard deviation of the sum of AC. Note that μ and σ are determined by the parameter set that reflects wildtype and is kept constant for simulations of the mutant phenotypes, i.e. the first steady state in the wildtype scenario determines the parameters of (5.18) for all subsequent simulations including the mutants. However, between varying parameter sets, μ and σ will differ, depending on the amount of AC.

5.5.5 Fitting the model to qPCR data

For each of the species in the model, expression levels have been determined in the respective mutant backgrounds. In Supplementary table 1, the expression levels of *ttg1*, *cpc*, *try*, *gl1*, *gl3* and *trycpc* are given relative to wildtype levels. As shown in Figure 5.4, we simplify these levels to one of three states: down-regulated, up-regulated or unchanged. It is this pattern of three possible levels among 6 mutants that we try to match with the model, where we assume that these discrete, relative differences on the mRNA level are reflected by the protein levels as they are described by the model.

Given that some of these mutant phenotypes are glabrous (no trichomes are formed), we check that these conditions are also true for the parameter sets considered. This means that parameter sets that simulate the knock-out of GL3 (glabrous leaves) should not produce any patterns, i.e. through using linear stability analysis we determine that these parameter sets fall outside of the Turing space [123]. To still obtain a prediction on up- or down-regulation, we simulate these glabrous mutants in a single-cell model and compare this to the average across the grid of the wildtype simulation. For all other cases where mutants do form a pattern, we average over 12 simulations on the grid with randomized initial conditions.

5.6 Supplementary Information

5.6.1 Supplmentary table 1

Table 5.1: qPCR data relative to wildtype for different mutant backgrounds.

	GL1+MYB23	GL3+EGL3	TTG1	CPC	TRY
ttg1-21	0.201	0.951	0.000	0.244	0.358
cpc-2	1.143	1.426	0.905	0.000	1.049
try	2.555	0.617	0.987	2.570	0.000
gl1 myb23	0.000	0.241	1.045	0.240	1.917
gl3-3 egl3	0.073	0.000	0.878	0.077	0.388
cpc-2 try	3.803	0.541	0.497	0.000	0.000

Chapter 6

Genetic and molecular analysis of trichome development in *Arabis alpina*

Divykriti Chopra, Mona Mapar, Lisa Stephan, Maria C. Albani, Anna Deneer, George Coupland, Eva-Maria Willing, Swen Schellmann, Korbinian Schneeberger, Christian Fleck, Andrea Schrader, Martin Hülskamp

Published in: PNAS 116, 24 (2019).
www.pnas.org/cgi/doi/10.1073/pnas.1819440116

6.1 Abstract

The genetic and molecular analysis of trichome development in *Arabidopsis thaliana* has generated a detailed knowledge about the underlying regulatory genes and networks. However, how rapidly these mechanisms diverge during evolution is unknown. To address this problem, we used an unbiased forward genetic approach to identify most genes involved in trichome development in the related crucifer species *Arabis alpina*. In general, we found most trichome mutant classes known in *A. thaliana*. We identified orthologous genes of the relevant *A. thaliana* genes by sequence similarity and synteny and sequenced candidate genes in the *A. alpina* mutants. While in most cases we found a highly similar gene-phenotype relationship as known from *Arabidopsis*, there were also striking differences in the regulation of trichome patterning, differentiation and morphogenesis. Our analysis of trichome patterning suggests that the formation of two classes of trichomes is regulated differentially by the homeodomain transcription factor *AaGL2*. Moreover, we show that overexpression of the GL3 bHLH transcription factor in *A. alpina* leads to the opposite phenotype as described in *A. thaliana*. Mathematical modeling indicates that this non-intuitive behavior can be explained by different ratios of GL3 and GL1 in the two species.

6.2 Significance statement

Typically, comparative developmental studies focus on the question whether key genes and the underlying processes are conserved over large evolutionary distances. How the underlying gene regulatory network of a whole developmental process evolves is virtually unknown. This requires a functional comparison (e.g. by genetics) of homologous processes in species that are closely related enough to find similarities and distant enough to see functional changes. We approach this question by genetically and molecularly comparing trichome development in *Arabidopsis thaliana* with that in *Arabis alpina*. We show which steps are regulated similarly and which are different. A closer analysis of trichome patterning revealed fascinating changes in the gene regulatory network. In particular the finding that overexpression of one key regulator leads to opposite phenotypes in the two species suggested fundamental changes.

6.3 Introduction

Evolutionary genetics has been widely used to study how molecular changes of gene functions or regulatory networks lead to phenotypic differences [74, 80, 192]. To understand the source of molecular variation that ultimately results in phenotypic changes, micro-evolutionary approaches [79] and comparative studies of closely related species have been used [74, 80]. These approaches enable an understanding of functional changes of genes in developmental processes. In this work we study the evolution of leaf trichome development. Leaf trichome development is best studied in *A. thaliana*. The characterization of individual mutants in closely related species suggests that trichome development has a common basis in crucifer [72, 73]. Yet, within the crucifer family, trichomes exhibit a wide range of trichome density and morphology phenotypes that correlate well with the phylogenic tree [193]. It is therefore conceivable that the basic regulatory machinery for trichome development is conserved in the crucifer family but exhibits significant variation. We therefore reasoned that functional comparison of trichome development of *A. thaliana* with that of another evolutionarily distant crucifer species should enable us to recognize functional differences or diversifications. Towards this end we chose *A. alpina* as a second genetic model for trichome development. *A. alpina* is diverged about 26 to 40 million years from *A. thaliana* and is in a distinct clade in the Brassicaceae family [194–196]. These species might therefore be distant enough to find differences in the regula-

tion of trichome development, while being closely enough related to compare orthologous gene functions. To approach this problem, we decided to first perform a genetic screen aiming to identify gene functions based on mutant phenotypes and then to compare the spectrum of phenotypes to that known in *A. thaliana*. In *A. thaliana*, trichomes develop on young leaves in regular patterns and are separated by protodermal cells [34]. Incipient trichomes proceed through several rounds of DNA replication in the absence of cell divisions (endoreduplication) and emerge from the leaf surface [34]. Typically three or four branches are formed in a regular arrangement before the cell expands to its final size [34]. Genetic screens and subsequent molecular analyses have revealed most genes relevant for the initiation and development of trichomes [133, 169, 197]. Based on mutant phenotypes, various developmental steps have been defined [34]. The first step is the selection of individual epidermal cells among others to adopt a trichome cell fate, which involves positive and negative regulators of trichome development. They are considered to generate a trichome pattern through intercellular interactions [35, 36]. The initiation of trichome differentiation is a separate step and requires the GLABRA2 gene [63]. Various genes are important for the proper endoreduplication driven cell enlargement [197–200]. Trichome branching is governed by several genes that regulate either the number or the position of branches [201, 202]. Regular cell expansion is defined by a group of mutants that are characterized by grossly distorted trichomes [203]. Finally, maturation of trichomes as recognized by the formation of small papilla on their outer surface is dependent on genes of the glassy group [34, 204].

Leaf trichome development in *A. alpina* is reminiscent to that in*A. thaliana* [73]. In contrast to *A. thaliana*, two types of trichomes are formed on *A. alpina* leaves. One class of small trichomes is regularly distributed and a second class of much larger trichomes is interspersed with regular distances to each other [73]. Leaf trichome development is readily accessible to genetic approaches [73] and the molecular analysis is facilitated by the availability of a fully sequenced genome [196, 205] and because *A. alpina* can be transformed by Agrobacterium mediated floral dip [206].

In this study, we systematically screened an EMS mutagenized population of *A. alpina* plants for trichome mutants. We identified mutants affecting most steps of trichome development, as previously reported for *A. thaliana*. Orthologous genes in *A. alpina* were identified based on sequence similarity and synteny and candidate genes were sequenced in *A. alpina* mutants. While most gene-phenotype relations in *A. alpina* were the same as in *A. thaliana*,

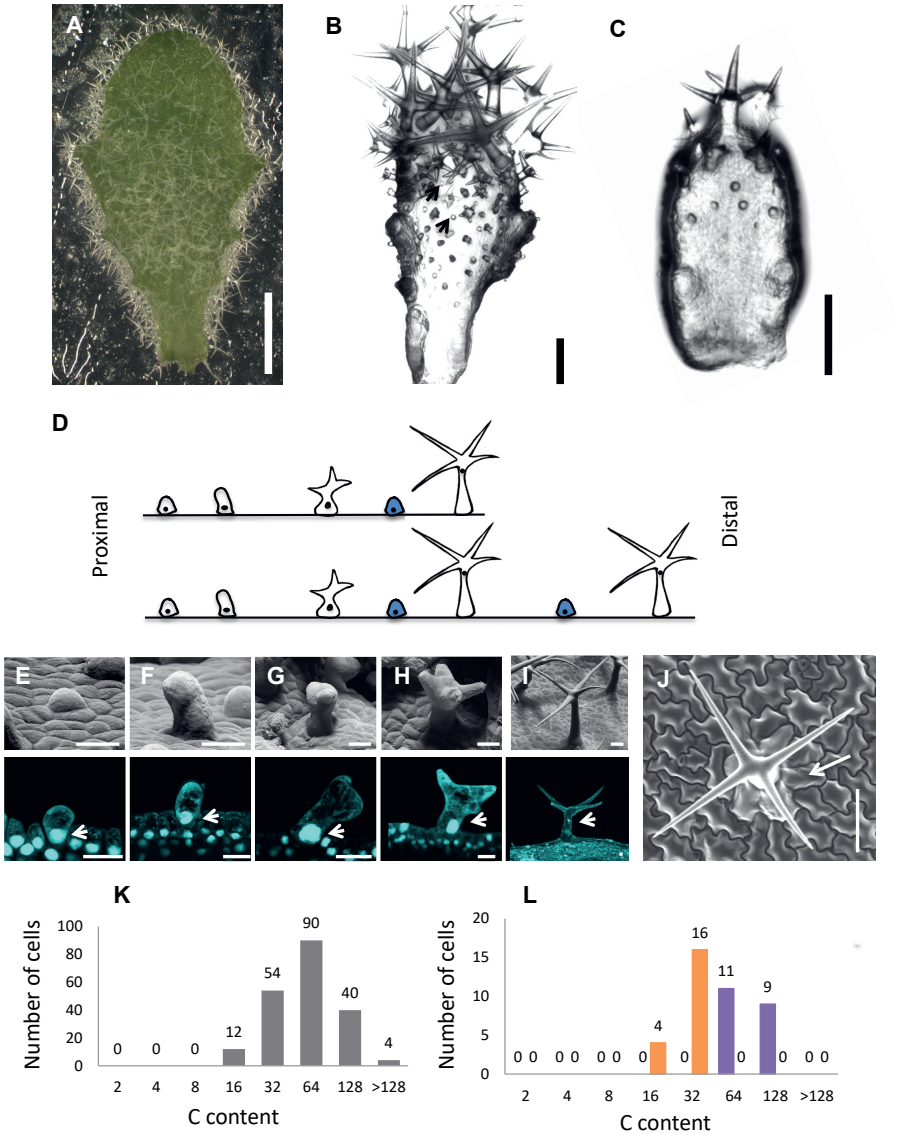
we found some notable differences in the regulation of trichome patterning, differentiation and morphogenesis.

6.4 Results

6.4.1 Trichome development in *A. alpina*

Trichomes in *A. alpina* densely cover the surface of adult leaves during vegetative development (Figure 6.1A). They are initiated at the base of young leaves and in more distal regions between already existing trichomes ([73], Figure 6.1B-D). At the leaf base, incipient trichomes are typically formed 4-5 cells away from each other. Subsequently, trichomes are spatially separated due to expansion and cell divisions of the intervening epidermal cells. In more distal regions, new trichomes are formed between more advanced trichome stages. These are typically formed in a similar distance as between incipient trichomes at the leaf base (typically 4-5 cells).

Similar as in *A. thaliana* [34, 207], trichome development proceeds through a series of characteristic stages (Figure 6.1E-I). Incipient trichomes begin to expand and grow perpendicular to the leaf surface (Figure 6.1E,F). Subsequently, trichomes typically undergo three branching events (Figure 6.1G,H) followed by extensive elongation growth. During maturation, numerous papillae are formed on the trichome surface. It is noteworthy, that the cells immediately next to a mature trichome are shaped like normal epidermal pavement cells (Figure 6.1J). This suggests that in contrast to *A. thaliana* [208], these cells do not differentiate into socket cells. Cell growth in *A. thaliana* occurs concomitant with an increase in nuclear size and DNA content due to several rounds of endoreduplication, leading to an average DNA content of 32C in mature trichome cells [199, 209]. We used DAPI staining to study the nuclear size in different stages of trichome development in *A. alpina*. Incipient unbranched trichomes already show increased nuclear sizes (Figure 6.1E,F). A further increase in nuclear size was observed during branch formation (Figure 6.1G,H). Similar as in *A. thaliana* [201, 210], the nucleus moves up to the branch points at this stage. We determined the relative DNA content of nuclei in mature trichomes by comparing the DAPI fluorescence in trichomes with that in stomata. In *A. thaliana*, stomata have a DNA content of 2C (30) and were therefore used as a reference to judge the C value of trichome cells [210]. Using this strategy, we found an average C-value of 64C in *A. alpina* trichomes (Figure 6.1K). When analyzing the population of the larger trichome class separately from the smaller trichome class, we found a clear separation such that



(Caption on next page.)

Figure 6.1: **Trichome development on *A. alpina* rosette leaves. .** (A) Mature leaf; trichomes densely cover the whole surface. (B) In slightly older stages, incipient trichomes are found between older trichomes (arrowheads). (C) On very young leaves incipient trichomes are found at the leaf base and advanced developmental stages in more distal regions. (D) Schematic representation of the trichome distribution along the proximal-distal axis on a very young leaf as shown in (C top) and an older leaf as shown in (B bottom). Blue colored trichomes are intercalating between already existing ones. (E-J) Scanning Electron micrographs (upper pictures) and optical section of DAPI stained trichomes (lower pictures) at different developmental stages. (E) Incipient trichomes beginning to expand. (F) Unbranched trichomes. (G) Two-branched trichomes. (H) Four-branched trichomes. (I) Mature trichomes. (J) Top view of a mature trichome. Note that the immediately adjacent cells are shaped like pavement cells. (K) Distribution of the C content of all mature trichomes. It ranges between 16C to 128C (n=200). (L) Small (orange) and large trichome classes (purple). n=20.Scale bars=10µm (A, B), 20µm (C, D), 10µm (E-I), 100µm (J).

the smaller trichome class has on average 16C and 32C and the larger trichome class 64 to 128C (Figure 6.1L). This suggests that the small trichomes undergo on average four endoreduplication rounds and the larger trichomes five or six cycles.

6.4.2 Isolation of trichome mutants in *A. alpina*

We took a forward genetic approach to dissect trichome development into functional steps. Towards this end, seeds of five M1 plants from two independent EMS populations were pooled and about 48 M2 plants from each pool were screened for trichome phenotypes. The first screen was done in an *A. alpina* Pajares population representing 4165 M1 plants [206]. The second screen was done in the *A. alpina* flowering time pep1-1 background that does not have an obvious trichome phenotype compared to Pajares [206, 211–213]. Here we screened an M2 population derived from 6800 M1 plants. Trichome mutant phenotypes were confirmed in the M3 generation. Our screens yielded 49 trichome mutants. The mutant spectrum of patterning and morphogenesis mutants was similar to that in *A. thaliana* [210] though we did no find higher trichome density mutants and glassy mutants that are otherwise normal in shape were not found (Table S1).

We identified 13 mutant lines in which trichome initiation and/or their distribution is affected. Eleven mutants show no trichomes on the leaf blade of four-week old rosette leaves and variable numbers of trichomes at the leaf

margin. Two lines were already reported to be *Aattg1* alleles [73]. We excluded the presence of very small and aborted trichomes by analyzing the leaf surface at a high magnification. The second class of patterning mutants comprises two lines in which trichomes occur in clusters with the individual trichomes being larger and over-branched (Figure 6.2B,D). Trichome clusters contained two to three trichomes immediately next to each other. We also found trichome clusters on young leaves indicating that the corresponding gene is important to single out one trichome cell during trichome initiation (data not shown).

We identified one mutant that initially appeared glabrous (Figure 6.2E), however on closer inspection revealed two classes of trichomes: extremely small and aborted trichomes and wild-type like small trichomes (Figure 6.2F). The small trichomes were reminiscent of young unbranched trichomes. Their shape, however, differed from young wild-type trichomes in that their cell body exhibited a puzzle piece like form, that is characteristic of epidermal pavement cells (Figure 6.2G). This suggests that trichome initiation was not affected and that trichomes have a mixed trichome/pavement fate. The wild-type trichomes were regularly distributed on the leaf surface at large distances with typically one or two small trichome cells in between (Figure 6.2H). This phenotype suggests the presence of two types of genetically distinct trichomes in *A. alpina* that have not been observed in *A. thaliana*.

In addition, we isolated a number of mutants affecting the morphogenesis of trichomes. One class comprising 24 mutants exhibited defects in branch number or their spatial arrangement. A second class showed distorted and twisted trichomes (Supplemental text, Figure S1,2, Table S2)

6.4.3 Identification of trichome genes in *A. alpina*

Given that *A. alpina* and *A. thaliana* are crucifers that diverged only about 26 to 40 million years ago [194, 195], we assumed that the majority of the mutant phenotypes found in our screen are caused by mutations in the genes orthologous to the respective *A. thaliana* genes. In a first step, we searched for putative orthologous genes in the *A. alpina* genome based on sequence similarity and synteny [214]. By these criteria, we identified orthologs of the selected *Arabidopsis* genes in the *A. alpina* genome (Supplemental text, Figure S3,4, Table S3).

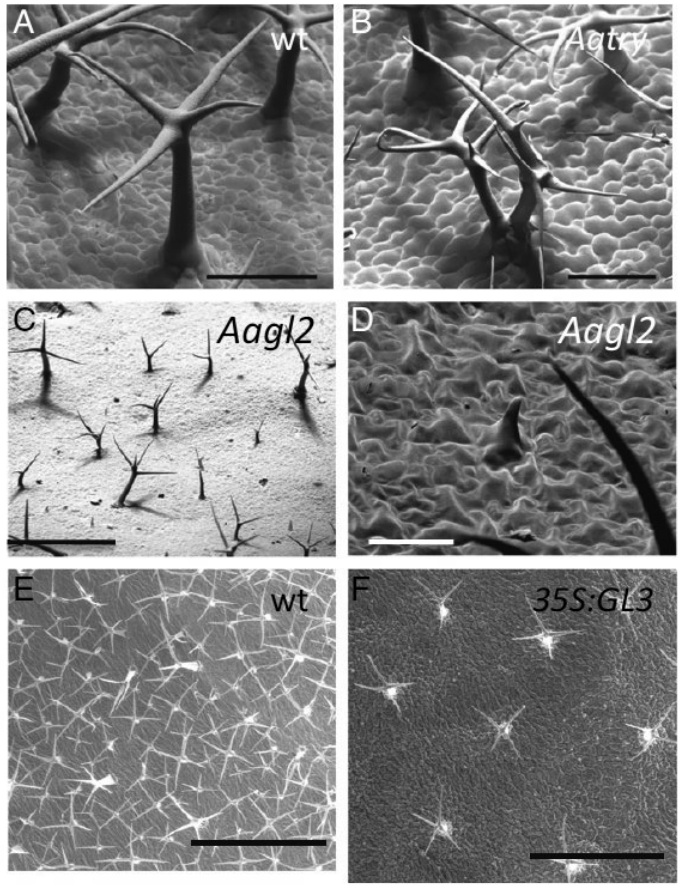


Figure 6.2: **Phenotypes of patterning mutants.** (A) Scanning electron microscope (SEM) picture of a mature wild-type trichome. (B) SEM picture of a mature *Aatry* mutant trichome. (C) SEM picture of an *Aagl2* mutant leaf showing the distribution of large wild-type shaped trichomes and small aborted trichomes. (D) SEM picture of an aborted *Aagl2* mutant trichome. (E) Wild-type *A. alpina* (Pajares) leaf with large and small trichomes. (F) 35S:AaGL3 with large trichomes. (Scale bars: A and B: 100 μm; C: 500 μm; D: 50 μm; E and F: 1 mm.)

6.4.4 Identification of mutant-specific alleles

Under the assumption that mutations in *A. alpina* trichome genes lead to the same phenotype as in *A. thaliana*, one would expect to find relevant mutations

in the respective genes in the trichome mutants. We therefore sequenced candidate genes in selected trichome mutants focusing on the unambiguous phenotypes, similar as done before with *AaTTG1* [73]. As summarized in Table S4 we found 23 mutant-specific alleles.

The analysis of the eleven glabrous mutants revealed three lines with mutations in the *AaTTG1* gene, as reported before [73]. In the remaining eight mutants we discovered relevant mutations in the *AaGL3* gene (Figure S5). Mutations included changes leading to acceptor splice site, premature STOP codons or amino acid exchanges. These findings were unexpected, as Atgl3 mutants have trichomes. Atgl3 Ategl3 double mutants are completely glabrous indicating that GL3 and EGL3 act in a partially redundant manner in *A. thaliana* [136]. Sequencing the *AaEGL3* gene in four gl3 alleles revealed no mutations (Table S5). Thus in *A. alpina*, AaGL3 possesses the full bHLH function in trichome patterning and AaEGL3 does not appear to be functionally relevant in this context. However, rescue experiments show that the *AaEGL3* protein can rescue the *Atgl3 Ategl3 Arabidopsis* mutant efficiently when expressed under the 35S promoter indicating that the protein is fully functional in this context (Figure S6). We also sequenced the AaGL1 gene in all glabrous mutants. None displayed mutations in the *AaGL1* gene (Table S5).

Two patterning mutants exhibited trichome clusters reminiscent to try mutants in *A. thaliana*. Sequence analysis revealed a premature STOP codon in one mutant and a Leucine to Phenylalanine change in a conserved region in the other. Both mutations are expected to lead to severe defects in protein function, indicating that the two mutants are *Aatry* alleles (Figure S5).

The differentiation mutant displaying large and fairly normal trichomes plus small underdeveloped trichomes has no clear counterpart in *A. thaliana*. However, the population of underdeveloped trichomes shares similarities with *gl2* trichomes in *A. thaliana*. When sequencing the *AaGL2* gene in this line, we found a deletion of one base pair in exon 4 leading to a premature STOP codon at amino acid position 348 (Figure S5).

The molecular analysis of the morphogenesis mutants also revealed mutations in several known trichome morphogenesis genes (Supplemental text, Figure S5).

6.4.5 Analysis of the GL3 function in *A. alpina*

Our finding that mutations in GL3 result in a strong trichome phenotype in *A. alpina* whereas in *A. thaliana* an additional mutation in EGL3 is required for a glabrous phenotype, suggests a functional divergence of GL3 and EGL3 between the two species. In *A. thaliana* it was reported that *Atgl3* single mutants exhibit a weak trichome phenotype but none of the other *ttg1* traits. The *Atgl3 Ategl3* double mutants strongly enhance the trichome phenotype, show ectopic root hairs, lack anthocyanin and show a weak defect in seed coat mucilage production [136]. Seed color is normal in *Atgl3 Ategl3*. Together this suggests that *AtGL3* and *AtEGL3* act redundant in four TTG1 traits. As *Aagl3* single mutants in *A. alpina* display the same trichome phenotype as the *Atgl3 Ategl3* mutant in *A. thaliana*, we determined whether representative *Aagl3* alleles of *A. alpina*, *Aagl3-1* and *Aagl3-2*, show additional phenotypes. Seed color was normal in both *Aagl3* alleles (Figure S7A-C). The analysis of root hair formation revealed hairy roots (Figure S7 D-F). A closer inspection showed that all files including the N-files produced almost only root hairs. Thus, similar as observed for trichomes, *Aagl3* mutants produce the full AaTTG1 root hair phenotype suggesting that EGL3 has also no relevant function in root hair patterning. In contrast, seed coat mucilage production and anthocyanin production in seedlings grown on 3% sucrose was unaffected in both alleles (Figure S7 G-L).

To support the idea that the observed mutant phenotypes of *Aagl3-1* and *Aagl3-2* alleles are caused by mutations in the GL3 gene, we performed a complementation test. F1 plants showed a glabrous trichome phenotype and hairy roots (Figure S8), indicating that the causal mutations are allelic.

To better understand the function of *AaGL3* in *A. alpina*, we generated transgenic plants expressing the *AaGL3* gene under the 35S promoter. In *A. thaliana* overexpression of *AtGL3* leads to extra trichome formation [44] and reduced root hair numbers [69] supporting the idea that *AtGL3* promotes trichome formation and non-root hair formation. We recovered two plants in which we found 30 fold and 10 fold higher expression levels of *AaGL3* as compared to wild type in qRT-PCR experiments (Table S6-8). In the next generation, single plants were analyzed with respect to the trichome and root hair phenotypes (Table S9-11). In four individual plants we confirmed the overexpression of *AaGL3* by qRT-PCR (Table S7,8). In both independent lines we found reduced root hair formation (Figure S9) further supporting that *AaGL3* promotes non-root hair fate also in *A. alpina*. Trichome density was deter-

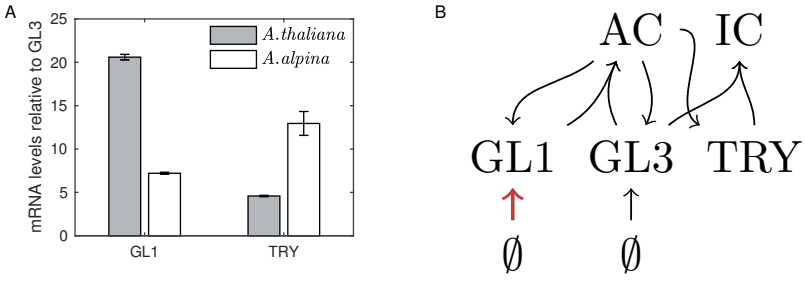


Figure 6.3: **Modeling *A. thaliana* and *A. alpina* wild-type and 35S:GL3 phenotypes.** (A) mRNA levels determined by qPCR of GL1 and TRY relative to GL3 for *A. thaliana* (gray bars) and *A. alpina* (white bars). Note that in *Arabis* we find relatively low levels of GL1 and relatively higher levels of TRY compared with *Arabidopsis*. (B) Schematic network of the model. The GL1 basal production parameter is decreased (red arrow) to reproduce *A. alpina* phenotypes.

mined on the third leaf when they reached a length of 0.6 to 1 cm. Trichome density was reduced to 10% of that found in wild type (Table S9-11). This is in sharp contrast to *A. thaliana* where overexpression of *AtGL3* causes extra trichome formation [44]. The reduced trichome density was also observed on older leaves. Trichomes were fairly similar in size and regularly distributed (Figure 6.2).

We reasoned that this non-intuitive difference in the response to GL3 overexpression might be correlated with differences in the relative expression levels of the core patterning genes GL1, GL3 and TRY. To test this assumption, we compared their expression between *Arabidopsis* and *Arabis* in young leaves corresponding to stages shown in Figure 6.1B by qPCR analysis. When comparing the relative expression levels of GL1 with that of GL3 we found a large difference (Figure 6.3A). As compared to *Arabidopsis* the relative levels of GL1 levels were strongly reduced. To explore whether reduced GL1 levels relative to GL3 are sufficient to explain the different responses to GL3 overexpression in *Arabis*, we used a modeling approach. We use a previously published model that includes GL1 and GL3 that together form the activator complex (AC) [38]. The AC activates TRY, which in turn forms the inhibitor complex (IC) together with GL3 (Figure 6.3B) (35).

Using this model, we screened for parameter sets that lead to increased tri-

chome numbers upon a simulated overexpression of GL3 (*A. thaliana* situation (Figure 6.4A,C,E)). These sets were subjected to a second screen for parameter sets in which the expression of GL1 was decreased and overexpression of GL3 leads to less trichomes (*A. alpina* situation (Figure 6.4B,D,F)). Our results show that the relative expression difference of GL1 compared to GL3 can explain the different response to the overexpression of GL3 in the two species. A closer analysis of the model revealed that in both, the *Arabidopsis* and the *Arabis* situation, overexpression of GL3 leads to an increase of the inhibitor and activator activity, however, in the *Arabis* situation, the relative increase of the inhibitor activity is more pronounced around the activator peaks (Figure S10). This is consistent with the qPCR results in which the relative levels of TRY are higher in *Arabis*. As a result, in the *Arabidopsis* situation overexpression of GL3 effectively leads to less repression and therefore a higher density. By contrast, in the *Arabis* situation, only the strongest peaks survive and the lower peaks are suppressed over time. This is reminiscent to presence of the large trichomes and the lack of small trichomes in 35S:GL3 *Arabis* plants.

6.5 Discussion

In this study, we set out to systematically identify all genes involved in trichome development in *A. alpina* by a forward genetic mutagenesis screen to enable a comparison with the genetic inventory in *A. thaliana*. We reasoned that the evolutionary distance between the two species is close enough to compare orthologous processes and genes, yet distant enough to expect differences. Our phenotypic comparison supports this hypothesis as trichome patterning and morphogenesis in *A. alpina* is almost indistinguishable from that in *A. thaliana* except for the presence of two differently sized trichome classes and the absence of morphologically distinct accessory cells. We therefore expected genetic and molecular differences explaining the two classes of trichomes. For all other patterning and morphogenetic aspects, we assumed not to find major differences in the function of the involved genes.

6.5.1 How well does the mutant spectrum reflect the gene inventory?

EMS mutagenesis screens can be considered to induce random mutations in the genome and as a consequence the frequency of deleterious mutations should be similar in all genes. It is therefore in principle possible to judge the saturation of the screen by the allele frequency. Our sequence analysis of candidate genes in the respective mutants revealed on average 2.6 mutations

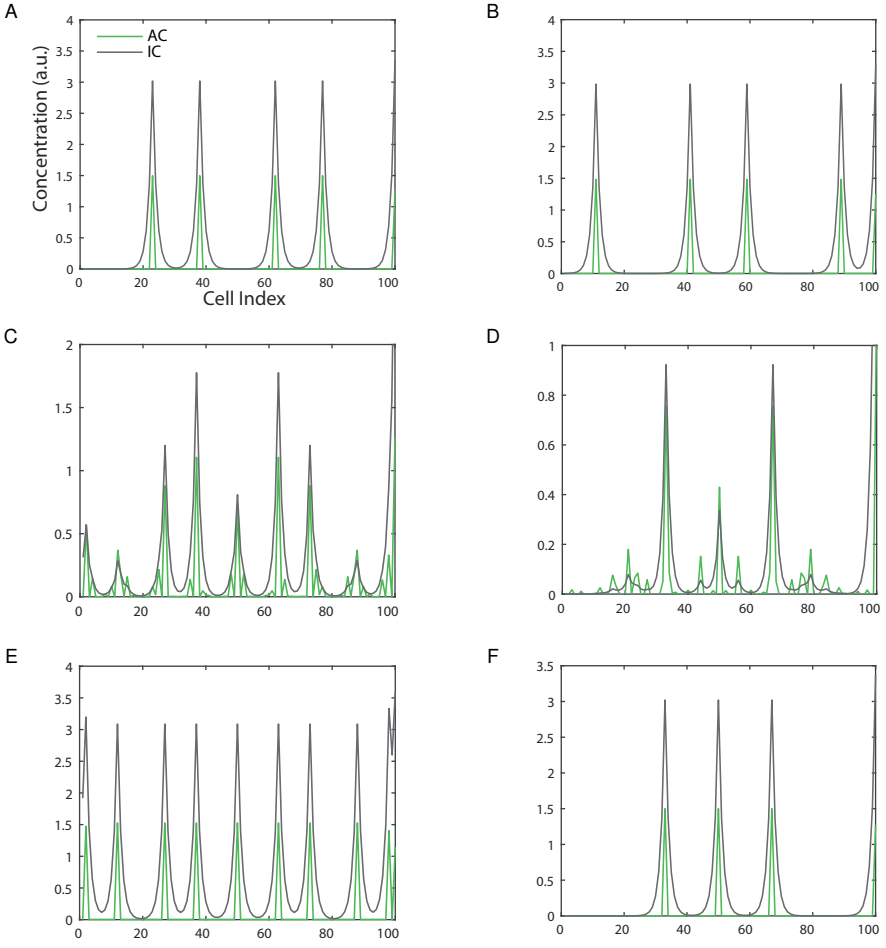


Figure 6.4: **Simulations of *A. thaliana* and *A. alpina* wild-type and 35S:GL3 phenotypes.** Simulation examples of (A) *A. thaliana* wild-type and (B) *A. alpina* wild-type on a 1D grid discretized into 100 cells (x-axis). Concentrations of activating complex (AC) and inactive complex (IC) are indicated by green and black lines, respectively. C and E show the development of 35S:GL3 patterns for *A. thaliana* conditions over time, where (C) is intermediate and (E) is the final state. D and F show the development of 35S:GL3 patterns for *A. alpina* conditions over time, where D is intermediate and (F) is the final state.

for each considered gene. Although, as in other studies, the distribution of alleles is very asymmetric, the allele frequency lies in a similar range as in other systematic genetic screens [215]. In this context it is noteworthy, that we also found several alleles of fairly small genes such as TRY. This suggests that we have identified a fairly representative set of mutants.

6.5.2 Trichome patterning and differentiation: Differences and similarities between *A. alpina* and *A. thaliana*

In *A. thaliana*, trichome initiation is governed by the WD40 gene AtTTG1, the MYB gene AtGL1 and the redundantly acting bHLH genes AtGL3 and AtEGL3 [8, 197]. Orthologous genes of these regulators are also present in the *A. alpina* genome as judged by sequence similarity and synteny. It was therefore surprising that we found no *gl1* mutant in our screen though this may be due to statistical reasons. Our attempts to create transgenic lines suppressing *AaGL1* expression failed as we could not recover any transgenic plants. All glabrous mutants carried either relevant mutations in *AaTTG1* [73] or in *AaGL3*. The latter is striking, as this indicates that *Aagl3* mutants are completely devoid of trichomes. This is in sharp contrast to *A. thaliana*, where only the additional loss of AtEGL3 results in a glabrous phenotype [136]. Our result indicates that the *AaEGL3* gene in *A. alpina* has no function in trichome patterning. Since *AaEGL3* can efficiently rescue the *A. thaliana Atgl3 Ategl3* mutant when expressed from a heterologous promoter, it is likely that the functional change is due to differences in the regulation of expression. Also the function of *AaGL3* in trichome patterning of *A. alpina* appears to be different as judged by our overexpression data. While overexpression of AtGL3 in *A. thaliana* Columbia causes a higher trichome density, 7 to 30 fold overexpression of *AaGL3* in *A. alpina* leads to a reduction of trichome density. On the first glance this is difficult to understand. One possibility is that in *A. alpina* overexpression of *AaGL3* counteracts intercalating trichome formation. In support of this, we found only similar sized trichomes on older leaves. However, as both trichome size classes depend on *AaGL3* and *AaTTG1*, it is difficult to envision why only one patterning event should be sensitive to *AaGL3* overexpression. In principle it could be due to changes in the gene regulatory network structure or additional regulators. A simpler explanation would explain the difference by changes of the parameters in the *Arabidopsis* network. The latter was suggested to us by the finding that the relative expression levels of GL1 and GL3 differs between the two species. Mathematical modeling revealed that a reduction of GL1 relative to GL3 is sufficient to explain the different responses to GL3 overexpression. It is important to

note that this mathematical explanation of the phenotypes is already possible with a simple model capturing the protein-protein interactions and that this conclusion is therefore also valid for any models extended by additional genes and/or interactions.

Trichome patterning in *A. thaliana* involves also seven redundantly acting inhibitor genes including *AtTRY*, *AtCPC*, *AtETC1*, *AtETC2*, *AtETC3*, *AtTCL1* and *AtTCL2*. Among these, *AtTRY* and *AtCPC* are the most relevant as judged by the single mutant phenotypes, though they show qualitatively different phenotypes. While *Attry* mutants exhibit trichome clusters, *Atcpc* mutants show a higher trichome density. It is not surprising that we did not find *Aacpc* mutants in our screens as wild-type *A. alpina* plants display a very dense trichome pattern and higher densities can easily be missed under screening conditions. We did, however, find two *Aatry* alleles. Both show trichome clusters and larger, more-branched trichomes. The latter is interesting as it indicates that the dual function of the TRY gene in trichome patterning and branching is evolutionary conserved and therefore possibly functionally relevant.

In *A. thaliana*, trichome differentiation is regulated by the AtGL2 gene [198]. This is suggested by the *Atgl2* mutant trichome phenotype: trichomes are smaller, less branched, and in extreme cases, trichomes have a puzzle piece like shape with a little bump suggesting that trichomes are initiated but lost their trichome fate. The *Arabis Aagl2* mutant found in this study differs in that only the small trichomes but not the large ones are affected. As our attempts to verify the phenotype by miRNA suppression or rescue of the *Aagl2* mutant failed because we could not recover any transformants the phenotype needs to be interpretated with caution. Within this limits one possible explanation is that *AaGL2* is not necessary for the proper differentiation of large trichomes implying that small and large trichomes are genetically distinct. An alternative explanation is that increased size can compensate the requirement for *AaGL2*. This scenario is supported by the previous finding that Atgl2 *A. thaliana* mutants can be rescued in genetic situations in which the DNA content and thereby trichome size is increased [185].

6.5.3 Trichome morphogenesis in *A. alpina*

Our phenotypic analysis of trichome morphogenesis revealed that the development in *A. alpina* is very similar to that in *A. thaliana*. Also the range of mutants is similar, as we identified under- and over-branched mutants, mutants with irregular branching, larger and over-branched mutants and several

distorted mutants. Moreover, the molecular analysis revealed that morphogenesis mutants exhibit specific mutations in genes known to be required in *Arabidopsis* for the respective processes. However, several distorted mutants showed no relevant mutations in the tested known DIS-genes. It is therefore possible that mutations in additional, not yet identified genes lead to a distorted phenotype in *A. alpina*.

6.5.4 Perspective

Our forward genetic approach in *A. alpina* has enabled us to recognize the developmental processes and the underlying genes that are similar to that in *A. thaliana*, as well as those that are different. Based on this work it will be possible to unravel the molecular basis of the evolutionary differences of trichome development between *Arabidopsis* and *Arabis*.

6.6 Methods

6.6.1 Plant material and growth conditions

All *A. alpina* mutants were isolated from EMS-mutagenized *A. alpina* Pajares and pep1-1 populations [206]. *A. alpina* seeds on soil were stratified in darkness at 4°C for five days and then placed in growth chambers under long day (LD; 16 h light, 8 h darkness) conditions at 21°C. Pajares required twelve weeks of vernalization to flower whereas *pep1-1* flowered in 80 days without vernalization. For root hair file analysis, seeds were surface sterilized with 70% (v/v) ethanol (5 min) and 2% sodium hypochlorite (w/v, 8 min). Sterilized seeds were sown on 1x Murashige-Skoog plates lacking sucrose and stratified at 4°C for five days. Plants were grown on vertically positioned plates for seven days under long day (LD) conditions (16h light, 8h darkness) at 21°C. Atgl3 Ategl3 double mutants [136] and *A. thaliana* Col-0 were used for inter-species rescue experiments.

6.6.2 Sequence and synteny analysis

A. alpina gene sequences were analysed with the CLC DNA Workbench 5.6.1 (CLC bio, Aarhus, Denmark) by comparison with the coding sequences of the relevant *A. thaliana* genes obtained from TAIR 10 (`www.arabidopsis.org`). NCBI Blastn (2.2.28; `http://blast.ncbi.nlm.nih.gov/Blast.cgi`) [216] along with the assembled *A. alpina* genome was used to confirm the synteny of the selected genes using conserved order and appearance of the neighbouring genes. For

sequence analysis, primers were designed outside the CDS of a given *A. alpina* gene to sequence it in wild type and mutants. NCBI's conserved domain database (CDD; http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) was used to search for conserved domains within the protein sequence [217]. The Net2-Gene server was used to predict splicing sites (http://www.cbs.dtu.dk/services/NetGene2/) [218].

6.6.3 Trichome and root hair analysis

Leaves were analysed when leaf number 6 was approximately 2 cm long, using the TrichEratops software [145]. The distances between trichomes were calculated with R (http://www.r-project.org/). Leaf number three from 10-14 days old *A. alpina* seedlings was dissected and mounted on agarose and analysed as described previously [145]. The position of H-files was determined by their position with respect to the underlying cortex cells on 7 day old plate-grown seedlings. Between 10 and 15 H- and N-file cells were analysed in each root.

6.6.4 Constructs and stable plant transformation

The binary vector pAMPAT-CaMV35S-GW was used to create pAMPAT-CaMV35S:AaGL3 and pAMPAT- CaMV35S:AaEGL3 using the Gateway® system with PCR amplified *AaGL3* and *AaEGL3* coding sequences, respectively (Table S12). The constructs were introduced in the *A. thaliana gl3 egl3* double mutant [136] and in the *Arabis alpina pep1-1* background [206] by Agrobacterium-mediated transformation (strain GV3101-pMP90RK) using floral dip [219]. qPCR analysis was done with established reference primers (Supplemental text [220]).

6.6.5 Mathematical model

The model used to simulate the *A. thaliana* and *A. alpina* phenotypes was based on the activator-inhibitor model as described before [38]. A small adjustment was made to this model by including the saturation of activation by

the activating complex (AC), resulting in the following set of equations:

$$\partial_t [GL1]_j = k_1 + \frac{k_2 k_3 [AC]_j}{k_3 + [AC]_j} - [GL1]_j (1 + [GL3]_j) \tag{6.1}$$

$$\partial_t [GL3]_j = k_4 + k_5 [AC]_j - [GL3]_j (k_6 + [GL1]_j + k_7 [TRY]_j) \tag{6.2}$$

$$\partial_t [TRY]_j = \frac{k_8 k_9 [AC]_j^2}{k_9 + [AC]_j^2} - [TRY]_j (k_{10} + k_7 [GL3]_j) + k_{10} k_{11} \hat{D} [TRY]_j \tag{6.3}$$

$$\partial_t [AC]_j = [GL1]_j [GL3]_j - k_{12} [AC]_j \tag{6.4}$$

These equations were solved for every cell j on a one-dimensional grid of 100 cells total. $\hat{D}$ indicates the diffusion operator between cells. Note that a change introduced to the previously published model is the inclusion of saturating terms for the activation of GL1 and TRY by AC. This is needed to prevent amounts of TRY and GL1 to increase very strongly as a result of the feedback loops. In biology this saturation is realized by a limited number of binding sites for the AC to the promoters of GL1 and TRY. Levels of AC are used as an indicator for whether a cell is considered a trichome or non-trichome cell. If AC levels are higher than half maximum a cell is designated as a trichome.

For the parameter screen we employed a quasi-Monte-Carlo approach and generated 5 million sets of 12 parameters using a scrambled Sobol sequence. All of these sets were then checked for Turing instability [123], i.e. the capability to form a Turing pattern [3]. If a parameter set was capable of patterning, the 35S:AtGL3 phenotype was tested by increasing the GL3 basal production (k_4) 2- to 10-fold. Then we checked whether overexpression of GL3 resulted in an increase of trichome density of at least 20% compared to wild-type, as observed in the *A. thaliana* 35S:GL3 phenotype. Next, the parameter for GL1 basal production (k_1) is reduced by at least 5% and down to at most a 10-fold reduction, where we start by testing the smallest parameter changes first and move logarithmically down to the strongest reduction in k_1. If, after reducing k_1, the network still forms a pattern, the GL3 overexpression is again simulated as described above. If in this case the density is decreased by at least 10%, this is considered to fit the *A. alpina* phenotype. Note that we do not try to exactly reproduce the experimentally observed densities, given that we cannot determine average densities from the data. Furthermore, the aim of the model is not producing a quantitative fit, but rather explore the opposite responses to GL3 overexpression. Therefore, this parameter search

is based on the qualitative difference in behavior upon GL3 overexpression.

Densities are determined by numerical simulation, using randomized initial conditions for each simulation. The initial conditions are given by the steady state of a single-cell model (i.e. $\hat{D} = 0$) plus small inhomogeneous perturbations, sampled from the standard uniform distribution. The density is determined from an average over 100 simulations. The density obtained for the 35S:GL3 cases is analyzed in comparison with corresponding the wild-type density as described above.

6.7 Supplementary Information

The supplementary information for this chapter can be found online at doi.org/10.1073/pnas.1819440116.

Chapter 7

Comparative expression analysis in three Brassicaceae species revealed compensatory changes of the underlying gene regulatory network

Jessica Pietsch*, Anna Deneer*, Christian Fleck and Martin Hülskamp
* Contributed equally

7.1 Introduction

Evolutionary differences and adaptive strategies within plants are driven by the structure and function of the underlying gene regulatory networks (GRNs) ([221–223]). Even minute changes in a GRN can result in striking differences between species ([224, 225]). In evolutionary developmental approaches, such differences are studied in order to gain insight into the genetic basis of phenotypic diversity ([74, 80, 192]). A system that is well-suited for such an approach is trichome patterning in *Arabidopsis thaliana* and other Brassicaceae species ([7, 37, 184]). Genetic analysis of trichome patterning in *A. thaliana* has revealed a complex GRN that controls the regular distribution of trichomes in the leaf epidermis ([9, 34, 169]). Most of the genes found in *A. thaliana* are present in A. alpina and appear to have the same function in regulating trichome patterning ([73]). This suggests that the core of the GRN found in *A. thaliana* might be operating in other Brassicaceae as well. Therefore, the evolutionary analysis of trichome patterning in different Brassicaceae species may enable the identification of subtle changes of the underlying GRN.

In *A. thaliana*, trichomes are initiated in a regular pattern early in leaf development. Genetic analysis identified mutants in which regular pattern formation was disturbed and subsequent molecular analysis revealed the rel-

evant genes ([34]). One group of genes promotes trichome formation and a second group inhibits trichome formation ([35]). The core of the network is a group of three genes, the R2R3MYB protein GLABRA1 (GL1) ([45, 134, 136]), the bHLH protein GLABRA3 (GL3) ([44, 136, 160]) and the WD40 protein TRANSPARANT TESTA GLABRA1 (TTG1) ([43, 162, 163]). The respective proteins form a complex in which GL1 and TTG1 both bind to GL3 ([44, 64]). This so-called MBW (**M**YB, **b**HLH, **W**D40) complex promotes trichome development ([44]). In addition, MYB23 and EGL3 were found to act redundantly with GL1 and GL3, respectively ([134–136]). A second group of genes act as inhibitors of trichome formation. These are all encoded by small R3MYB transcription factors including TRIPTYCHON (TRY) ([50, 52, 56, 139]), CAPRICE (CPC) ([50, 51]), ENHANCER OF TRY and CPC1, 2 and 3 (ETC1, ETC2, ETC3) ([54, 59, 134]). TRY and CPC seem to be the major players as the single mutants exhibit clear phenotypes which is enhanced in combinations with the others suggesting redundant action ([50]). These inhibitors repress the function of the MBW complex by competing with GL1 for binding to GL3/EGL3. The detailed analysis of the function of these genes led to two principles that can explain the generation of trichome spacing patterns without pre-existing information (*de novo* patterning). In short, the first principle is an activator inhibitor model ([5]): the three MBW proteins activate the expression of the inhibitors, that can move within the tissue and repress the MBW function ([38, 64]). The second principle is an activator depletion model ([5]). Here, the activator TTG1 is mobile and captured by GL3 in trichome precursors, which in turn leads to a depletion of TTG1 in the neighbouring cells and thereby inhibition of trichome formation ([40, 141]). It is likely that both principles act in parallel ([183]). Mathematical models have been developed to study the behaviour of these principles in more detail ([38, 40, 42, 183]). These principles can explain how a trichome cell is selected. In the selected trichome cell, the homeobox transcription factor gene GLABRA2 (GL2) is turned on and considered to initiate the differentiation into a trichome cell ([61, 138, 207]).

The systematic forward genetic screen for trichome mutants in *A. alpina* has enabled the identification and functional characterization of trichome patterning genes in this species ([184, 220]). *Arabis alpina* diverged from *A. thaliana* between 26 and 40 million years ago ([194–196]). At this evolutionary distance it was possible to identify the gene orthologs to those in *A. thaliana* by synteny on the chromosomes ([184]). It was therefore possible to unambiguously recognize not only the homologous genes, but also that two of the seven inhibitor genes, TCL1 and ETC2, are missing. The genetic analysis revealed two interesting changes in the GRN. First, the GL3 gene in *A. alpina*

does not appear to act redundantly with EGL3. While in *A. alpina* the *gl3* single mutant is completely devoid of trichomes, it requires the *gl3 egl3* double mutant in *A. thaliana* to express the full phenotype ([184]). Other than that, the structure of the GRN in *A. alpina* seems to the same as in *A. thaliana*. It is, however, noteworthy, that the response of the network to overexpression of GL3 is very different such that this causes the production of more trichomes in *A. thaliana* and less in *A. alpina*. This difference in the behaviour can be explained by different relative expression levels of two key genes, GL1 and TRY, in the two species ([184]). Modelling revealed that this change in the parameters is sufficient to explain the different responses to GL3 overexpression.

In this work, we compared the relative expression levels of trichome patterning genes in *A. thaliana*, *A. alpina* and *C. hirsuta*. The additional species *Cardamine hirsuta* is estimated to have diverged between 13 and 43 million years ago from *A. thaliana* ([226, 227]). For comparison between the species, the expression levels of the trichome patterning genes were considered relative to GL3/EGL3 as all patterning proteins bind to GL3 thereby regulating its activity. We observed striking differences raising the question whether and how the GRN established in *A. thaliana* has adapted to this. We analysed the differences by mathematical modelling and determined which parameters (i.e. interactions and regulations in the GRN) could explain the observed differences in the relative expression levels.

7.2 Methods

7.2.1 Primer establishment and validation

qPCR primers must meet particular requirements. Preferably intron-spanning primers were designed using GenScript Real-time PCR Primer Design (`www.genscript.com`) with an optimal melting temperature of 60±2°C and sequence specific amplicons of ideally 150-200 bp. They exhibit one single band of the expected size in agarose gel electrophoresis and a single peak in the melting curve. Amplification efficiency and correlation were determined based on serial cDNA dilution steps (1:10, 1:20, 1:40, 1:80, 1:160, 1:320). Cq and log10 values of the dilution series were used to calculate the slope Δ by

$$\Delta = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}, \tag{7.1}$$

where N is the number of dilution steps. The slope served to calculate the primer efficiency E by

$$E = 100 \cdot (10^{\frac{-1}{\Delta}} - 1). \tag{7.2}$$

The R^2 correlation of the Cq and the log10 values was calculated using

$$\rho_{x,y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \sigma_y}. \tag{7.3}$$

Amplification efficiencies of 100% ± 20 for genes of interest and 100% ±10 for reference genes as well as a linear standard curve with a correlation of ≥ 0.99 were accepted. Sequences were taken from TAIR (www.arabidopsis.org), from Genomic resources for *Arabis alpina* (`www.arabis-alpina.org`) and from *Cardamine hirsuta* genetic and genomic resource (`http://chi.mpipz.mpg.de`).

7.2.2 Sample preparation

Cotyledons as well as juvenile leaves (leaf one and two for *Arabidopsis thaliana* and *A. alpina*, additionally leaf three for *C. hirsuta*) of plant seedlings were removed to gather 200-400 μm sized leaves with on-going trichome patterning machinery. Material of up to 45 plants was collected per biological replicate, frozen in liquid nitrogen and stored at −80°C until further processing. RNA extraction was performed using the Tri-Reagent method including DNaseI treatment and quality control was ensured via bleach gel and photometry. cDNA synthesis was carried out according to the manufacturer's protocol (RevertAid First Strand cDNA Synthesis Kit; Thermo Fisher Scientific) using 1.5 μg RNA per sample because pre-tests had revealed that 1 μL undiluted cDNA based on 1 μg RNA were required to obtain Cq values < 30. qPCR protocols were standardized using three biological as well as three technical replicates, master mixes and always both reference genes on each plate.

7.2.3 Analysis qPCR data

A two-sided Grubbs test ($\alpha = 0.05$) was performed to identify outliers. Normalization of the data was conducted according to the geNorm manual ([228]), describing gene expressions relatively to each other. Special considerations are given to normalization factors and the individual primer efficiencies ϵ. Thereby not a generalized gene duplication per cycle (1+1) is assumed, but the individual amplification rate (1+ϵ) is used for further calculations. The expression data of each species was normalized by two different reference genes. Using the variability of the reference genes and not the Cq values,

allows interspecies comparison even with different reference genes for each organism.

7.2.4 Compiling GL1 synteny

Arabidopsis thaliana was used as reference to elaborate the synteny of GL1 comparing it with *A. alpina* and *C. hirsuta*. The AtGL1 sequence was used to perform a BLAST search against the *C. hirsuta* CDS database (`http://bioinfo.mpipz.mpg.de/blast/cgi-bin/public_blast_cs.cgi`). More than a dozen of loci spanning the first three highest ranked genes were blasted against *Arabidopsis thaliana*. The AaGL1 ortholog as well as its adjoining genes were identified using the 1x1 orthologs table from the *Arabis alpina* website (`http://www.arabis-alpina.org/data/ArabisAlpina/data/Aa_At_ortho_1x1.txt`).

7.2.5 Mathematical modelling

The model consists of 8 components, which are modelled in the form of a system of coupled ordinary differential equations. These components include the proteins TTG1, GL1, GL3, TRY, CPC and ETC. Note that the species designated by GL1 and ETC are assumed to be the combined behavior of GL1, MYB23 and ETC1, ETC2 and ETC3, respectively. Additionally, the complex formation between GL3 and TTG1 and GL3 and GL1 is explicitly modelled, whereas the binding between GL3 and the inhibitors TRY, CPC and ETC is implicitly modelled since these do not feed back into the system. This model consists of 31 parameters that describe processes such as degradation, binding, activation and transport. This model is based on previously published versions and is extended by the inclusion of ETC ([38, 40, 183, 184]). The system of equations is

$$\partial_t [TTG1]_j = \theta_1 - [TTG1]_j(\theta_2 + \theta_3[GL3]_j) + \theta_2\theta_4\hat{L}[TTG1]_j \tag{7.4}$$

$$\partial_t [GL1]_j = \theta_5 + \theta_6[AC2]_j - [GL1]_j(\theta_7 + \theta_8[GL3]_j) + \theta_7\theta_{30}\hat{L}[GL1]_j \tag{7.5}$$

$$\begin{aligned}\partial_t [GL3]_j = \theta_9 + \frac{\theta_{10}\theta_{11}[AC1]_j^2}{\theta_{11} + [AC1]_j^2} + \frac{\theta_{12}\theta_{13}[AC2]_j^2}{\theta_{13} + [AC2]_j^2} - \\ [GL3]_j(\theta_{14} + \theta_3[TTG1]_j + \theta_8[GL1]_j + \theta_{15}[TRY]_j + \end{aligned} \tag{7.6}$$

$$\theta_{16}[CPC]_j + \theta_{17}[ETC]_j) + \theta_{14}\theta_{31}\hat{L}[GL3]_j \tag{7.7}$$

$$\begin{aligned}\partial_t [TRY]_j = \theta_{18}[AC1]_j^2 - [TRY]_j(\theta_{19} + \theta_{15}[GL3]_j) + \\ \theta_{19}\theta_{20}\hat{L}[TRY]_j\end{aligned} \tag{7.8}$$

$$\partial_t[CPC]_j = \theta_{21}[AC2]_j^2 - [CPC]_j(\theta_{22} + \theta_{16}[GL3]_j) + \theta_{22}\theta_{23}\hat{L}[CPC]_j \tag{7.9}$$

$$\partial_t[ETC]_j = \theta_{24}[AC1]_j^2 + \theta_{25}[AC2]_j^2 - [ETC]_j(\theta_{26} - \theta_{17}[GL3]_j) + \theta_{26}\theta_{27}\hat{L}[ETC]_j \tag{7.10}$$

$$\partial_t[AC1]_j = \theta_3[GL3]_j[TTG1]_j - \theta_{28}[AC1]_j \tag{7.11}$$

$$\partial_t[AC2]_j = \theta_8[GL3]_j[GL1]_j - \theta_{29}[AC2]_j \tag{7.12}$$

where $\hat{L}$ indicates the coupling equation between cells, given by

$$\hat{L}[\chi]_{x,y} = [\chi]_{y-1,x} + [\chi]_{y+1,x} + [\chi]_{y,x-1} + [\chi]_{y,x+1} + [\chi]_{y+1,x-1} + [\chi]_{y-1,x+1} - 6[\chi]_{y,x}. \tag{7.13}$$

for any species χ and cell at coordinates (x, y). Patterns were simulated on a grid of 20-by-20 cells with hexagonal connectivity on a domain with zero-flux boundary conditions. The initial conditions are given by the steady state of a single-cell model (i.e. $\hat{L} = 0$) plus small inhomogeneous perturbations, sampled from the standard uniform distribution. The trichomes on the grid are identified by cells that have relatively high amounts of active complex (AC1 + AC2), specifically, cells that have more than the half-maximum of total AC are designated as trichomes. The cluster density of trichomes was averaged over 10 simulations, each with randomized initial conditions. Parameter sets that produced less than 10% clusters are used for further analysis.

Given that we are only interested in parameter sets that form patterns, we apply linear stability anaylsis to identify these sets ([123]). In the domain of interest, a diffusion-driven instability (Turing instability) occurs ([3]), resulting in an inhomogeneous patterning state. In linear stability analysis, the stability of a uniform steady state is verified by determining whether effects of small perturbations to the ODE system decay over time. Turing instability was tested by the following criteria: starting from a uniform steady state (i) the steady state in the absence of diffusion is stable and (ii) the steady state in the presence of diffusion is unstable ([123]). For criterion i this means that all eigenvalues of the Jacobian of the system in (7.4) - (7.12) evaluated at steady state must be negative. To perform the same test for criterion ii we decoupled the system by Fourier transformation and analysed the eigenvalues ([38, 40, 183]), where the real part of at least one of the eigenvalues must be positive.

7.2.6 Parameter estimation

The parameters of the model are estimated through an optimization routine where the qPCR data is used in a cost function. The goal is to arrive at a distribution of values for these parameters for each of the species. Note that this problem suffers from non-identifiability ([178]), i.e., no unique value or bounded confidence interval can be determined for the parameters; for this, additional data would be required that is simply not available. Nonetheless, through a multi-start optimization routine that ensures multiple optimal solutions, it is possible to deal with the uncertainty in the system and arrive at predictions about possible genetic adaptations on a regulatory level that differentiates the three species from each other.

The analysis used here requires solving a constrained multivariable minimization problem ([229]). Specifically, we aim to find the minimum of the problem specified by

$$\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) \text{ such that } \begin{cases} c(\boldsymbol{\theta}) \geq 0 \\ \text{lb} \leq \boldsymbol{\theta} \leq \text{ub} \end{cases} \tag{7.14}$$

where $c(\boldsymbol{\theta})$ is a non-linear constraint function, $f(\boldsymbol{\theta})$ is a scalar cost-function and lb and ub are the lower- and upper-bounds of the parameter vector $\boldsymbol{\theta}$. The cost function f is a normalized sum-of-squares given by

$$f(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{(\bar{y}_i(\boldsymbol{\theta}) - y_i)^2}{y_i^2} \tag{7.15}$$

where $\bar{y}_i(\boldsymbol{\theta})$ is the expression level of the i-th gene out of N total genes predicted by the model and y_i is the corresponding datapoint. Given that the model simulates the concentration in a tissue, $\bar{y}_i(\boldsymbol{\theta})$ is the average of gene i across the tissue, relative to the sum of GL3 and EGL3, similar as the data.

The constraint function $c(\boldsymbol{\theta})$ is chosen such that the parameters $\boldsymbol{\theta}$ fall in the Turing Space, i.e. are capable of patterning. To achieve this, we make use of linear stability analysis as described above and determine the eigenvalues of the Jacobian of the system of equations. $c(\boldsymbol{\theta})$ is given by

$$c(\boldsymbol{\theta}) = \text{Re}(\lambda_{\text{max}}) \tag{7.16}$$

where λ_{max} is the largest eigenvalue of the Jacobian. By determining whether the real part of the largest eigenvalue is positive (i.e. $c(\boldsymbol{\theta}) \geq 0$), we learn that

the parameter set $\boldsymbol{\theta}$ can form a pattern, which constrains the allowable range of parameters. Note that this range is also constrained by the choice of bounds (lb and ub) of the optimization problem. In this case, we set the interval for the each of the parameters in $\boldsymbol{\theta}$ to $[0.01, 100]$, to allow a range of multiple orders of magnitude. One further constraint, which is applied in post-processing, is that the pattern produced by $\boldsymbol{\theta}$ must not show any clusters of trichomes, as is the case in all the patterns formed by the three species. As such, we limit the range of parameter values to those that simulate a realistic pattern and produce the best possible fit to the data according to $f(\boldsymbol{\theta})$. Finally, the optimization problem is started from multiple, randomly generated initial points. This set of initial points is generated by a Sobol sequence to ensure a good coverage of the parameter space ([90]). Starting from these randomly generated initial points, the optimization converges to a local minimum that satisfies the constraints, leading to a distribution of optimal parameter sets $\hat{\boldsymbol{\theta}}$ that correspond to the local minima. This procedure is followed until 100 optimized parameter sets are obtained for each of the species.

These distributions are then used in a statistical analysis to determine which of the parameter distributions are statistically different between the species. Towards this end, we use the two-sample Kolmogorov-Smirnov (KS) goodness-of-fit hypothesis test to determine if two empirical distributions are drawn from the same (unknown) underlying population cumulative distribution functions ([230]).

7.2.7 Sensitivity analysis

The sensitivity of the trichome density to the each of the individual parameters is determined, using a variation on the elementary effects (EE) test ([119, 231]). The EE is a one-at-a-time screening method, i.e., only one parameter is varied at a time and the variation in the output is measured ([119, 231]). For a model with N parameters, each parameter θ_i, $i = 1,\dots N$, is assumed to vary across p selected levels in the parameter space. The region of experimentation Ω is an N-dimensional p-level grid. In standard sensitivity practices, parameters are assumed to be uniformly distributed in $[0, 1]$ and then transformed from the unit hypercube to their known distributions ([119]). In this case, we adapt this region to ensure that the parameters fall within the Turing Space. The lower limit of Ω is by default chosen to be 10^{-1} and the upper limit 10, where every point Δ in the grid is the perturbation applied to θ_i for which the EE_i is determined. In the case that either limit would shift $\boldsymbol{\theta}$ outside of the Turing space, then the lower limit is adjusted to the smallest

value between $[10^{-1}, 1]$ that according to linear stability analysis falls within the Turing space, and the upper limit is the largest value between $[1, 10]$ that falls within the Turing space. This means that the p-level grid in Ω can have different upper and lower limits, depending on the allowable range according to linear stability analysis, but always consists of the same number of grid-points. Furthermore, these grid-points are chosen such that they are logarithmically spaced.

For the trichome patterning model, we have $N = 31$ and choose $p = 10$. We perform the EE sensitivity analysis for the top 10 best-fitting parameter sets $\hat{\boldsymbol{\theta}}$ resulting from the optimization routine. For a given set $\hat{\boldsymbol{\theta}}^k$, $k = 1,\dots,10$ the EE of the i-th parameter is defined as:

$$\mathrm{EE}_i(\hat{\boldsymbol{\theta}}^k) = \frac{Y(\theta_1,\dots,\theta_{i-1},\theta_i,\theta_{i+1},\dots,\theta_N) - Y(\hat{\boldsymbol{\theta}}^k)}{\theta_i \cdot \Delta - \theta_i} \tag{7.17}$$

where Δ is a value in the p-level grid with the limits chosen as described above. Then, the absolute values of the EE_i, computed at p different grid points, are averaged to get

$$\bar{\mathrm{EE}}_i = \frac{1}{p}\sum_{j=1}^{p} |\mathrm{EE}_i^j|. \tag{7.18}$$

Finally, we average over all $\bar{\mathrm{EE}}_i$ for every $\hat{\boldsymbol{\theta}}^k$.

7.3 Results

7.3.1 Comparison of trichome gene expression in different species

All three species considered here, *A. thaliana*, *A. alpina* and *C. hirsuta* produce regularly spaced trichomes on leaves ([32, 184]). The trichome density differs such that *C. hirsuta* has a lower density, and *A. alpina* a higher density as compared to *A. thaliana*. A meaningful quantitative comparison of trichome density appears not to be possible as leaf sizes, growth dynamics and the juvenile-to-adult transition differ making it arbitrary to choose the proper mature leaves for comparison. We therefore focused on qualitative and ratiometric comparisons in this study.

A direct comparison of the expression levels of trichome genes between species by qPCR is not possible for various reasons. In particular, the primers

are different for the same genes and normalization was done with difference reference genes. We therefore compared the expression levels between species by normalizing the expression to the bHLH genes. The bHLH protein is the central component of the MBW complex to which the activators TTG1 and R2R3 MYB proteins bind ([44]) and on which the R3 MYB negative regulators exert their repressive effect by competitive binding with the R2R3 MYBs ([38]). It is conceivable that the outcome of this GRN depends on the concentrations of the other patterning proteins relative to the bHLH. We therefore considered the bHLH expression levels a good reference to judge and compare to the relative changes of all other patterning genes. We combined GL3 and EGL3 for the comparison between the species because the two genes act redundantly in *A. thaliana* and have similar molecular roles in trichome patterning ([61, 69, 136]).

Another aspect to enable a comparison between the species is the choice of plant material. For our qPCR experiments we used young leaves at developmental stages in which trichome patterning was still ongoing as recognized by the presence of incipient and young stages of trichome development at the base of leaves. These stages could be unambiguously identified in all three species.

7.3.2 Relative trichome gene expression differs in *A. thaliana*, *A. alpina* and *C. hirsuta*

In a first step, we identified the bonafide orthologs of the *Arabidopsis* trichome patterning genes in *A. alpina* and *C. hirsuta* by sequence similarity and synteny (Suppl. Figure S1). Primers were designed to meet the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines ([232]).

Plants were grown on soil and young leaves were harvested at stages at which incipient developing trichomes were seen. The first two leaves and the hypocotyl were removed. Quantitative Real-Time PCR experiments were done with three biological replicas and normalized to a set of species-specific reference genes. To enable a comparison between species we normalized all expressions with GL3/EGL3. Figure 2.1 shows the relative expression levels of patterning genes normalized to the combined transcript levels of GL3 and EGL3 which was set to one (see also Table S1). In *A. thaliana*, GL3/EGL3 appear to be the limiting factors among the other trichome activating genes (Figure 7.1). TTG1 expression is about14 fold higher and the expression of GL1

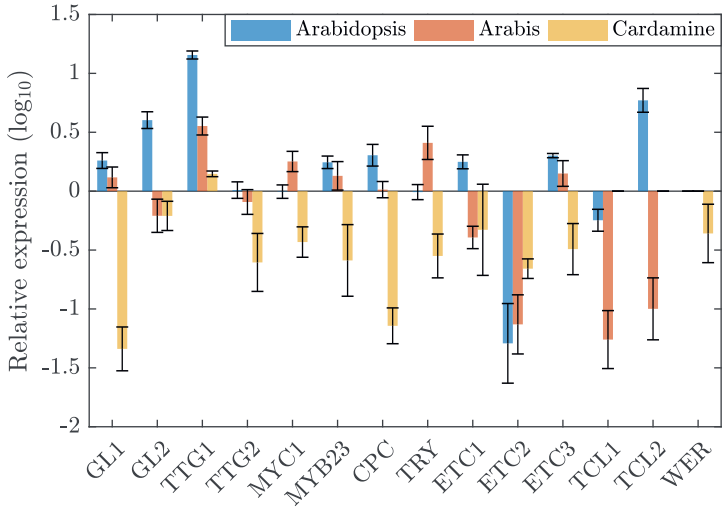


Figure 7.1: **Comparative patterning gene expression.** Depicted are the expressions and fold changes of 15 patterning genes in *A. thaliana* (blue), *A. alpina* (red), and *C. hirsuta* (yellow) relative to the sum of GL3 and EGL3 in the respective species.

and MYB23 both show an about 2-fold higher expression. Consistent with the previous finding that GL1 and MYB23 act redundantly in *Arabidopsis* ([135]), they show a similar expression level and we combined their expression for the modelling approach to reduce the complexity (see below). The relative expression levels of the inhibitors were different with CPC, ETC1, ETC3 and TCL2 being higher and ETC2 and TCL1 lower than GL3/EGL3.

The expression profile in *A. alpina* is fairly similar to that in *A. thaliana*. In *C. hirsuta* we found a strikingly different pattern of the relative expression of trichome patterning genes. Here, most of the patterning genes exhibit lower expressions as compared to GL3/EGL3. In particular, the expression of GL1 and CPC were drastically lower as compared to the other two species.

7.3.3 GL1, MYB23 and WER expression differs in *Arabidopsis thaliana*, *A. alpina* and *C. hirsuta*

The very low relative and also absolute expression levels of GL1 in *C. hirsuta* raised the question whether the function of GL1 is redundantly provided by MYB23 ([135]) or even WER ([233]). To study this in more detail, we com-

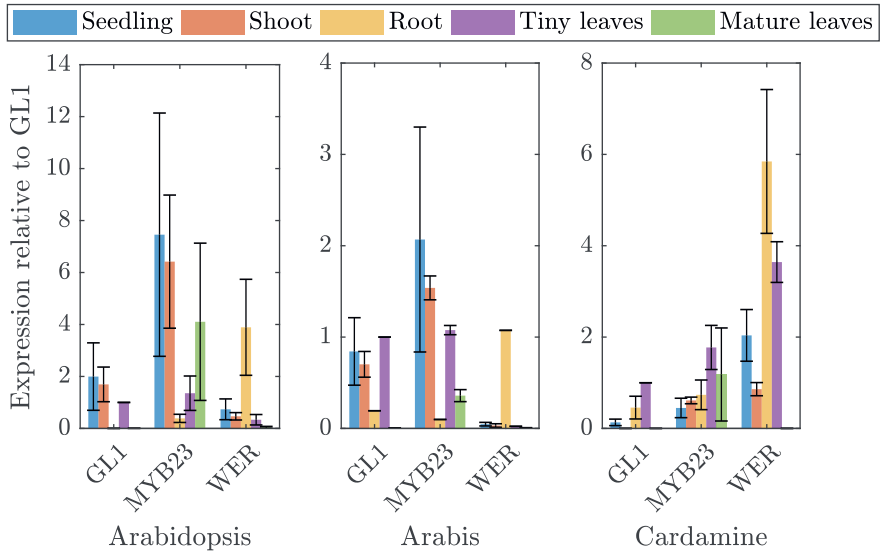


Figure 7.2: **Quantitative expression analysis of three MYB homologs in different species and tissues.** Depicted is the expression of GL1, MYB23, and WER in *Arabidopsis thaliana*, *A. alpina*, and *C. hirsuta* in seedlings (blue), shoots (red), roots (yellow), tiny leaves (purple), and mature leaves (green), relative to GL1 expression in tiny leaves of the respective species.

pared the expression of the three genes in five different tissues in the three species (Figure 7.2). To facilitate a comparison in the context of trichome patterning, we normalized the expression levels with respect to the GL1 expression in young leaves. As expected, GL1 and MYB23 are expressed in most aerial tissues in *Arabidopsis thaliana* but not in the root whereas WER expression was detected strongly in the root. In *A. alpina*, GL1 and MYB23 expression levels are similarly high in seedlings, shoot, young leaves and roots. WER expression was root specific. In *C. hirsuta*, GL1 and MYB23 expression was absent or low in all tissues. Surprisingly, WER expression was not only high in roots, but also in young leaves. Here, WER expression was 2.6 fold higher than that of GL1. These findings suggest that the tissue specific functions of GL1, MYB23 and WER might be different in the three species. Given that it GL1 and WER proteins have equivalent function during trichome initiation in *A. thaliana* ([134]) it is conceivable that the higher expression of WER can substitute the low levels of GL1.

7.3.4 Modelling to predict the molecular adaptations to relative expression differences between the three species

The functional comparison of the regulation between *A. thaliana* and *A. alpina* suggests that the core of the underlying regulatory network of trichome patterning is conserved ([184]). Consistent with this, all relevant orthologs of the relevant *A. thaliana* genes are also found in *C. hirsuta*. The qPCR data show striking differences in the relative expression levels. Given that trichome initiation is driven by the activity of the MBW complex, in which the components undergo competitive binding, it is surprising that the GRN can tolerate such differences ([38, 64]). We used mathematical modelling to explore whether the *Arabidopsis*-based GRN is capable of coping with such differing relative expression patterns. And if so, which parameters can compensate for changes in the relative expression levels and to which of these is the pattern most sensitive? Towards this end, we developed a model based on previous versions ([38, 40, 183]) and consisting of TTG1, TRY, CPC, ETC1/ETC2/ETC3, GL3/EGL3 and GL1/MYB23/WER (Figure 7.3A).

We define two criteria for the model to fulfil. First, the parameter set has to reproduce the same relative differences as found for the patterning genes. Second, it has to simulate realistic trichome patterns, i.e., patterns must not show any clustering of trichomes ([32]). We varied all 31 parameters using a non-linear optimization routine such that the model most accurately reproduces the expression data (Figure 3). This is surprisingly well possible for the expression data sets of all three species with many different parameter sets.

The identification of a large number of parameter combinations for each species enabled us to compare the distributions of the parameters between the three species in search for striking differences. Towards this end, we used a Kolmogorov-Smirnov test to identify parameter distributions that are significantly different between the species. Out of the 31 parameters only 13 fulfilled this criterion and were considered parameters that are relevant for compensating different expression ratios in all three species. The distributions are shown in Figure 4. The 13 parameters have significantly different distributions for at least two of the species, indicating that the *Arabidopsis*-based model can cope with different relative expression levels by compensatory changes of different parameters. For the other 18 parameters we did not find a significant difference (Suppl. Figure S2).

Following from this approach are three distributions (one for every species) for each of the 31 parameters. To identify the most relevant differences, we

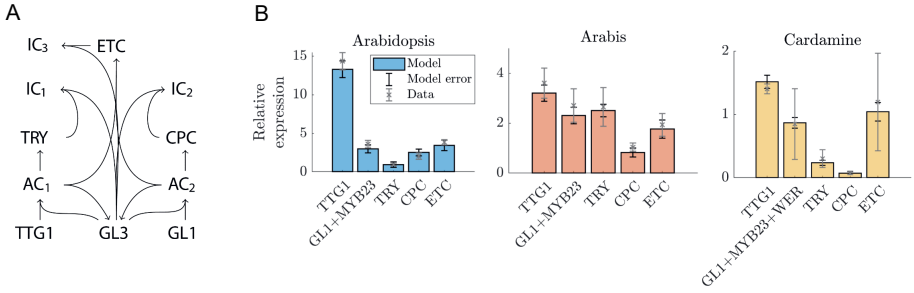


Figure 7.3: **Expression levels in the model compared to qPCR data.** A: Schematic representation of model network. AC1 and AC2 are the active complexes TTG1 GL3 and GL1 GL3; IC1 and IC2 are the inactive complexes TRY GL3 and CPC GL3. B: Expression levels of genes in *A. thaliana*, *A. alpina*, and *C. hirsuta* relative to GL3+EGL3. The qPCR data is indicated by grey crosses with error bars (representing biological replicates) and the model expression levels are indicated by the bars (blue for *A. thaliana*, red for *A. alpina*, yellow for *C. hirsuta*), the model error (black error bars) is the error over the 10 best fitting parameter sets.

focus only on a subset of 13 parameters (Figure 7.4). These 13 parameters have significantly different distributions for at least two of the species as determined by the Kolmogorov-Smirnov test ([230]). For the remaining 18 parameters we did not find a significant difference (Suppl. Figure S2).

This analysis revealed three interesting differences. First, all parameters regulating the activity of TTG1 in *Arabidopsis* differ from the distributions of *Arabis* and *Cardamine* (except for the diffusion rate of TTG1). It is conceivable that this is due to the relatively high expression of TTG in *Arabidopsis* that could be compensated by parameters changes reducing its activity such as the basal production rate and degradation rate. Second, the parameters regulating TRY activity differ between *Arabis* and the other two species. The model predicts a higher activation rate of TRY by the GL3 TTG1 complex in *Arabis* that would explain the relatively high levels of TRY in this species. Third, the regulation of CPC in *Cardamine* differs from the other two species. We found significant differences for the activation of CPC by the GL3 Gl1 complex and its degradation rate. Both would compensate for the relative low amount of CPC in Cardamine. These three cases exemplify the versatility of the trichome patterning network and show how, despite the varying underlying differences between regulatory mechanisms, the same core network is capable of robustly producing a realistic trichome pattern in all three species.

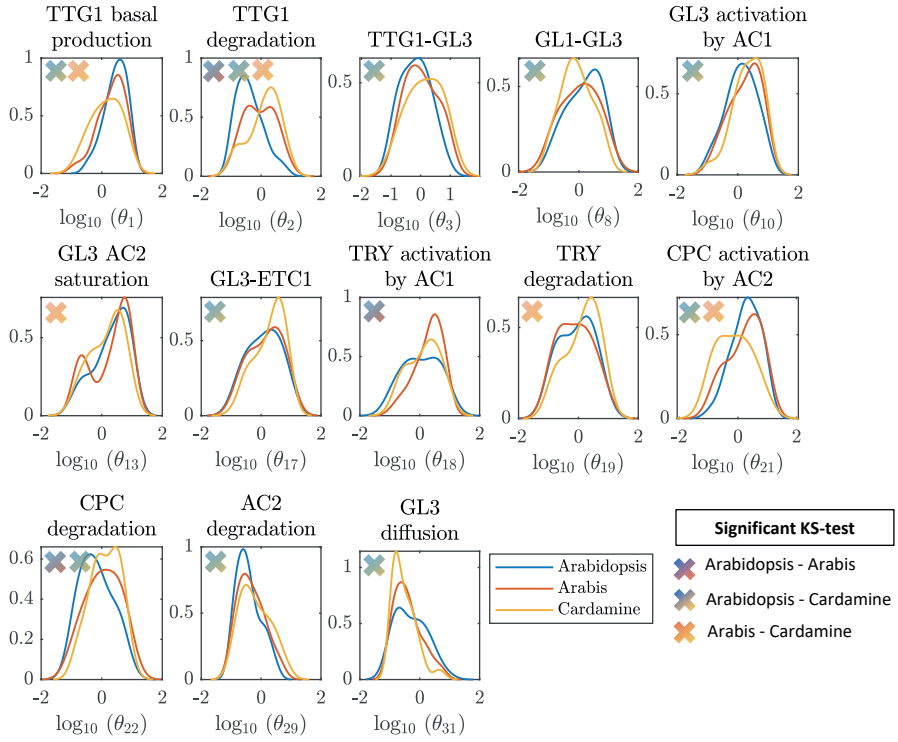


Figure 7.4: **Parameter profile densities.** Parameter distributions that differed between the species according to a Kolmogorov-Smirnov test, obtained from fitting the model output to the qPCR data. The crosses indicate between which pair of species the distributions were found to statistically differ. The titles indicate the biological interpretation behind the parameter θ_i on the x-axis.

The comparison of parameter distributions provides insight into the adaptability of the network to the different expression levels but does not immediately provide information on the effects on trichome patterning. This is possible by determining the sensitivity of trichome density to changes in each of the parameters in all three species (Figure 7.5). This allows predicitions on which parameter is most influential in patterning and whether this varies between species.

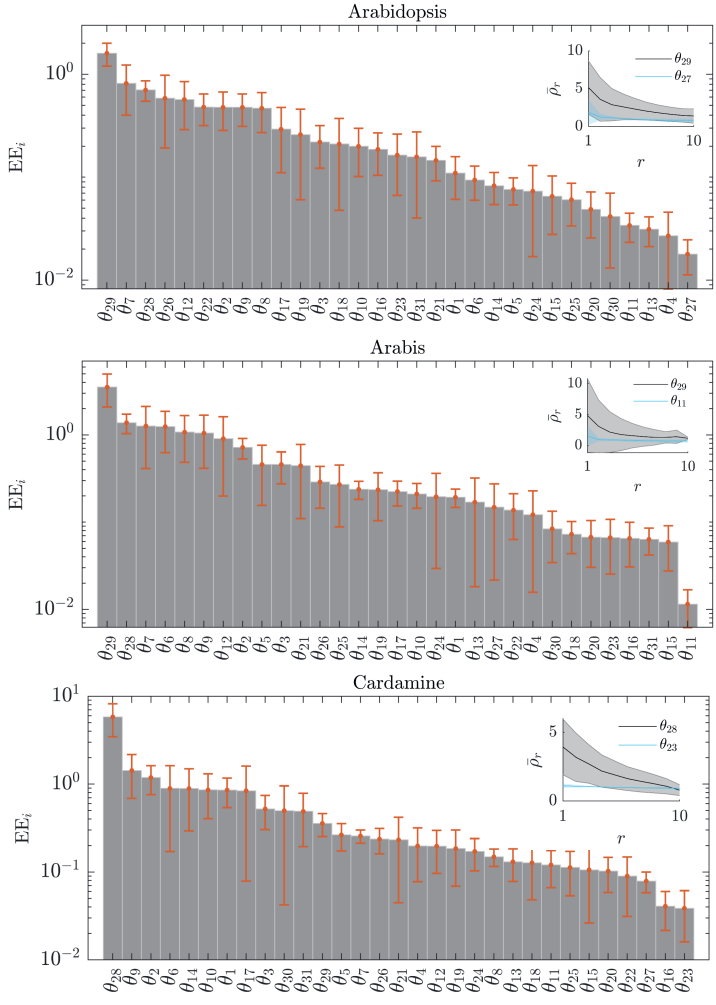


Figure 7.5: **Sensitivity of parameters to trichome density.** The elementary effects (EEs) of each of the parameters θ_i in the model indicates the sensitivity of the trichome density to changes in θ_i, sorted by the respective EE value. The error bars indicate the standard deviation in the EE for the ten best-fitting parameter sets. The inset shows the average and the spread (shaded region) of the trichome density ($\bar{\rho}_r$) for the ten best-fitting parameter sets at each of the ten grid points r of the EE test, for the most sensitive parameter (grey) and the least sensitive parameter (blue).

Our sensitivity analysis predicts that for all three species the stability of the TTG1 GL3 complexes (θ_{28}) is one of the most sensitive parameters. In *A. thaliana* and *A. alpina* the stability of the GL1 GL3 complex (θ_{29}) and the degradation rate of GL1 (θ_7) are among the most sensitive parameters. In *C. hirsuta*, the basal production of GL3/EGL3 is relevant (θ_9) and the degradation of TTG1 (θ_2). Taken together, the sensitivity analysis predicts that the amount of the active complexes most strongly influences the trichome density and that this is a common feature in all three species. *C. hirsuta* differs in that the role of GL3-TTG1 is more relevant than that of GL3-GL1. For an overview of the biological interpretation of all other parameters in Figure 2.5 see Suppl. Table 2.

7.4 Discussion

In this study, we have compared the expression levels trichome patterning genes in the three closely related Brassicaceae species *A. thaliana*, *A. alpina* and *C. hirsuta*. We aimed to use the variation of the relative expression levels to understand the potential of the GRN underlying trichome patterning. For our mathematical modelling approach, we used a complex model that considers the genetic interactions and simulates concentrations on the protein level to consider differential complex formations ([64]). The details of transcription and translation are not explicitly modelled as this would add another layer of complexity and thereby more parameters without gaining an extra value in the absence of additional data. This approach enabled us to evaluate the parameter changes with respect to many different aspects of the patterning process including transcriptional regulation, differential complex formation and stability of the transcript/protein. The possible downside is that we have to consider a 31-dimensional parameter space making it necessary to use statistical approaches to monitor the effect of different parameters and their combinations.

What did we learn? First of all, the structure of the GRN network established in *A. thaliana* is sufficient to generate a trichome pattern even if the relative expression levels show an order of magnitude difference. Second, the GRN compensates for differences in the relative expression patterns by changing other parameters. Not all, but only a subset of the parameters is important for this. Third, one type of parameter – the stability of the MBW complexes – is among the most important in all three species. These predictions might be instrumental for future experiments as they help to focus on aspects of the

GRN network that have not been studied so far.

A second unexpected result followed from our analysis of the relative transcript levels of the three R2R3MYBs GL1, MYB23 and WER. The three *Arabidopsis thaliana* genes cluster together in one MYB subgroup and are characterised by a unique amino acid motif in the eponymous MYB domain ([234]) and are likely to be the result of gene duplications. GL1 and MYB23 act redundantly in the regulation of trichome patterning but have distinct functions in the regulation of trichome branching ([134]). Both are only important for trichome formation but not involved in root hair patterning, which is specifically regulated by WER. This trait specificity is due to differences in the transcriptional regulation as WER and GL1 are equivalent proteins ([235]). Also, overexpression of MYB23 can rescue the *wer* mutant phenotype indicating that the protein can substitute WER in this context ([236]). In *Arabidopsis* we found that GL1 and MYB23 are most strongly expressed in young leaves. WER expression was detectable but very low. When considering that the expression of the three genes is important for their function and that the proteins are functionally similar, it is conceivable that WER might have a function in trichome development in *Cardamine*. In fact, by this argument, WER would be most important in root and leaf patterning as it is also most prominently expressed in roots. We therefore hypothesize that our findings reflect the evolutionary sub-functionalization of the three homologous MYB genes in trichome and root hair regulation. Functional assays, ideally including mutant analysis in *Cardamine* will be required to test this hypothesis.

Our comparison of the relative transcript levels of the three R2R3MYBs GL1, MYB23 and WER revealed surprising differences. While the two genetically relevant genes for trichome pattern GL1 and MYB23 are expressed most strongly in young leaves, both are clearly expressed at a lower level as the WER gene. Given that the expression of these three genes is important for their function and that the proteins are functionally similar ([134]), this suggests that WER might have a function in trichome development in *Cardamine*. In fact, by this argument, WER would be most important in root and leaf patterning as it is also most prominently expressed in roots ([8, 233]). We therefore hypothesize that our findings reflect the evolutionary sub-functionalization of the three homologous MYB genes in trichome and root hair regulation. Functional assays, ideally including mutant analysis in *Cardamine* will be required to test this hypothesis.

7.5 Supplementary information

7.5.1 Figures

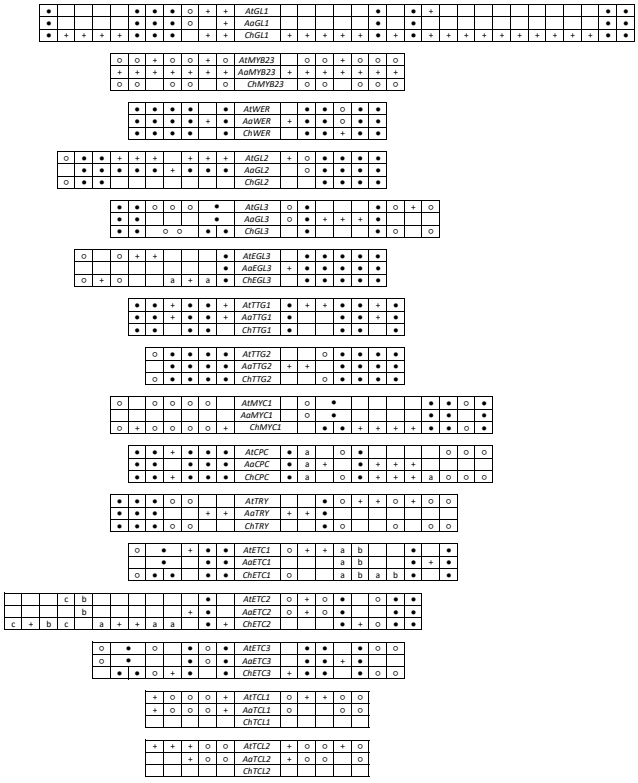


Figure 7.6: **Synteny of all patterning genes in *A. thaliana*, *A. alpina*, and *C. hirsuta*.** Because the three species are closely related, it is possible to identify ortholog genes based on the arrangement of neighboring genes. Each column represents one gene locus. Filled circles display gene orthologs in all three species, empty circles in two species. Pluses indicate additional genes and same letters indicate same genes. Merged cells mean that two or more loci correspond to one gene.

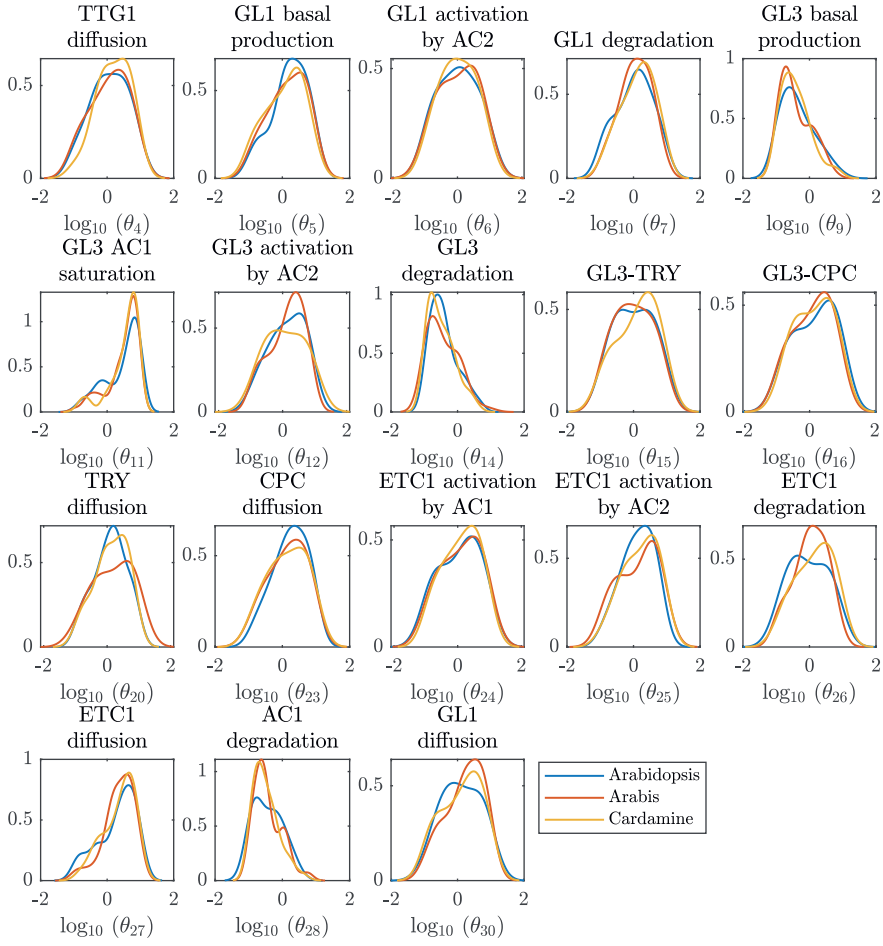


Figure 7.7: **Parameter profile densities.** Parameter distributions for which no significant difference between the species according to a Kolmogorov-Smirnov test, obtained from fitting the model output to the qPCR data.

7.5.2 Tables

Table 7.1: **Summary of all Cq values genes** Listed are all Cq values of the reference genes and investigated patterning genes in *A. thaliana*, *A. alpina*, and *C. hirsuta*

	Replicate	Reference genes		Genes of interest																	
Arabidopsis thaliana		*18S rRNA*	*EF1α*	*GL1*	*GL2*	*GL3*	*EGL3*	*TTG1*	*TTG2*	*MYC1*	*MYB23*	*CPC*	*TRY*	*ETC1*	*ETC2*	*ETC3*	*TCL1*	*TCL2*
	1.1	9.594	15.664	22.280	21.307	24.992	23.868	19.534	23.106	23.438	22.184	22.615	23.657	22.800	28.174	22.920	24.199	21.068
	1.2	9.285	15.435	22.379	21.351	25.124	23.845	19.321	23.400	23.369	22.452	22.795	24.051	22.762	31.964	22.855	23.949	21.099
	1.3	9.475	15.341	22.244	21.708	25.222	23.827	19.347	23.113	23.539	22.697	22.619	23.960	22.815	29.763	22.877	24.209	21.090
	2.1	9.127	15.122	22.642	21.451	25.139	23.945	19.562	23.271	23.720	22.851	22.804	24.056	23.122	27.391	23.424	24.645	21.617
	2.2	9.096	15.287	22.593	21.283	24.887	24.072	19.563	23.755	23.518	22.772	22.863	24.054	23.058	28.200	23.282	24.793	21.532
	2.3	8.959	15.455	22.844	21.405	25.172	24.208	19.372	23.177	23.455	22.836	23.097	24.247	23.041	27.134	23.342	24.786	21.553
	3.1	9.208	15.610	22.145	21.089	24.992	23.599	19.245	23.249	23.393	22.735	22.661	23.727	22.749	28.028	22.943	24.365	21.260
	3.2	9.240	15.562	22.025	21.348	24.891	23.604	18.865	23.027	23.026	22.220	22.694	23.403	22.646	28.655	22.563	24.305	21.164
	3.3	9.340	15.323	22.341	21.165	24.917	23.673	19.075	23.128	23.272	22.442	22.593	24.256	22.531	27.978	22.579	24.247	21.178
Arabis alpina		*RAN3*	*PP2A*	*GL1*	*GL2*	*GL3*	*EGL3*	*TTG1*	*TTG2*	*MYC1*	*MYB23*	*CPC*	*TRY*	*ETC1*	*ETC2*	*ETC3*	*TCL1*	*TCL2*
	1.1	15.410	19.341	19.681	20.210	19.933	21.844	18.264	20.260	19.160	19.374	19.809	17.913	21.410	23.799	18.967	23.821	23.112
	1.2	15.478	19.579	19.486	20.527	20.411	21.847	18.322	20.378	19.391	19.707	19.691	18.277	21.275	23.856	18.798	23.876	23.359
	1.3	15.342	19.311	19.666	20.851	20.236	22.448	18.327	20.394	19.513	19.507	19.860	18.241	21.582	24.178	19.428	24.058	23.229
	2.1	15.374	19.185	19.860	21.214	20.346	22.307	18.075	20.406	19.107	19.673	19.986	18.666	21.488	24.761	19.531	24.833	24.416
	2.2	15.160	19.229	19.833	21.144	20.431	22.323	17.702	20.458	19.198	19.935	19.840	18.835	21.555	24.883	19.340	24.902	24.476
	2.3	14.754	19.143	19.664	20.995	20.380	22.337	18.340	20.416	19.275	19.821	19.829	18.778	21.646	24.992	19.483	25.174	24.525
	3.1	15.488	18.936	19.378	20.190	20.222	22.270	18.131	19.923	18.969	19.190	19.770	18.185	21.083	23.109	19.221	23.391	22.890
	3.2	15.445	19.278	19.413	20.326	20.177	22.141	18.038	20.007	18.887	18.938	19.384	17.864	20.933	23.265	18.899	23.457	22.644
	3.3	15.239	19.245	19.358	20.618	19.909	22.230	17.834	19.952	18.993	19.185	19.402	17.943	20.920	23.136	18.664	23.472	22.785
Cardamine hirsuta		*GAPDH*	*TPX1*	*GL1*	*GL2*	*GL3*	*EGL3*	*TTG1*	*TTG2*	*MYC1*	*MYB23*	*CPC*	*TRY*	*ETC1*	*ETC2*	*ETC3*		
	1.1	14.861	18.789	23.876	20.523	20.364	19.986	19.804	21.357	21.400	20.999	23.560	21.231	19.895	22.623	21.247		
	1.2	14.890	18.862	23.666	20.573	20.490	20.452	19.649	21.519	21.500	21.239	23.254	21.322	19.872	22.765	21.345		
	1.3	14.618	19.034	23.656	20.507	20.618	14.858	19.622	21.439	21.627	21.302	23.423	21.376	20.042	20.768	21.368		
	2.1	15.370	19.127	24.866	21.542	21.775	21.408	19.966	22.569	22.039	23.152	24.382	22.388	21.925	23.113	23.130		
	2.2	15.426	19.147	24.665	21.369	21.906	21.481	19.911	22.695	21.949	23.315	24.245	22.358	21.982	23.124	22.919		
	2.3	15.467	19.119	24.693	21.413	21.921	21.426	20.015	22.559	22.074	23.295	24.223	22.339	22.063	23.125	22.917		
	3.1	15.991	18.917	24.955	21.040	22.914	21.846	19.878	23.147	22.461	22.828	24.537	22.618	22.549	22.722	22.746		
	3.2	15.796	18.947	24.850	21.328	22.920	22.003	19.943	23.094	22.571	22.809	24.446	22.555	22.653	22.722	22.723		
	3.3	15.992	19.082	24.892	20.935	22.925	21.994	19.922	23.219	22.547	23.037	24.245	22.584	22.563	22.797	22.661		

Table 7.2: Overview of parameters in the model given in (7.4) - (7.12) and their biological interpretation.

Parameter	Interpretation
θ_1	TTG1 basal production
θ_2	TTG1 degradation
θ_3	TTG1-GL3 binding
θ_4	TTG1 diffusion
θ_5	GL1 basal production
θ_6	GL1 activation by AC2
θ_7	GL1 degradation
θ_8	GL1-GL3 binding
θ_9	GL3 basal production
θ_{10}	GL3 activation by AC1
θ_{11}	Saturation of GL3 AC1 activation
θ_{12}	GL3 activation by AC2
θ_{13}	Saturation of GL3 AC2 activation
θ_{14}	GL3 degradation
θ_{15}	GL3-TRY binding
θ_{16}	GL3-CPC binding
θ_{17}	GL3-ETC binding
θ_{18}	TRY activation by AC1
θ_{19}	TRY degradation
θ_{20}	TRY diffusion
θ_{21}	CPC activation by AC2
θ_{22}	CPC degradation
θ_{23}	CPC diffusion
θ_{24}	ETC activation by AC1
θ_{25}	ETC activation by AC2
θ_{26}	ETC degradation
θ_{27}	ETC diffusion
θ_{28}	AC1 degradation
θ_{29}	AC2 degradation
θ_{30}	GL1 diffusion
θ_{31}	GL3 diffusion

Chapter 8

Discussion

Every chapter of this thesis has treated different aspects of trichome patterning. We have seen how each aspect may be described by a mathematical model. In **Chapters 3** and **5** we have studied mutant phenotypes and developed models that provide predictions based on the mechanistic causes behind those phenotypes. In **Chapter 4** we have shown how biochemical interactions can be quantitatively analysed through the use of various models that describe different modes of protein binding. And, finally, in **Chapters 6** and **7** we studied the evolution of the gene regulatory networks behind trichome patterning and we have used modelling approaches that have provided insight into the observed differences between species. A common challenge in the aforementioned chapters is dealing with uncertainty, which will be discussed in the first part of this chapter. Next, we discuss the simplifying assumptions that were made to develop the models in this thesis and why these assumptions were chosen. Consequently, we will discuss specific aspects of the novel insights into trichome patterning that were found, followed by a discussion on the generality of those findings.

8.1 Meaningful modelling under uncertainty

A common challenge in all of the modelling approaches in the thesis is dealing with uncertainty. This topic has been treated in detail in **Chapter 2**. Specifically, the most prominent source of uncertainty we have to deal with in modelling trichome patterning is that of incomplete knowledge and insufficient data. The consequence thereof is that almost all of the models have unidentifiable parameters [102, 237], with the exception of the models in **Chapter 4**, which describe the binding behaviour of the protein complex driving trichome initiation. Although quantification of these binding constants was also plagued by uncertainty in the form of inaccurate measurements (e.g. estimating protein ratios through Western blot intensity), it still turned out to

be possible to identify the relevant parameters using the available data. Note that these binding models and corresponding data were on a much simpler scale than the spatial patterning systems treated in the other chapters, making it all the more challenging to arrive at meaningful models in these cases. In the following, we will discuss how we have dealt with uncertainty for these more challenging cases.

We have made use of global approaches in order to arrive at minimal networks that could explain experimental observations like the loss of TTG1 depletion in the *ttg1-9* mutant in **Chapter 3** and the clustering behaviour in the inhibitor mutants in **Chapter 5**. Such global constrained approaches depend on a set of criteria that are derived from experimental data with the aim of deriving a model that is as simple as possible (i.e., constrained by the data) and for which all (or most) of the parameters have to be estimated by global sampling procedures. The problem remains, however, that none of the parameter sets that followed from these global searches are unique, i.e. the models were structurally unidentifiable. The question is whether any prediction made or hypothesis tested by the model could still be considered meaningful when different combinations of parameters yield the same outcome, this means that there exist multiple parameter sets that satisfy the criteria (e.g., pattern quantities like trichome density) which the model should fulfil. Note that these different parameter values may correspond to differences in biological processes, as illustrated by the following example: a parameter set with an increased production of the activator GL1 shows the same pattern as a parameter set with a decreased activity of the inhibitor TRY, namely in both cases the trichome density is increased compared to 'wild-type'. So, with respect to the research question "*What mechanism may cause an increase in trichome density?*", the model cannot provide an unambiguous answer. As a way to deal with these issues, throughout this thesis we have focussed on developing models that contained only the core components, where the choice of what to include was guided by experimental observations and the model output was strictly screened for multiple criteria through global parameter searches. What these components are and how they follow from the chosen criteria differs per research question. To illustrate this point, we next consider the modelling approach that we followed in **Chapter 3** to explain the *ttg1-9* mutant.

In **Chapter 3** we started with a substrate-depletion model that contained only GL3 and TTG1 and soon found that it would be impossible to reproduce patterns as observed in *ttg1-9* mutants. Given that the two models driv-

ing trichome patterning (activator inhibitor and substrate-depletion) contain overlapping genes, it logically followed to include an activator-inhibitor motif into the existing substrate-depletion model. While this extended model was able to match the mutant phenotype, it could not reproduce the observed loss of TTG1 depletion. This finally led us to include a second inhibitor into the model, thus allowing differential inhibitor behaviour and increased robustness (i.e. pattern stability was less sensitive to the *ttg1-9* mutation). In following these steps, we arrived at a model that matched all the criteria that characterized the *ttg1-9* mutant. Although we could not pinpoint a single parameter set that uniquely defined these characteristics, we did show that out of all of the tested networks, this final version was the only one for which multiple parameter sets could be found that matched *all* the criteria. While the initial hypothesis was that the loss of depletion was the main mechanism behind the mutant phenotype [141], we showed that a model relying exclusively on the depletion of TTG1 via its interaction with GL3 was not sufficient and that there were more mechanisms at play, like the differential dimer formation and differing ranges in effect of the inhibitors. And so, despite a large amount of uncertainty, we found a core network that could robustly (i.e., multiple points in parameter space) reproduce specific aspects of the mutant, where smaller, simpler models could not. Note that this consideration of 'robustness' in terms of multiple parameter sets is meant to capture the expected robustness of the trichome patterning system; given that all of the parameters are unknown, we are searching for a region or multiple regions in parameter space that can reproduce the pattern, which increases the confidence that the core network has been found as opposed to a single or only a couple of points in parameter space. In conclusion, the strength in this approach did not lie in the identification of parameters, but rather to show that a perturbation as seemingly small as a point mutation may lead to a cascade of unexpected and non-linear events that ultimately describe the full complexity behind the *ttg1-9* phenotype. Such concepts would be hard to grasp without the aid of a model and a global approach such as the one employed in this chapter, which is a common problem in studies on patterning and development [238, 239]. In **Chapter 2** we have described a method which facilitates parameter searches such as the Monte Carlo (MC) approach used in **Chapter 3**. Where MC is a straightforward brute-force method that requires high sampling numbers, the spectral methods in **Chapter 2** provide a smarter and more efficient choice of points for which the model output should be sampled. As we have shown in the example of the Schnakenberg patterning model in **Chapter 2**, the convergence of polynomial chaos expansions (PCEs) is slow around bifurcations. As a way to circumvent these issues, we have

applied Haar wavelets as alternative basis functions to the PCE. In the case of patterning models, such discontinuities are to be expected and so an appropriate basis function set could be chosen beforehand. However, in some cases it is not immediately obvious and the initial use of PCE could show slow convergence. In the last of example of **Chapter 2** we have applied this method to a trichome model and show that it achieved the same accuracy as MC sampling 20 times faster. In order to use them most effectively, spectral expansions require more *a priori* knowledge about the model behaviour than MC approaches, like the existence of bifurcations mentioned before. In general, spectral and other meta-modelling methods show great promise when it comes to global systems approaches such as those described in this thesis. While the use of spectral methods requires more background knowledge (e.g., appropriate basis functions) than the brute-force MC, we have shown that it can be as simple as following a set of steps as described in **Chapter 2**. In our view, this makes spectral methods widely applicable in the field of systems biology.

8.2 Simplifying assumptions

The art of developing a model requires careful consideration of the underlying biological mechanisms. The relevant aspects are included into the model and other details are abstracted away. Typically, these simplifactions are based on certain assumptions that follow from biological knowledge. In the following, we will discuss the relevant assumptions for trichome patterning and whether these are valid. In particular, we focus on steady state assumptions, grid design and stochasticity.

8.2.1 Do steady state assumptions hold?

Trichome cells develop asynchronously across the leaf. The youngest trichomes are found at the base of the leaf and the most mature ones at the tip [34]. However, there is no gradient of development across the axis of the leaf, since newer trichomes can intercalate in between older ones [34]. The final pattern that results from these differences in initiation timing is one where trichomes are regularly spaced on the leaf [32, 33]. In our model, we simulate patterns on a static domain and quantify pattern-features like density and regularity at steady state. This is an obvious simplification from the dynamic process of trichome initiation, where in later developmental stages the aforementioned intercalation influences the pattern [34]. This means that our simulations reflect the initial pattern at the base of young leaf and assume that

this functions as a pre-pattern for the mature leaf, thereby ignoring effects like intercalation. Similarly, due to our static domain, we ignore the impact of leaf growth and shape on the pattern. It has been shown that patterning by reaction-diffusion models can be influenced by domain effects like growth and geometry [20] and it is therefore conceivable that this plays a role in the later stages of leaf development. Futhermore, in the phenotypes considered in this thesis, there appeared to be no preferred localization towards certain parts of the leaf [33], which is therefore not considered. Another assumption underlying this simplification is that the patterning proteins in the model function together as a type of precursor pattern [240], that acts on a faster time-scale than processes like cell-division and tissue growth, i.e., these processes on the protein-level are assumed to be in equilibrium at the moment of trichome initiation. Given the difference in the relevant rates (e.g. the rate cell division is in the order of hours, compared to protein binding rates which are in the order of seconds), we consider this a plausible assumption. So far, a static domain and steady state simulations have proven useful in elucidating trichome mutant phenotypes. The models in this thesis were simple enough to be able to capture the essence of the core patterning principles. However, there is evidence that the trichome patterning module is linked to that of cell-cycle and -division [185]. To gain more understanding about how this influences the pattern it would be required to extend the analysis to models that include cell growth and division. Specialized packages for modelling plants, e.g., Virtual Laboratory and L-studio [241], OpenAlea [242] and Virtual Leaf [243], could facilitate the development of such a model.

8.2.2 Does the simulated grid capture the relevant aspects of plant tissue?

In our simulations, we simulate the tissue of epidermal cells as a hexagonal grid. This is based on the topological properties of the tissue which leads to cells tending to have six neighbours on average [244, 245]. Our model is a discretized version of the reaction-diffusion model, where every discretized unit represents a cell or compartment and the change of concentration of molecules within that cell is described by an ODE, i.e. reactions are localized to a specific cell. This means that the system consists of $N_{\text{species}} \times N_{\text{cell}}$ number of coupled ODEs, where N_{species} is the number of chemical species and N_{cell} the total number of cells on the grid. By rescaling the concentrations by the size of the cells h, we arrive at a mathematically equivalent system that is independent of h [246]. There is no experimental evidence that epidermal cells receive signalling from underlying cell layers that influence trichome

patterning [10]. Therefore, the simulations have zero-flux boundary conditions, which means that none of the proteins can diffuse out of the domain and there are no influences from outside the domain, i.e. the epidermis is considered to be an isolated system in which trichome pattenring occurs. All of these details together represent a simplified leaf epidermis. As a result of these simplifications, the connectivity between cells plays an important role in patterning.

In our models, cells are connected to their neighbours and components which are assumed to be non-cell-autonomous can move between connected cells. This effect is modelled as jumps between cells (similar as in e.g., Erban and Chapman [247]). This movement is uniform, i.e., the probability of a molecule 'jumping' to a neighbouring cell is equal for each of the six neighbours. Given that the movement of patterning proteins is assumed to occur via symplastic transport by the plasmodesmata, this deterministic transport is a simplification. The permeability as well as the distribution of plasmodesmata are ways of controlling the transport through the channels [248], which has been shown to be influential in the patterning of lateral root formation [249]. Not much is known yet about the role of plasmodesmata in trichome patterning, but given the importance of the mobility of activators and inhibitors, it is likely that the regulation of transport can be a crucial factor in patterning [250]. In the models in this thesis, the focus has been on the interactions between patterning genes and proteins. Therefore, details like regulated transport have not been considered. However, it could be an interesting aspect for future studies, especially in the case of cluster formation as seen in the *try cpc* double mutant, where mobility plays a crucial role. An additional transport-regulating mechanism which has not been included in trichome patterning models so far is the sub-cellular localization of proteins. It has been shown in *A. thaliana* that the protein MYC1 regulates the transport of activators and inhibitors to and from the nucleus [60]. As a result of shuttling proteins to the nucleus, they are sequestered in that cell, meaning that the nuclear shuttling by MYC1 may function as a way of regulating transport and activity of activators and inhibitors.

8.2.3 The role of stochasticity in trichome patterning

In our model as well as in reaction-diffusion models in general, the initial conditions are random perturbations about the homogeneous steady state [3, 123]. This means that from the same point in parameter space, different patterns can arise due to the slight variations in initial conditions. In the case of

trichome patterning, this could be one of the sources that causes the differences between individual leaves and plants, where no two trichome patterns are exactly the same [32]. However, this is not the only source of stochasticity that one can expect to play a role in pattern variability. In *Arabidopsis*, gene expression has been shown to be noisy [32, 33, 251] which may be a part of the quasi-regularity seen in trichome patterns. For example, stochastic fluctuations in gene expression have been shown by Okamoto *et al.* to increase the variability in the trichome distribution pattern [39]. As has been shown for Turing's reaction-diffusion model, this stochasticity could improve robustness [20], which likely holds relevance for trichome patterning as well.

So far, we have only treated gene fluctuations in the initial conditions and not in the individual components. In **Chapter 3** we have quantified the extent of irregularity of the *ttg1-9* pattern which at first glance appeared randomized. We have shown that seemingly random aspects of the pattern were not characteristic of a homogeneous random patterning process, e.g. the probability of cluster formation in a random process depends inversely on density (the opposite of what is seen in the mutant). We also found correlation dimensions smaller than expected for a random process, which indicated that the observed variability seen in the pattern depended for a considerable part on deterministic factors [146]. In this case, we reasoned that the added complexity of stochastic gene expression would not be necessary to capture the core principles behind the mutant phenotype. It is conceivable that there are mutants which will be sensitive or dependent on stochastic processes. For example, in the pattern formation of giant cells in *Arabidopsis*, it has been found that fluctuations in the AtML1 transcription factor may be critical for the patterning process [252, 253]. Given that the MBW complex plays a role in both the formation of giant cells and trichomes, a similar process may be relevant for trichome patterning as well.

8.3 Findings on mechanisms of trichome patterning

In the previous sections we have seen the influence of uncertainty and the assumptions underlying the models in the thesis. Here, we will discuss the most important novel findings that followed from our modelling approaches. First, we discuss the potential role of cell differentiation on patterning, as highlighted in **Chapter 5**. Building upon this, we discuss the finding that the mobility of patterning proteins was crucial in **Chapter 5**. Finally, we discuss the role of the complex formation that reoccurs in every model in this thesis

and describe the newly found interactions that are relevant for patterning.

8.3.1 Cell differentiation and patterning

In **Chapter 5** we have dealt with similar issues surrounding the inhibitor mutants. The development of the model was not merely a matter of adding additional biologically relevant components to the network (as was the case in **Chapter 3**), but rather the introduction of an additional mechanism. When we discovered that the hypothesized network could not reproduce the double mutant phenotype, we did not think the answer lies in extending the network further, but rather in the lack of a missing element. For this reason, we turned back to the biological evidence and asked what the difference was between the double mutant patterns in the model and those seen in *Arabidopsis*. Schellmann *et al.* observed that trichomes in the double mutant clusters appeared to show an age difference, suggesting that there is some link between the inhibitors and the timing of trichome initiation [50]. In our model, we postulated that this link could be the effect of cell differentiation, which affects the regulation of the patterning genes after trichome cell fate has been established. The model predicts that patterning genes are down-regulated or completely turned off upon differentiation, which in the case of the double mutant may lead to clustering behaviour. We also tested the opposite: due to endoreduplication the DNA content and thus the amount of patterning proteins is increased. For this type of perturbation we found no effect, i.e. the pattern at steady state was not changed after the perturbation. Bramsiepe *et al.* have shown that a reduction in endoreduplication can override the patterning defect in the *try cpc* double mutant, which indicates that endoreduplication is likely to play a role in the mutant cluster formation [185]. However, in our model, merely the effect of increased DNA content as a result of endoreduplication is not the factor that leads to clusters with different initiation times of individual trichomes within the cluster. Following the network postulated by Bramsiepe *et al.*[185], it is likely that endoreduplication events play a role after initiation and during trichome development, not exclusively at the moment of differentiation as it is modelled in **Chapter 5**. Although in our model it was sufficient to simulate a mechanism of negative regulation at steady state to provide an explanation for the *try cpc* phenotype, it is easy to envision a role for endoreduplication as a fine-tuning mechanism influenced by patterning proteins which stabilizes cell fate, functioning in parallel with the down-regulation of genes upon differentiation as we simulated it on our model.

8.3.2 Mobility of patterning proteins

An additional mechanism that was influential in reproducing the *try cpc* phenotype was the mobility of the activator proteins. This has been shown for some of the components in the network, but for now only assumed for the others. Initial bombardment experiments in leek cells have shown promising results for GL1 (on-going research) where movement between cells was detected, but requires further validation to show that this occurs in leaves of *A. thaliana* as well. Given that transport of these proteins likely occurs via plasmodesmata, we have only included mobility terms for the single proteins, not the complexes between them. These complexes are transcriptionally active and are therefore likely found in the nucleus [60]. Note that this assumption of plasmodesmatal transport comes with a certain complexity (e.g., diffusion is not uniform as described previously) that is currently not captured by any of the existing trichome patterning models, which was discussed in more detail in Section 8.2 above.

8.3.3 The role of the MBW complex in patterning

The **M**YB-**b**HLH-**W**D40 (MBW) complex plays a crucial role in trichome patterning. In **Chapter 4** we have studied the interaction of the MBW components in closer detail. Through binding assays by LUMIER, we have quantified the relative dissociation constants for combinations of GL3, TTG1, GL1, TRY and CPC. Surprisingly, we found that the activators showed stronger binding than the inhibitors, raising the question how inhibitors compensate for this effect in exerting their repressive function. Using the estimated dissociation constants, we predicted that TRY requires to be present in at least four-fold higher amounts than GL3 in order have 50% inactive complexes. This would suggest that in cells in which trichome cell fate is repressed, inhibitors would have to be present in abundance compared to activators. It is unlikely that this is the only mechanism by which inhibition is regulated, e.g., regulation of sub-cellular localization of activators and inhibitors could also play an important role [60]. In addition to the mobility of the activators, the inhibitor mobility is just as crucial. As described in the Introduction, activator-inhibitor models show stringent demands on the differing diffusion rates of activators and inhibitors. It has been shown [23], however, that extensions to the original two-component system can alleviate this stringency, i.e., a broader range of diffusion rates may still lead to stable patterns . The results from our model are in line with these findings. In the case of the *try cpc* mutant, it is the difference in the mobility of the individual inhibitors that plays a crucial role in

regulating cluster- and trichome-density, where we found varying ranges of mobility in activators. As their single mutant phenotypes suggest [50], CPC acts on a long range and TRY on short ranges, which was in line with the parameters that reproduced the criteria for the double mutant in our model.

Finally, we have seen how this MBW complex is capable of driving trichome patterning under differing expression levels as seen in the comparative genetic studies in **Chapters 6** and **7**. These studies showed that the core patterning network is surprisingly flexible under varying conditions. What has typically been attributed to such robustness is the functional redundancy in plant GRNs [254–256], also in the case of trichome patterning [35]. In comparing between closely related species like the Brassicaceae considered in **Chapters 6** and **7**, the core network consisting of three inhibitors and three activators was enough to reproduce the observed evolutionary differences on the gene expression levels. The role of redundancy will likely become even more important when considering individual mutations between species in potential future studies, as we have already seen in **Chapter 6** for the opposite phenotypes in *A. thaliana* and *A. alpina*.

8.4 Generality of trichome patterning models

Several basic principles can be identified in trichome patterning. For one, aspects like lateral inhibition and local activation are based on models by Gierer and Meinhardt, and the preceding reaction-diffusion model of Turing [3, 5]. In those models, patterning is assumed to start with a field of cells that have equal amounts of concentrations in activators and inhibitors, where stochastic fluctuations break this equilibrium, followed by the enhancement of small, local differences by autocatalytic loops. This in turn leads to local accumulation of activators and inhibitors, thus specifying which cells may eventually obtain trichome identity. In trichome patterning we also consider this to be the mechanism by which patterns are formed, given that these patterns develop without any pre-existing information. The same principle is also found in other patterning systems, such as Delta-Notch signalling in e.g., the *Drosophila* wing vein system [257], and the formation of stripes in zebrafish [258]. In plants, patterning systems based on similar principles have also been found, like the vascular patterning of plant stems [259], and root hair patterning, which shows a large genetic overlap with trichome patterning [8]. Although it is apparent that different systems based on the same patterning principles have already been described, there are certain advantages to the trichome system

that make it especially suited to studies on these principles. One advantage is the ease of genetic manipulation, as is evident from the isolation of many patterning mutants [34]. Another advantage is the accessibility of the network itself, which appears to consist of relatively simple interactions, thus allowing an analysis beyond qualitative and intuitive understanding, towards one that can be quantified. In such a quantitative approach, mathematical models such as the ones applied in this thesis and other publications [38–42] are essential in capturing the essence of the core aspects of spatial patterning. Specific to the trichome network, the existence of the autocatalytic feedback-loop – a crucial mechanism for breaking the initial symmetry in reaction-diffusion models – is yet to be shown experimentally. Progress towards this is being made through optogenetics approaches in single cells where activators can be controlled in spatial and temporal precision, such that feedback mechanisms could potentially be identified. If the activator auto-catalytic feedback-loops could be identified in that way, it would confirm one of the crucial mechanisms predicted by the models. This mechanism is the requirement for the development of patterns in activator-inhibitor models by the dominance of the activator peak over the inhibitor peak. So far, experiments have shown that the activator and inhibitor peaks overlap in position in trichome patterning, however, it has not been confirmed that the respective production rates are such that activator amounts dominate inhibitor amounts at those positions.

Another general feature of the trichome system is the role of the MBW complex. This complex is known to regulate multiple processes in plants, such as anthocyanin pigmentation [260], regulation of the flavonoid pathway [67], root hair formation [8], seed coat pigmentation, seed coat mucilage production [261] and stomata- and pavement cell differentiation [66]. Findings like the difference in dissociation constants of the MBW-components in **Chapter 4** and the mechanisms of competitive binding that underlie trichome patterning [38, 64], are relevant for these other contexts as well. Individual differences are expected, e.g., the role of WER in root patterning [233], but the molecular basis of MBW interactions constitutes general features. As we have seen in **Chapters 6** and **7** and as is reviewed by Robinson and Roeder [66], these features are conserved in different species, making it a suitable system for studies on evolutionary comparisons.

8.5 Conclusion and outlook

The way natural systems organise themselves into functional shape, form and structure is a phenomenon that has evoked the interest of biologists for centuries. More recently, also mathematicians and physicists have started to contribute to questions on pattern formation in living organisms through modelling approaches. In this thesis, trichome pattern formation in *Arabidopsis thaliana* and related species has been used as a model system to study spatial patterning in plant development. In answering questions about mutant phenotypes, molecular and genetic interactions, and evolutionary differences, we have used mathematical models to help unravel the complexity behind the trichome patterning system.

We have used mathematical models to test hypotheses (e.g. the loss of depletion results in patterning defects in TTG1 mutants in **Chapter 3**) as well as arriving at new hypotheses about relevant mechanisms in patterning (e.g. down-regulation of patterning genes upon differentiation results in clusters in inhibitor mutants in **Chapter 5**). We have shown validation of models (e.g. rescue of *ttg1-9* clustering phenotype by expressing TRY under the CPC promoter in **Chapter 3**) as well as deriving predictions that still require experimental verification (e.g. mobility of activators in **Chapter 5**). All in all, the models in each chapter have aided our understanding of trichome patterning and the general principles behind it, despite being faced with challenges such as large amounts of uncertainty. Many questions still remain, such as the impact of leaf growth and stochasticity in gene expression and transport, in which models will certainly prove to be instrumental.

From an evolutionary perspective, combined genetic and modelling approaches such as the ones developed in **Chapter 6** and **7**, could be useful in identifying the relevant genes in other plant species. Such comparative studies can reveal how gene regulatory networks evolve and adapt to differing circumstances and genetic influences, as seen in different accessions of *Arabidopsis* [39, 262–264]. Such findings could be relevant in species that hold value for crop industry, for example the trichomes of cotton plants which are used in textile industry [76], or the glandular trichomes of *Artemisia annua* which produce artimisinin, used in drugs for malaria treatment [77].

The aim of this thesis was to provide minimal models for a conceptual description that captures the essential features of trichome pattern formation. In this we tackled experimental and theoretical challenges and contributed to

unravelling mechanistic characteristics of the patterning model, which consists of principles that are generic to many other systems. Like other patterning systems, the modelling of trichome patterning comes with certain challenges, such as dealing with large amounts of uncertainty. Despite this, it has proven to be an excellent model system due to its relative simplicity and ease of genetic manipulability, and will undoubtedly continue to prove its worth in future research on pattern formation and plant development.

Bibliography

[1] Lewis Wolpert. Positional information and the spatial pattern of cellular differentiation. *Journal of theoretical biology*, 25(1):1–47, 1969. ISSN 0022-5193.

[2] Darcy Wentworth Thompson and D'Arcy W Thompson. *On growth and form*, volume 2. Cambridge university press Cambridge, 1942.

[3] Alan Mathison Turing. The chemical basis of morphogenesis. *Bulletin of mathematical biology*, 52(1):153–197, 1952.

[4] Shigeru Kondo and Takashi Miura. Reaction-diffusion model as a framework for understanding biological pattern formation. *science*, 329 (5999):1616–1620, 2010. ISSN 0036-8075.

[5] Hans Meinhardt and Alfred Gierer. Pattern formation by local self-activation and lateral inhibition. *Bioessays*, 22(8):753–760, 2000. ISSN 0265-9247.

[6] Donald A Levin. The role of trichomes in plant defense. *The quarterly review of biology*, 48(1, Part 1):3–15, 1973. ISSN 0033-5770.

[7] Martin Hülskamp. Plant trichomes: a model for cell differentiation. *Nature Reviews Molecular Cell Biology*, 5(6):471–480, 2004.

[8] Tetsuya Ishida, Tetsuya Kurata, Kiyotaka Okada, and Takuji Wada. A Genetic Regulatory Network in the Development of Trichomes and Root Hairs. *Annual Review of Plant Biology*, 59(1):365–386, apr 2008. ISSN 1543-5008. doi: 10.1146/annurev.arplant.59.032607.092949. URL http://dx.doi.org/10.1146/annurev.arplant.59.032607.092949.

[9] Sitakanta Pattanaik, Barunava Patra, Sanjay Kumar Singh, and Ling Yuan. An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. *Frontiers in Plant Science*, 5(JUN):1–8, 2014. ISSN 1664462X. doi: 10.3389/fpls.2014.00259.

[10] J C Larkin, N Young, M Prigge, and M D Marks. The control of trichome spacing and number in Arabidopsis. *Development*, 122(3):997

LP – 1005, mar 1996. URL http://dev.biologists.org/content/122/3/997.abstract.

[11] Vincent Castets, Etiennette Dulos, Jacques Boissonade, and Patrick De Kepper. Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern. *Physical review letters*, 64(24):2953, 1990.

[12] Eric Dessaud, Andrew P McMahon, and James Briscoe. Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. 2008. ISSN 1477-9129.

[13] Pascal Dollé, Juan-Carlos Izpisúa-Belmonte, Hildegard Falkenstein, Armand Renucci, and Denis Duboule. Coordinate expression of the murine Hox-5 complex homoeobox-containing genes during limb pattern formation. *Nature*, 342(6251):767–772, 1989. ISSN 1476-4687.

[14] Randy L Johnson and Clifford J Tabin. Molecular models for vertebrate limb development. *Cell*, 90(6):979–990, 1997. ISSN 0092-8674.

[15] Katherine W Rogers and Alexander F Schier. Morphogen gradients: from generation to interpretation. *Annual review of cell and developmental biology*, 27(1):377–407, 2011.

[16] Neil H Shubin and Pere Alberch. A morphogenetic approach to the origin and basic organization of the tetrapod limb. In *Evolutionary biology*, pages 319–387. Springer, 1986.

[17] Hans Meinhardt and A Gierer. Applications of a theory of biological pattern formation based on lateral inhibition. *Journal of cell science*, 15 (2):321–346, 1974.

[18] James D Murray. *Mathematical biology II: spatial models and biomedical applications*, volume 3. Springer New York, 2001.

[19] Leah Edelstein-Keshet. *Mathematical models in biology*. SIAM, 2005. ISBN 0898715547.

[20] Philip K Maini, Thomas E Woolley, Ruth E Baker, Eamonn A Gaffney, and S Seirin Lee. Turing's model for biological pattern formation and the robustness problem. *Interface focus*, 2(4):487–496, 2012. ISSN 2042-8898.

[21] Iain Barrass, Edmund J Crampin, and Philip K Maini. Mode transitions in a model reaction–diffusion system driven by domain growth and noise. *Bulletin of mathematical biology*, 68(5):981–995, 2006. ISSN 1522-9602.

[22] Natalie S Scholes, David Schnoerr, Mark Isalan, and Michael P H Stumpf. A comprehensive network atlas reveals that turing patterns are common but not robust. *Cell systems*, 9(3):243–257, 2019. ISSN 2405-4712.

[23] Luciano Marcon, Xavier Diego, James Sharpe, and Patrick Müller. High-throughput mathematical analysis identifies Turing networks for patterning with equally diffusing signals. *Elife*, 5:e14022, 2016. ISSN 2050-084X.

[24] Thomas E Woolley, Ruth E Baker, Eamonn A Gaffney, and Philip K Maini. Stochastic reaction and diffusion on growing domains: understanding the breakdown of robust pattern formation. *Physical Review E*, 84(4):46216, 2011.

[25] M Mocarlo Zheng, Bin Shao, and Qi Ouyang. Identifying network topologies that can generate turing pattern. *Journal of theoretical biology*, 408:88–96, 2016. ISSN 0022-5193.

[26] J D Murray. Parameter space for Turing instability in reaction diffusion mechanisms: a comparison of models. *Journal of theoretical biology*, 98 (1):143–163, 1982. ISSN 0022-5193.

[27] Edmund J Crampin, Eamonn A Gaffney, and Philip K Maini. Reaction and diffusion on growing domains: scenarios for robust pattern formation. *Bulletin of mathematical biology*, 61(6):1093–1120, 1999. ISSN 0092-8240.

[28] A J Koch and Hans Meinhardt. Biological pattern formation: from basic mechanisms to complex structures. *Reviews of modern physics*, 66(4): 1481, 1994.

[29] Alfred Gierer and Hans Meinhardt. A theory of biological pattern formation. *Kybernetik*, 12(1):30–39, 1972. ISSN 1432-0770.

[30] Hans Meinhardt. *The algorithmic beauty of sea shells*. Springer Science & Business Media, 2009. ISBN 3540921427.

[31] Martina Pesch and Martin Hülskamp. Creating a two-dimensional pattern de novo during Arabidopsis trichome and root hair initiation. *Current opinion in genetics & development*, 14(4):422–427, 2004. doi: 10.1016/j.gde.2004.06.007.

[32] Bettina Greese, K Wester, Robert Bensch, Olaf Ronneberger, J Timmer, M Hu Lskamp, and C Fleck. Influence of cell-to-cell variability on spatial pattern formation. *IET systems biology*, 6(4):143–153, 2012.

[33] Bettina Greese, Martin Hülskamp, and Christian Fleck. Quantification of variability in trichome patterns. *Frontiers in plant science*, 5:596, 2014. ISSN 1664-462X.

[34] Martin Hülskamp, Simon Miséra, and Gerd Jürgens. Genetic dissection of trichome cell development in Arabidopsis. *Cell*, 76(3):555–566, 1994.

[35] Marina Pesch and Martin Hülskamp. One, two, three... models for trichome patterning in Arabidopsis? *Current opinion in plant biology*, 12(5):587–592, 2009.

[36] Markus Grebe. The patterning of epidermal hairs in Arabidopsis—updated. *Current opinion in plant biology*, 15(1):31–37, 2012.

[37] Alexey V Doroshkov, Dmitrii K Konstantinov, Dmitrij A Afonnikov, and Konstantin V Gunbin. The evolution of gene regulatory networks controlling Arabidopsis thaliana L. trichome development. *BMC plant biology*, 19(1):71–85, 2019. ISSN 1471-2229.

[38] Simona Digiuni, Swen Schellmann, Florian Geier, Bettina Greese, Martina Pesch, Katja Wester, Burcu Dartan, Valerie Mach, Bhylahalli Purushottam Srinivas, Jens Timmer, Christian Fleck, and Martin Hulskamp. A competitive complex formation mechanism underlies trichome patterning on Arabidopsis leaves. *Molecular Systems Biology*, 1 (217):1–11, 2008. doi: 10.1038/msb.2008.54.

[39] Shotaro Okamoto, Kohei Negishi, Yuko Toyama, Takeo Ushijima, and Kengo Morohashi. Leaf trichome distribution pattern in arabidopsis reveals gene expression variation associated with environmental adaptation. *Plants*, 9(7):909, 2020.

[40] Daniel Bouyer, Florian Geier, Friedrich Kragler, Arp Schnittger, Martina Pesch, Katja Wester, Rachappa Balkunde, Jens Timmer, Christian Fleck, and Martin Hülskamp. Two-Dimensional Patterning by

a Trapping / Depletion Mechanism : The Role of TTG1 and GL3 in Arabidopsis Trichome Formation. *PLoS Biology*, 6(6), 2008. doi: 10.1371/journal.pbio.0060141.

[41] Mariana Benitez, Carlos Espinosa-Soto, Pablo Padilla-Longoria, Jose Diaz, and Elena R Alvarez-Buylla. Equivalent genetic regulatory recover contrasting spatial cell networks in different contexts patterns that resemble those in Arabidopsis root and leaf epidermis: a dynamic model. 2007.

[42] Mariana Benitez, Carlos Espinosa-Soto, Pablo Padilla-Longoria, and Elena R Alvarez-Buylla. Interlinked nonlinear subnetworks underlie the formation of robust cellular patterns in Arabidopsis epidermis: a dynamic spatial model. *BMC Systems Biology*, 2(1):1–16, 2008.

[43] Mingzhe Zhao, Kengo Morohashi, Greg Hatlestad, Erich Grotewold, and Alan Lloyd. The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. *Development*, 135(11):1991–1999, 2008.

[44] C Thomas Payne, Fan Zhang, and Alan M Lloyd. GL3 encodes a bHLH protein that regulates trichome development in Arabidopsis through interaction with GL1 and TTG1. *Genetics*, 156(3):1349–1362, 2000. ISSN 0016-6731.

[45] David G Oppenheimer, Patricia L Herman, Shan Sivakumaran, Jeffrey Esch, and M David Marks. A myb gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. *Cell*, 67(3):483–493, 1991.

[46] M Koornneeff, L W M Dellaert, and J H der Veen. EMS-and relation-induced mutation frequencies at individual loci in Arabidopsis thaliana (L.) Heynh. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 93(1):109–123, 1982.

[47] Moira E Galway, James D Masucci, Alan M Lloyd, Virginia Walbot, Ronald W Davis, and John W Schiefelbein. The TTG gene is required to specify epidermal cell fate and cell patterning in the Arabidopsis root. *Developmental biology*, 166(2):740–754, 1994.

[48] M Koornneef. The complex syndrome of ttg mutants. *Arabidopsis information service*, 1981. ISSN 0066-5657.

[49] Amanda R Walker, Paul A Davison, Agnese C Bolognesi-Winfield, Celia M James, N Srinivasan, Tom L Blundell, Jeffrey J Esch, M David Marks, and John C Gray. The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. *The Plant Cell*, 11(7): 1337–1349, 1999. ISSN 1532-298X.

[50] S Schellmann, A Schnittger, V Kirik, T Wada, K Okada, A Beermann, J Thumfahrt, G Jürgens, and M Hülskamp. TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in Arabidopsis. *The EMBO journal*, 21(19):5036–5046, 2002. ISSN 0261-4189.

[51] Takuji Wada, Tatsuhiko Tachibana, Yoshiro Shimura, and Kiyotaka Okada. Epidermal cell differentiation in Arabidopsis determined by a Myb homolog, CPC. *Science*, 277(5329):1113–1116, 1997.

[52] Martina Pesch and Martin Hülskamp. Role of TRIPTYCHON in trichome patterning in Arabidopsis. *BMC Plant Biology*, 11(1):1–14, 2011.

[53] Lijun Gan, Kai Xia, Jin-Gui Chen, and Shucai Wang. Functional characterization of TRICHOMELESS2, a new single-repeat R3 MYB transcription factor in the regulation of trichome patterning in Arabidopsis. *BMC plant biology*, 11(1):1–12, 2011.

[54] Victor Kirik, Marissa Simon, Martin Huelskamp, and John Schiefelbein. The ENHANCER OF TRY AND CPC1 gene acts redundantly with TRIPTYCHON and CAPRICE in trichome and root hair cell patterning in Arabidopsis. *Developmental biology*, 268(2):506–513, 2004.

[55] Rumi Tominaga, Mineko Iwata, Ryosuke Sano, Kayoko Inoue, Kiyotaka Okada, and Takuji Wada. Arabidopsis CAPRICE-LIKE MYB 3 (CPL3) controls endoreduplication and flowering development in addition to trichome and root hair formation. 2008.

[56] Shucai Wang, Leah Hubbard, Ying Chang, Jianjun Guo, John Schiefelbein, and Jin-Gui Chen. Comprehensive analysis of single-repeat R3 MYB proteins in epidermal cell patterning and their transcriptional regulation in Arabidopsis. *BMC plant biology*, 8(1):1–13, 2008.

[57] Shucai Wang, Christa Barron, John Schiefelbein, and Jin-Gui Chen. Distinct relationships between GLABRA2 and single-repeat R3 MYB transcription factors in the regulation of trichome and root hair patterning in Arabidopsis. *New phytologist*, 185(2):387–400, 2010.

[58] Shucai Wang, Su-Hwan Kwak, Qingning Zeng, Brian E Ellis, Xiao-Ya Chen, John Schiefelbein, and Jin-Gui Chen. TRICHOMELESS1 regulates trichome patterning by suppressing GLABRA1 in Arabidopsis. 2007.

[59] Katja Wester, Simona Digiuni, Florian Geier, Jens Timmer, Christian Fleck, and Martin Hulskamp. Functional diversity of R3 single-repeat genes in trichome development. *Development*, 136(9):1487–1496, 2009.

[60] Martina Pesch, Ilka Schultheiß, Simona Digiuni, Joachim F Uhrig, and Martin Hülskamp. Mutual control of intracellular localisation of the patterning proteins AtMYC1, GL1 and TRY/CPC in Arabidopsis. *Development*, 140(16):3456–3467, 2013.

[61] Kengo Morohashi, Mingzhe Zhao, Manli Yang, Betsy Read, Alan Lloyd, Rebecca Lamb, and Erich Grotewold. Participation of the Arabidopsis bHLH factor GL3 in trichome initiation regulatory events. *Plant physiology*, 145(3):736–746, 2007. ISSN 1532-2548.

[62] Kengo Morohashi and Erich Grotewold. A Systems Approach Reveals Regulatory Circuitry for Arabidopsis Trichome Initiation by the GL3 and GL1 Selectors. *PLOS Genetics*, 5(2):e1000396, feb 2009. URL `http://dx.doi.org/10.1371{%}2Fjournal.pgen.1000396`.

[63] William G Rerie, Kenneth A Feldmann, and M David Marks. The GLABRA2 gene encodes a homeo domain protein required for normal trichome development in Arabidopsis. *Genes & development*, 8(12): 1388–1399, 1994. ISSN 0890-9369.

[64] Martina Pesch, Ilka Schultheiß, Karsten Klopffleisch, Joachim F Uhrig, Manfred Koegl, Christoph S Clemen, Rüdiger Simon, Stefanie Weidtkamp-Peters, and Martin Hülskamp. TRANSPARENT TESTA GLABRA1 and GLABRA1 compete for binding to GLABRA3 in Arabidopsis. *Plant Physiology*, 168(2):584–597, 2015.

[65] Nicola A. Ramsay and Beverley J. Glover. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science*, 10(2):63–70, 2005. ISSN 13601385. doi: 10.1016/j.tplants.2004.12.011.

[66] Dana Olivia Robinson and Adrienne H.K. Roeder. Themes and variations in cell type patterning in the plant epidermis. *Current Opinion in Genetics and Development*, 32:55–65, 2015. ISSN 18790380. doi:

10.1016/j.gde.2015.01.008. URL http://dx.doi.org/10.1016/j.gde.2015.01.008.

[67] Wenjia Xu, Christian Dubos, and Loïc Lepiniec. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends in Plant Science*, 20(3):176–185, 2015. ISSN 13601385. doi: 10.1016/j.tplants.2014.12.001.

[68] Bipei Zhang, Divykriti Chopra, Andrea Schrader, and Martin Hülskamp. Evolutionary comparison of competitive protein-complex formation of MYB, bHLH, and WDR proteins in plants. *Journal of Experimental Botany*, 70(12):3197–3209, 2019. ISSN 14602431. doi: 10.1093/jxb/erz155.

[69] Christine Bernhardt, Myeong Min Lee, Antonio Gonzalez, Fan Zhang, Alan Lloyd, and John Schiefelbein. The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root. *Development*, 130(26):6431–6439, 2003.

[70] Fred Berger, Jim Haseloff, John Schiefelbein, and Liam Dolan. Positional information in root epidermis is defined during embryogenesis and acts in domains with strict boundaries. *Current Biology*, 8(8):421–430, 1998. ISSN 0960-9822.

[71] Takuji Wada, Tetsuya Kurata, Rumi Tominaga, Yoshihiro Koshino-Kimura, Tatsuhiko Tachibana, Koji Goto, M David Marks, Yoshiro Shimura, and Kiyotaka Okada. Role of a positive regulator of root hair development, CAPRICE, in Arabidopsis root epidermal cell differentiation. 2002. ISSN 1477-9129.

[72] N K Nayidu, S Kagale, A Taheri, T S Withana-Gamage, I A P Parkin, A G Sharpe, and M Y Gruber. Comparison of Five Major Trichome Regulatory Genes in Brassica villosa with Orthologues within the Brassicaceae. *Plos One*, 9(4), 2014. doi: ARTNe9587710.1371/journal.pone.0095877.

[73] D Chopra, H Wolff, J Span, S Schellmann, G Coupland, M C Albani, A Schrader, and M Hulskamp. Analysis of TTG1 function in Arabis alpina. *BMC Plant Biol*, 14:16, 2014. doi: 10.1186/1471-2229-14-16. URL https://www.ncbi.nlm.nih.gov/pubmed/24406039.

[74] M Fernandez-Mazuecos and B J Glover. The evo-devo of plant speciation. *Nature Ecology & Evolution*, 1(4), 2017. doi: UNSP011010.1038/s41559-017-0110.

[75] Xiaojing Wang, Chao Shen, Pinghong Meng, Guofei Tan, and Litang Lv. Analysis and review of trichomes in plants. *BMC Plant Biology*, 21(1): 70, 2021. ISSN 1471-2229. doi: 10.1186/s12870-021-02840-x. URL https://doi.org/10.1186/s12870-021-02840-x.

[76] Shui Wang, Jia-Wei Wang, Nan Yu, Chun-Hong Li, Bin Luo, Jin-Ying Gou, Ling-Jian Wang, and Xiao-Ya Chen. Control of plant trichome development by a cotton fiber MYB gene. *The Plant Cell*, 16(9):2323–2334, 2004. ISSN 1532-298X.

[77] Lies Maes, Filip C W Van Nieuwerburgh, Yansheng Zhang, Darwin W Reed, Jacob Pollier, Sofie R F Vande Casteele, Dirk Inzé, Patrick S Covello, Dieter L D Deforce, and Alain Goossens. Dissection of the phytohormonal regulation of trichome formation and biosynthesis of the antimalarial compound artemisinin in Artemisia annua plants. *New Phytologist*, 189(1):176–189, 2011. ISSN 0028-646X.

[78] Thomas Payne, John Clement, David Arnold, and Alan Lloyd. Heterologous myb genes distinct from GL1 enhance trichome production when overexpressed in Nicotiana tabacum. *Development*, 126(4):671–682, 1999. ISSN 1477-9129.

[79] M D S Nunes, S Arif, C Schlotterer, and A P McGregor. A Perspective on Micro-Evo-Devo: Progress and Potential. *Genetics*, 195(3):625–634, 2013. doi: 10.1534/genetics.113.156463.

[80] P Simpson. Evolution of development in closely related species of flies and worms. *Nature Reviews Genetics*, 3(12):907–917, 2002. doi: 10.1038/nrg947.

[81] Ben Scheres. Non-linear signaling for pattern formation? *Current opinion in plant biology*, 3(5):412–417, 2000.

[82] M David Marks. Molecular genetic analysis of trichome development in Arabidopsis. *Annual review of plant biology*, 48(1):137–163, 1997. ISSN 1040-2519.

[83] T J Sullivan. *Introduction to Uncertainty Quantification*. Springer International Publishing, 2015. ISBN ISBN 978-3-319-23395-6.

[84] William J Blake, Mads Kærn, Charles R Cantor, and James J Collins. Noise in eukaryotic gene expression. *Nature*, 422(6932):633–637, 2003. ISSN 1476-4687.

[85] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic Gene Expression in a Single Cell. *Science (New York, NY)*, 297 (5584):1183–1186, aug 2002. doi: 10.1126/science.1070919.

[86] Mukund Thattai and Alexander Van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, 2001. ISSN 0027-8424.

[87] Lev S Tsimring. Noise in biology. *Reports on Progress in Physics*, 77(2): 26601, 2014. ISSN 0034-4885.

[88] F James. Monte Carlo Theory and Practice. *Reports on Progress in Physics*, 43(9):1145–1189, 1980.

[89] Adrian Barbu and Song-Chun Zhu. *Monte Carlo Methods*. Springer Singapore, Singapore, 2020. ISBN 9789811329708 9789811329715. doi: 10.1007/978-981-13-2971-5.

[90] Ilya M Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280, 2001.

[91] Ishaan L Dalal, Deian Stefan, and Jared Harwayne-Gidansky. Low discrepancy sequences for Monte Carlo simulations on reconfigurable platforms. In *2008 International Conference on Application-Specific Systems, Architectures and Processors*, pages 108–113. IEEE, 2008. ISBN 1424418976.

[92] Jon C Helton and Freddie Joe Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23–69, 2003. ISSN 0951-8320.

[93] Peter W Glynn and Donald L Iglehart. Importance sampling for stochastic simulations. *Management science*, 35(11):1367–1392, 1989. ISSN 0025-1909.

[94] Olivier Le Maitre and Omar M Knio. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media, 2010.

[95] Bruno Sudret. Uncertainty propagation and sensitivity analysis in mechanical models–Contributions to structural reliability and stochastic spectral methods. *Habilitationa diriger des recherches, Université Blaise Pascal, Clermont-Ferrand, France*, 147, 2007.

[96] Géraud Blatman and Bruno Sudret. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics*, 25(2):183–197, 2010.

[97] Joel A Paulson, Marc Martin-Casas, and Ali Mesbah. Fast uncertainty quantification for dynamic flux balance analysis using non-smooth polynomial chaos expansions. *PLoS computational biology*, 15 (8):e1007308, 2019.

[98] Liesbet Geris, David Gomez-Cabrero, and Others. *Uncertainty in biology*. Springer, 2016.

[99] Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3(10): e189, 2007.

[100] Eshan D Mitra and William S Hlavacek. Parameter estimation and uncertainty quantification for systems biology models. *Current opinion in systems biology*, 18:9–18, 2019.

[101] Simon van Mourik, Cajo Ter Braak, Hans Stigter, and Jaap Molenaar. Prediction uncertainty assessment of a systems biology model requires a sample of the full probability distribution of its parameters. *PeerJ*, 2: e433, 2014.

[102] Andreas Raue, Marcel Schilling, Julie Bachmann, Andrew Matteson, Max Schelke, Daniel Kaschek, Sabine Hug, Clemens Kreutz, Brian D Harms, Fabian J Theis, and Others. Lessons learned from quantitative dynamical modeling in systems biology. *PloS one*, 8(9):e74335, 2013.

[103] Roger Ghanem, David Higdon, and Houman Owhadi. *Handbook of uncertainty quantification*, volume 6. Springer, 2017.

[104] Paul Kirk, Daniel Silk, and Michael P H Stumpf. Reverse engineering under uncertainty. In *Uncertainty in biology*, pages 15–32. Springer, 2016.

[105] Areti Tsigkinopoulou, Aliah Hawari, Megan Uttley, and Rainer Breitling. Defining informative priors for ensemble modeling in systems biology. *Nature protocols*, 13(11):2643–2663, 2018.

[106] William Stafford Noble and Others. Support vector machine applications in computational biology. *Kernel methods in computational biology*, 71:92, 2004.

[107] Ryszard Tadeusiewicz. Neural networks as a tool for modeling of biological systems. *Bio-Algorithms and Med-Systems*, 11(3):135–144, 2015.

[108] Darren J Wilkinson. Bayesian methods in bioinformatics and computational systems biology. *Briefings in bioinformatics*, 8(2):109–116, 2007.

[109] Marc Martin-Casas and Ali Mesbah. Discrimination between competing model structures of biological systems in the presence of population heterogeneity. *IEEE life sciences letters*, 2(3):23–26, 2016.

[110] Stefan Streif, Felix Petzke, Ali Mesbah, Rolf Findeisen, and Richard D Braatz. Optimal experimental design for probabilistic model discrimination using polynomial chaos. *IFAC Proceedings Volumes*, 47(3):4103–4109, 2014.

[111] Marissa Renardy, Tau-Mu Yi, Dongbin Xiu, and Ching-Shan Chou. Parameter uncertainty quantification using surrogate models applied to a spatial model of yeast mating polarization. *PLoS computational biology*, 14(5):e1006181, 2018.

[112] Dongbin Xiu and George Em Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2):619–644, 2002.

[113] Michael Eldred, Clayton Webster, and Paul Constantine. Evaluation of non-intrusive approaches for Wiener-Askey generalized polynomial chaos. In *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 16th AIAA/ASME/AHS Adaptive Structures Conference, 10th AIAA Non-Deterministic Approaches Conference, 9th AIAA Gossamer Spacecraft Forum, 4th AIAA Multidisciplinary Des*, page 1892, 2008.

[114] Alexandre Joel Chorin. Gaussian fields and random flow. *Journal of Fluid Mechanics*, 63(1):21–32, 1974.

[115] O P Le Maitre, O M Knio, H N Najm, and R G Ghanem. Uncertainty propagation using Wiener–Haar expansions. *Journal of computational Physics*, 197(1):28–57, 2004.

[116] Dongbin Xiu. Efficient collocational approach for parametric uncertainty analysis. *Communications in computational physics*, 2(2):293–309, 2007.

[117] Brian P Ingalls. *Mathematical {{Modeling}} in {{Systems Biology}}*. An {{Introduction}}. MIT Press, jul 2013. ISBN 0-262-01888-8.

[118] Brian Ingalls. Sensitivity Analysis: From Model Parameters to System Behaviour. *Essays in Biochemistry*, 45:177–194, sep 2008. ISSN 0071-1365, 1744-1358. doi: 10.1042/bse0450177.

[119] A Saltelli, editor. *Global Sensitivity Analysis: The Primer*. John Wiley, Chichester, England ; Hoboken, NJ, 2008. ISBN 978-0-470-05997-5.

[120] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, 93(7):964–979, 2008.

[121] Géraud Blatman and Bruno Sudret. Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliability Engineering & System Safety*, 95(11):1216–1229, 2010.

[122] Steven H Strogatz. *Nonlinear Dynamics and {{Chaos}}: With Applications to Physics, Biology, Chemistry, and Engineering*. Studies in Nonlinearity. Addison-Wesley Pub, 1994. ISBN 978-0-201-54344-5.

[123] James D Murray. *Mathematical biology: I. An introduction*, volume 17. Springer Science & Business Media, 2001.

[124] Adrian W Bowman and Adelchi Azzalini. *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, volume 18. OUP Oxford, 1997.

[125] Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

[126] Fabio Nobile, Raúl Tempone, and Clayton G Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.

[127] Benno Hess and Arnold Boiteux. Mechanism of Glycolytic Oscillation in Yeast, I. Aerobic and anaerobic growth conditions for obtaining glycolytic oscillation. *Biological Chemistry*, 349(2):1567–1574, 1968.

[128] Ping Liu, Junping Shi, Yuwen Wang, and Xiuhong Feng. Bifurcation analysis of reaction-diffusion Schnakenberg model. *Journal of Mathematical Chemistry*, 51(8):2001–2019, 2013.

[129] Gene H Golub and John H Welsch. Calculation of Gauss quadrature rules. *Mathematics of computation*, 23(106):221–230, 1969.

[130] Samih Zein, Benoit Colson, and François Glineur. An efficient sampling method for regression-based polynomial chaos expansion. *Communications in computational physics*, 13(4):1173–1188, 2013.

[131] Noura Fajraoui, S Marelli, and B Sudret. On optimal experimental designs for sparse polynomial chaos expansions. *arXiv preprint arXiv:1703.05312*, 2017.

[132] Rachappa Balkunde, Martina Pesch, and Martin Hülskamp. Trichome patterning in Arabidopsis thaliana: from genetic to molecular models. *Current topics in developmental biology*, 91:299–321, 2010.

[133] Rumi Tominaga-Wada, Tetsuya Ishida, and Takuji Wada. New insights into the mechanism of development of Arabidopsis root hairs and trichomes. *International review of cell and molecular biology*, 286:67–106, 2011.

[134] Victor Kirik, Myeong Min Lee, Katja Wester, Ullrich Herrmann, Zhengui Zheng, David Oppenheimer, John Schiefelbein, and Martin Hulskamp. Functional diversification of MYB23 and GL1 genes in trichome morphogenesis and initiation. 2005.

[135] Victor Kirik, Arp Schnittger, Volodymyr Radchuk, Klaus Adler, Martin Hülskamp, and Helmut Bäumlein. Ectopic expression of the Arabidopsis AtMYB23 gene induces differentiation of trichome cells. *Developmental biology*, 235(2):366–377, 2001.

[136] Fan Zhang, Antonio Gonzalez, Mingzhe Zhao, C Thomas Payne, and Alan Lloyd. A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. *Development*, 130(20): 4859–4869, 2003.

[137] Ying Gao, Ximing Gong, Wanhong Cao, Jinfeng Zhao, Liqin Fu, Xuechen Wang, Karen S Schumaker, and Yan Guo. SAD2 in Arabidopsis functions in trichome initiation through mediating GL3 function and regulating GL1, TTG1 and GL2 expression. *Journal of Integrative Plant Biology*, 50(7):906–917, 2008.

[138] Shucai Wang and Jin-Gui Chen. Arabidopsis transient expression analysis reveals that activation of GLABRA2 may require concurrent binding of GLABRA1 and GLABRA3 to the promoter of GLABRA2. *Plant and cell physiology*, 49(12):1792–1804, 2008.

[139] Ilona M Zimmermann, Marc A Heim, Bernd Weisshaar, and Joachim F Uhrig. Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal*, 40(1):22–34, 2004.

[140] Jeffrey J Esch, Margaret Chen, Mark Sanders, Matthew Hillestad, Sampson Ndkium, Brian Idelkope, James Neizer, and M David Marks. A contradictory GLABRA3 allele helps define gene interactions controlling trichome development in Arabidopsis. 2003.

[141] Rachappa Balkunde, Daniel Bouyer, and Martin Hülskamp. Nuclear trapping by GL3 controls intercellular transport and redistribution of TTG1 protein in Arabidopsis. *Development*, 138(22):5039–5048, 2011.

[142] John C Larkin, David G Oppenheimer, Alan M Lloyd, Ellen T Paparozzi, and M David Marks. Roles of the GLABROUS1 and TRANSPARENT TESTA GLABRA genes in Arabidopsis trichome development. *The Plant Cell*, 6(8):1065–1076, 1994.

[143] John C Larkin, Jason D Walker, Agnese C Bolognesi-Winfield, John C Gray, and Amanda R Walker. Allele-specific interactions between ttg and gl1 during trichome development in Arabidopsis thaliana. *Genetics*, 151(4):1591–1604, 1999.

[144] Yun Long and John Schiefelbein. Novel TTG1 mutants modify root-hair pattern formation in Arabidopsis. *Frontiers in plant science*, 11:383, 2020.

[145] Henrik Failmezger, Benjamin Jaegle, Andrea Schrader, Martin Hülskamp, and Achim Tresch. Semi-automated 3D leaf reconstruction and analysis of trichome patterning from light microscopic images. *PLoS computational biology*, 9(4):e1003029, 2013.

[146] James Theiler. Efficient algorithm for estimating the correlation dimension from a set of discrete points. *Physical review A*, 36(9):4456, 1987.

[147] Nobuhide Ueki, Tamaki Oda, Maiko Kondo, Kazuhiro Yano, Teruhisa Noguchi, and Masa-aki Muramatsu. Selection system for genes encoding nuclear-targeted proteins. *Nature biotechnology*, 16(13):1338–1342, 1998.

[148] Arp Schnittger, Ulrike Folkers, Birgit Schwab, Gerd Jürgens, and Martin Hülskamp. Generation of a spacing pattern: the role of TRIPTYCHON in trichome patterning in Arabidopsis. *The Plant Cell*, 11(6):1105–1116, 1999.

[149] Daniel B Szymanski and M David Marks. GLABROUS1 overexpression and TRIPTYCHON alter the cell cycle and trichome cell fate in Arabidopsis. *The Plant Cell*, 10(12):2047–2062, 1998.

[150] Tetsuya Kurata, Tetsuya Ishida, Chie Kawabata-Awai, Masahiro Noguchi, Sayoko Hattori, Ryosuke Sano, Ryoko Nagasaka, Rumi Tominaga, Yoshihiro Koshino-Kimura, Tomohiko Kato, and Others. Cell-to-cell movement of the CAPRICE protein in Arabidopsis root epidermal cell differentiation. 2005.

[151] Rumi Tominaga-Wada and Takuji Wada. Extended C termini of CPC-LIKE MYB proteins confer functional diversity in Arabidopsis epidermal cell differentiation. *Development*, 144(13):2375–2380, 2017.

[152] Tamara L Western, Debra J Skinner, and George W Haughn. Differentiation of mucilage secretory cells of the Arabidopsis seed coat. *Plant physiology*, 122(2):345–356, 2000.

[153] Allen Sessions, Detlef Weigel, and Martin F Yanofsky. The Arabidopsis thaliana MERISTEM LAYER 1 promoter specifies epidermal expression in meristems and young primordia. *The Plant Journal*, 20(2):259–263, 1999.

[154] R Daniel Gietz, Robert H Schiestl, Andrew R Willems, and Robin A Woods. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, 11(4):355–360, 1995.

[155] P Holgate. Tests of randomness based on distance methods. *Biometrika*, 52(3/4):345–353, 1965.

[156] Philip J Clark and Francis C Evans. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35(4):445–453, 1954. ISSN 0012-9658.

[157] Oded Sandler, Sivan Pearl Mizrahi, Noga Weiss, Oded Agam, Itamar Simon, and Nathalie Q Balaban. Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature*, 519(7544):468–471, 2015.

[158] James E Skinner. Low-dimensional chaos in biological systems. *Bio/technology*, 12(6):596–600, 1994.

[159] Peter Grassberger and Itamar Procaccia. Characterization of strange attractors. *Physical review letters*, 50(5):346, 1983.

[160] Antje Feller, Katja Machemer, Edward L Braun, and Erich Grotewold. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant Journal*, 66(1):94–116, apr 2011. ISSN 0960-7412. doi: https://doi.org/10.1111/j.1365-313X.2010.04459.x. URL https://doi.org/10.1111/j.1365-313X.2010.04459.x.

[161] Laura Serna and Cathie Martin. Trichomes: different regulatory networks lead to convergent structures. *Trends in plant science*, 11(6):274–280, 2006. ISSN 1360-1385.

[162] Steven van Nocker and Philip Ludwig. The WD-repeat protein superfamily in Arabidopsis: Conservation and divergence in structure and function. *BMC Genomics*, 4:1–11, 2003. ISSN 14712164. doi: 10.1186/1471-2164-4-50.

[163] Bipei Zhang and Andrea Schrader. TRANSPARENT TESTA GLABRA 1-dependent regulation of flavonoid biosynthesis. *Plants*, 6(4):1–30, 2017. ISSN 22237747. doi: 10.3390/plants6040065.

[164] Cornelis Spelt, Francesca Quattrocchio, Joseph Mol, and Ronald Koes. ANTHOCYANIN1 of petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms. *The Plant Cell*, 14(9):2121–2135, 2002. ISSN 1532-298X.

[165] Francesca Quattrocchio, Walter Verweij, Arthur Kroon, Cornelis Spelt, Joseph Mol, and Ronald Koes. PH4 of Petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *The Plant Cell*, 18(5):1274–1291, 2006. ISSN 1532-298X.

[166] Xiao-Xia Shangguan, Chang-Qing Yang, Xiu-Fang Zhang, and Ling-Jian Wang. Functional characterization of a basic helix-loop-helix

(bHLH) transcription factor GhDEL65 from cotton (Gossypium hirsutum). *Physiologia Plantarum*, 158(2):200–212, 2016. ISSN 0031-9317.

[167] Que Kong, Sitakanta Pattanaik, Antje Feller, Joshua R Werkman, Chenglin Chai, Yongqin Wang, Erich Grotewold, and Ling Yuan. Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. *Proceedings of the National Academy of Sciences*, 109(30):E2091–E2097, 2012. ISSN 0027-8424.

[168] Nick W Albert, Kevin M Davies, David H Lewis, Huaibi Zhang, Mirco Montefiori, Cyril Brendolise, Murray R Boase, Hanh Ngo, Paula E Jameson, and Kathy E Schwinn. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *The Plant Cell*, 26(3):962–980, 2014. ISSN 1532-298X.

[169] M D Marks, J Esch, P Herman, S Sivakumaran, and D Oppenheimer. A model for cell-type determination and differentiation in plants. *Symp Soc Exp Biol*, 45:77–87, 1991. URL https://www.ncbi.nlm.nih.gov/pubmed/1843415.

[170] Patricia L Herman and M David Marks. Trichome development in Arabidopsis thaliana. II. Isolation and complementation of the GLABROUS1 gene. *The Plant Cell*, 1(11):1051–1055, 1989. ISSN 1532-298X.

[171] Shucai Wang and Jin-Gui Chen. Regulation of cell fate determination by single-repeat R3 MYB transcription factors in Arabidopsis. *Frontiers in plant science*, 5:133, 2014. ISSN 1664-462X.

[172] Xuemei Dai, Limei Zhou, Wei Zhang, Ling Cai, Hongyan Guo, Hainan Tian, John Schiefelbein, and Shucai Wang. A single amino acid substitution in the R3 domain of GLABRA1 leads to inhibition of trichome formation in Arabidopsis without affecting its interaction with GLABRA3. *Plant, Cell & Environment*, 39(4):897–907, apr 2016. ISSN 0140-7791. doi: https://doi.org/10.1111/pce.12695. URL https://doi.org/10.1111/pce.12695.

[173] Sylwia Struk, Anse Jacobs, Elena Sánchez Martín-Fontecha, Kris Gevaert, Pilar Cubas, and Sofie Goormachtig. Exploring the protein–protein interaction landscape in plants. *Plant, Cell & Environment*, 42(2):387–409, feb 2019. ISSN 0140-7791. doi: https://doi.org/10.1111/pce.13433. URL https://doi.org/10.1111/pce.13433.

[174] Jelena Brkljacic and Erich Grotewold. Combinatorial control of plant gene expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1860(1):31–40, 2017. ISSN 1874-9399. doi: https://doi.org/10.1016/j.bbagrm.2016.07.005. URL https://www.sciencedirect.com/science/article/pii/S1874939916301377.

[175] Marco Fambrini and Claudio Pugliesi. The Dynamic Genetic-Hormonal Regulatory Network Controlling the Trichome Development in Leaves, 2019.

[176] Zelou Wei, Yalong Cheng, Chenchen Zhou, Dong Li, Xin Gao, Shuoxin Zhang, and Mingxun Chen. Genome-Wide Identification of Direct Targets of the TTG1–bHLH–MYB Complex in Regulating Trichome Formation and Flavonoid Accumulation in Arabidopsis Thaliana, 2019.

[177] Sonja Blasche and Manfred Koegl. Analysis of protein–protein interactions using LUMIER assays. In *Virus-Host Interactions*, pages 17–27. Springer, 2013.

[178] Clemens Kreutz, Andreas Raue, Daniel Kaschek, and Jens Timmer. Profile likelihood in systems biology. *The FEBS journal*, 280(11):2564–2571, 2013. ISSN 1742-464X.

[179] Gang Liang, Hua He, Yang Li, Qin Ai, and Diqiu Yu. MYB82 functions in regulation of trichome development in Arabidopsis. *Journal of experimental botany*, 65(12):3215–3223, jul 2014. ISSN 1460-2431. doi: 10.1093/jxb/eru179. URL https://pubmed.ncbi.nlm.nih.gov/24803498https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4071844/.

[180] Daniel Bouyer, Florian Geier, Friedrich Kragler, Arp Schnittger, Martina Pesch, Katja Wester, Rachappa Balkunde, Jens Timmer, Christian Fleck, and Martin Hülskamp. Two-dimensional patterning by a trapping/depletion mechanism: the role of TTG1 and GL3 in Arabidopsis trichome formation. *PLoS biology*, 6(6):e141, 2008.

[181] Michael R Green and Joseph Sambrook. Molecular cloning. *A Laboratory Manual 4th*, 2012.

[182] Zhi-Xin Wang. An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS letters*, 360(2):111–114, 1995. ISSN 0014-5793.

[183] Rachappa Balkunde, Anna Deneer, Hanna Bechtel, Bipei Zhang, Stefanie Herberth, Martina Pesch, Benjamin Jaegle, Christian Fleck, and Martin Hülskamp. Identification of the Trichome Patterning Core Network Using Data from Weak ttg1 Alleles to Constrain the Model Space. *Cell Reports*, 33(11), 2020. ISSN 22111247. doi: 10.1016/j.celrep.2020.108497.

[184] Divykriti Chopra, Mona Mapar, Lisa Stephan, Maria C. Albani, Anna Deneer, George Coupland, Eva Maria Willing, Swen Schellmann, Korbinian Schneeberger, Christian Fleck, Andrea Schrader, and Martin Hülskamp. Genetic and molecular analysis of trichome development in Arabis alpina. *Proceedings of the National Academy of Sciences of the United States of America*, 116(24):12078–12083, 2019. ISSN 10916490. doi: 10.1073/pnas.1819440116.

[185] J Bramsiepe, K Wester, C Weinl, F Roodbarkelari, R Kasili, J C Larkin, M Hulskamp, and A Schnittger. Endoreplication controls cell fate maintenance. *PLoS Genet*, 6(6):e1000996, 2010. doi: 10.1371/journal.pgen.1000996. URL https://www.ncbi.nlm.nih.gov/pubmed/20585618.

[186] Tianya Wang, Qiming Jia, Wei Wang, Saddam Hussain, Sajjad Ahmed, Dao-xiu Zhou, Zhongfu Ni, and Shucai Wang. GCN5 modulates trichome initiation in Arabidopsis by manipulating histone acetylation of core trichome initiation regulator genes. *Plant cell reports*, 38(6):755–765, 2019. ISSN 1432-203X.

[187] Stefanie Rosa, Vardis Ntoukakis, Nobuko Ohmido, Ali Pendle, Rita Abranches, and Peter Shaw. Cell differentiation and development in Arabidopsis are associated with changes in histone dynamics at the single-cell level. *The Plant Cell*, 26(12):4821–4833, 2014. ISSN 1532-298X.

[188] William J Lucas. Plasmodesmata: intercellular channels for macromolecular transport in plants. *Current opinion in cell biology*, 7(5):673–680, 1995. ISSN 0955-0674.

[189] Matt Geisler, Jeanette Nadeau, and Fred D Sack. Oriented asymmetric divisions that generate the stomatal spacing pattern in Arabidopsis are disrupted by the too many mouths mutation. *The Plant Cell*, 12(11):2075–2086, 2000. ISSN 1532-298X.

[190] Kalyanmoy Deb and J Sundar. Reference point based multi-objective optimization using evolutionary algorithms. In *Proceedings of the 8th*

annual conference on Genetic and evolutionary computation, pages 635–642, 2006.

[191] James Whitbread Lee Glaisher. XXXII. On a class of definite integrals. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 42(280):294–302, 1871. ISSN 1941-5982.

[192] M D Purugganan. The molecular evolution of development. *Bioessays*, 20(9):700–711, 1998. doi: Doi10.1002/(Sici)1521-1878(199809)20:9<700::Aid-Bies3>3.0.Co;2-K.

[193] M A Beilstein, I A Al-Shehbaz, and E A Kellogg. Brassicaceae phylogeny and trichome evolution. *Am J Bot*, 93(4):607–619, 2006. doi: 10.3732/ajb.93.4.607. URL https://www.ncbi.nlm.nih.gov/pubmed/21646222.

[194] M A Koch, C Kiefer, D Ehrich, J Vogel, C Brochmann, and K Mummenhoff. Three times out of Asia Minor: the phylogeography of Arabis alpina L. (Brassicaceae). *Molecular Ecology*, 15(3):825–839, 2006. doi: 10.1111/j.1365-294X.2005.02848.x. URL https://www.ncbi.nlm.nih.gov/pubmed/16499705.

[195] M A Beilstein, N S Nagalingum, M D Clements, S R Manchester, and S Mathews. Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43):18724–18728, 2010. doi: 10.1073/pnas.0909766107. URL https://www.ncbi.nlm.nih.gov/pubmed/20921408.

[196] E M Willing, V Rawat, T Mandakova, F Maumus, G V James, K J Nordstrom, C Becker, N Warthmann, C Chica, B Szarzynska, M Zytnicki, M C Albani, C Kiefer, S Bergonzi, L Castaings, J L Mateos, M C Berns, N Bujdoso, T Piofczyk, L de Lorenzo, C Barrero-Sicilia, I Mateos, M Piednoel, J Hagmann, R Chen-Min-Tao, R Iglesias-Fernandez, S C Schuster, C Alonso-Blanco, F Roudier, P Carbonero, J Paz-Ares, S J Davis, A Pecinka, H Quesneville, V Colot, M A Lysak, D Weigel, G Coupland, and K Schneeberger. Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. *Nature Plants*, 1:14023, 2015. doi: 10.1038/nplants.2014.23. URL https://www.ncbi.nlm.nih.gov/pubmed/27246759.

[197] M Hulskamp. Plant trichomes: a model for cell differentiation. *Nat Rev Mol Cell Biol*, 5(6):471–480, 2004. doi: 10.1038/nrm1404. URL https://www.ncbi.nlm.nih.gov/pubmed/15173826.

[198] H O Lee, J M Davidson, and R J Duronio. Endoreplication: polyploidy with purpose. *Genes Dev*, 23(21):2461–2477, 2009. doi: 10.1101/gad.1829209. URL https://www.ncbi.nlm.nih.gov/pubmed/19884253.

[199] J Traas, M Hulskamp, E Gendreau, and H Hofte. Endoreduplication and development: rule without dividing? *Curr Opin Plant Biol*, 1(6):498–503, 1998. URL https://www.ncbi.nlm.nih.gov/pubmed/10066638.

[200] S Guimil and C Dunand. Cell growth and differentiation in Arabidopsis epidermal cells. *J Exp Bot*, 58(14):3829–3840, 2007. doi: 10.1093/jxb/erm253. URL https://www.ncbi.nlm.nih.gov/pubmed/18162628.

[201] U Folkers, J Berger, and M Hulskamp. Cell morphogenesis of trichomes in Arabidopsis: differential control of primary and secondary branching by branch initiation regulators and cell growth. *Development*, 124(19):3779–3786, 1997. URL https://www.ncbi.nlm.nih.gov/pubmed/9367433.

[202] X Zhang, P H Grey, S Krishnakumar, and D G Oppenheimer. The IRREGULAR TRICHOME BRANCH loci regulate trichome elongation in Arabidopsis. *Plant Cell Physiol*, 46(9):1549–1560, 2005. doi: 10.1093/pcp/pci168. URL https://www.ncbi.nlm.nih.gov/pubmed/16043432.

[203] D B Szymanski. Breaking the WAVE complex: the point of Arabidopsis trichomes. *Curr Opin Plant Biol*, 8(1):103–112, 2005. doi: 10.1016/j.pbi.2004.11.004. URL https://www.ncbi.nlm.nih.gov/pubmed/15653407.

[204] C Fornero, B X Suo, M Zahde, K Juveland, and V Kirik. Papillae formation on trichome cell walls requires the function of the mediator complex subunit Med25. *Plant Molecular Biology*, 95(4-5):389–398, 2017. doi: 10.1007/s11103-017-0657-x.

[205] W B Jiao, G G Accinelli, B Hartwig, C Kiefer, D Baker, E Severing, E M Willing, M Piednoel, S Woetzel, E Madrid-Herrero, B Huettel, U Humann, R Reinhard, M A Koch, D Swan, B Clavijo, G Coupland, and K Schneeberger. Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. *Genome Res*, 27(5):778–786, 2017. doi: 10.1101/gr.213652.116. URL https://www.ncbi.nlm.nih.gov/pubmed/28159771.

[206] R Wang, S Farrona, C Vincent, A Joecker, H Schoof, F Turck, C Alonso-Blanco, G Coupland, and M C Albani. PEP1 regulates perennial flowering in Arabis alpina. *Nature*, 459(7245):423–427, 2009. doi: 10.1038/nature07988. URL https://www.ncbi.nlm.nih.gov/pubmed/19369938.

[207] D B Szymanski, R A Jilk, S M Pollock, and M D Marks. Control of GL2 expression in Arabidopsis leaves and trichomes. *Development*, 125(7):1161–1171, 1998. URL https://www.ncbi.nlm.nih.gov/pubmed/9477315.

[208] E R Valdivia, D Chevalier, J Sampedro, I Taylor, C E Niederhuth, and J C Walker. DVL genes play a role in the coordination of socket cell recruitment and differentiation. *J Exp Bot*, 63(3):1405–1412, 2012. doi: 10.1093/jxb/err378. URL https://www.ncbi.nlm.nih.gov/pubmed/22112938.

[209] J E Melaragno, B Mehrotra, and A W Coleman. Relationship between Endopolyploidy and Cell Size in Epidermal Tissue of Arabidopsis. *Plant Cell*, 5(11):1661–1668, 1993. doi: 10.1105/tpc.5.11.1661. URL https://www.ncbi.nlm.nih.gov/pubmed/12271050.

[210] M Hulskamp, S Misra, and G Jurgens. Genetic dissection of trichome cell development in Arabidopsis. *Cell*, 76(3):555–566, 1994. URL https://www.ncbi.nlm.nih.gov/pubmed/8313475.

[211] K J Nordstrom, M C Albani, G V James, C Gutjahr, B Hartwig, F Turck, U Paszkowski, G Coupland, and K Schneeberger. Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. *Nat Biotechnol*, 31(4):325–330, 2013. doi: 10.1038/nbt.2515. URL https://www.ncbi.nlm.nih.gov/pubmed/23475072.

[212] M C Albani, L Castaings, S Wotzel, J L Mateos, J Wunder, R Wang, M Reymond, and G Coupland. PEP1 of Arabis alpina is encoded by two overlapping genes that contribute to natural genetic variation in perennial flowering. *PLoS Genet*, 8(12):e1003130, 2012. doi: 10.1371/journal.pgen.1003130. URL https://www.ncbi.nlm.nih.gov/pubmed/23284298.

[213] S Bergonzi, M C Albani, E Ver Loren van Themaat, K J Nordstrom, R Wang, K Schneeberger, P D Moerland, and G Coupland. Mechanisms of age-dependent response to winter temperature in perennial flowering of Arabis alpina. *Science*, 340(6136):1094–1097, 2013. doi: 10.1126/science.1234116. URL https://www.ncbi.nlm.nih.gov/pubmed/23723236.

[214] E Lyons, B Pedersen, J Kane, M Alam, R Ming, H Tang, X Wang, J Bowers, A Paterson, D Lisch, and M Freeling. Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol*, 148(4):1772–1781, 2008. doi: 10.1104/pp.108.124867. URL https://www.ncbi.nlm.nih.gov/pubmed/18952863.

[215] D D Pollock and J C Larkin. Estimating the degree of saturation in mutant screens. *Genetics*, 168(1):489–502, 2004. doi: 10.1534/genetics.103.024430. URL https://www.ncbi.nlm.nih.gov/pubmed/15454559.

[216] S F Altschul, T L Madden, A A Schaffer, J Zhang, Z Zhang, W Miller, and D J Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17):3389–3402, 1997. URL https://www.ncbi.nlm.nih.gov/pubmed/9254694.

[217] A Marchler-Bauer, J B Anderson, P F Cherukuri, C DeWeese-Scott, L Y Geer, M Gwadz, S He, D I Hurwitz, J D Jackson, Z Ke, C J Lanczycki, C A Liebert, C Liu, F Lu, G H Marchler, M Mullokandov, B A Shoemaker, V Simonyan, J S Song, P A Thiessen, R A Yamashita, J J Yin, D Zhang, and S H Bryant. CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res*, 33(Database issue):D192–6, 2005. doi: 10.1093/nar/gki069. URL https://www.ncbi.nlm.nih.gov/pubmed/15608175.

[218] S M Hebsgaard, P G Korning, N Tolstrup, J Engelbrecht, P Rouze, and S Brunak. Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. *Nucleic Acids Res*, 24(17):3439–3452, 1996. URL https://www.ncbi.nlm.nih.gov/pubmed/8811101.

[219] Steven J Clough and Andrew F Bent. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The plant journal*, 16(6):735–743, 1998.

[220] L Stephan, V Tilmes, and M Hulskamp. Selection and validation of reference genes for quantitative Real-Time PCR in Arabis alpina. *PLoS One*, 14(3):e0211172, 2019. doi: 10.1371/journal.pone.0211172. URL https://www.ncbi.nlm.nih.gov/pubmed/30830921.

[221] Elena R Alvarez-Buylla, Mariana Benitéz, Enrique Balleza Dávila, Alvaro Chaos, Carlos Espinosa-Soto, and Pablo Padilla-Longoria. Gene regulatory network models for plant development. *Current Opinion in Plant Biology*, 10(1):83–91, 2007. ISSN 1369-5266.

[222] D Marc Jones and Klaas Vandepoele. Identification and evolution of gene regulatory networks: insights from comparative studies in plants. *Current opinion in plant biology*, 54:42–48, 2020. ISSN 1369-5266.

[223] Maria Katherine Mejia-Guerra, Marcelo Pomeranz, Kengo Morohashi, and Erich Grotewold. From plant gene regulatory grids to network dynamics. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1819(5):454–465, 2012. ISSN 1874-9399.

[224] Yipei Guo and Ariel Amir. Exploring the effect of network topology, mRNA and protein dynamics on gene regulatory network stability. *Nature communications*, 12(1):1–10, 2021. ISSN 2041-1723.

[225] Isabella Schember and Marc S Halfon. Common Themes and Future Challenges in Understanding Gene Regulatory Network Evolution. *Cells*, 11(3):510, 2022. ISSN 2073-4409.

[226] Marcus Koch, Bernhard Haubold, and Thomas Mitchell-Olds. Molecular systematics of the Brassicaceae: evidence from coding plastidic matK and nuclear Chs sequences. *American Journal of Botany*, 88(3): 534–544, 2001. ISSN 0002-9122.

[227] Mark A Beilstein, Ihsan A Al-Shehbaz, Sarah Mathews, and Elizabeth A Kellogg. Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: tribes and trichomes revisited. *American journal of botany*, 95(10):1307–1327, 2008. ISSN 0002-9122.

[228] Jo Vandesompele, Katleen De Preter, Filip Pattyn, Bruce Poppe, Nadine Van Roy, Anne De Paepe, and Frank Speleman. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, 3(7):1–12, 2002. ISSN 1474-760X.

[229] Kenneth D Boese, Andrew B Kahng, and Sudhakar Muddu. A new adaptive multi-start technique for combinatorial global optimizations. *Operations Research Letters*, 16(2):101–113, 1994. ISSN 0167-6377.

[230] M A Stephens. EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, 69(347):730–737, sep 1974. ISSN 0162-1459. doi: 10.1080/01621459.1974.10480196. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1974.10480196.

[231] Francesca Campolongo, Jessica Cariboni, and Andrea Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental modelling & software*, 22(10):1509–1518, 2007. ISSN 1364-8152.

[232] S A Bustin, V Benes, T Nolan, and M W Pfaffl. Quantitative real-time RT-PCR–a perspective. *Journal of molecular endocrinology*, 34(3):597–601, 2005. ISSN 1479-6813.

[233] Myeong Min Lee and John Schiefelbein. WEREWOLF, a MYB-related protein in Arabidopsis, is a position-dependent regulator of epidermal cell patterning. *Cell*, 99(5):473–483, 1999. ISSN 0092-8674.

[234] Ralf Stracke, Martin Werber, and Bernd Weisshaar. The r2r3-myb gene family in arabidopsis thaliana. *Current opinion in plant biology*, 4(5): 447–456, 2001.

[235] Myeong Min Lee and John Schiefelbein. Developmentally distinct myb genes encode functionally equivalent proteins in arabidopsis. *Development*, 128(9):1539–1546, 2001.

[236] Rumi Tominaga-Wada, Yuka Nukumizu, Shusei Sato, Tomohiko Kato, Satoshi Tabata, and Takuji Wada. Functional divergence of myb-related genes, werewolf and atmyb23 in arabidopsis. *Bioscience, biotechnology, and biochemistry*, 76(5):883–887, 2012.

[237] Joseph H A Guillaume, John D Jakeman, Stefano Marsili-Libelli, Michael Asher, Philip Brunner, Barry Croke, Mary C Hill, Anthony J Jakeman, Karel J Keesman, Saman Razavi, and Johannes D Stigter. Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. *Environmental Modelling & Software*, 119:418–432, 2019. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2019.07.007. URL https://www.sciencedirect.com/science/article/pii/S1364815218307278.

[238] Adrián López García de Lomana, Alex Gómez-Garrido, David Sportouch, and Jordi Villa-Freixa. Optimal experimental design in the modelling of pattern formation. In *International Conference on Computational Science*, pages 610–619. Springer, 2008.

[239] J D Stigter, D Joubert, and J Molenaar. Observability of Complex Systems: Finding the Gap. *Scientific Reports*, 7(1):16566, 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-16682-x. URL https://doi.org/10.1038/s41598-017-16682-x.

[240] PhilipáK Maini, KevináJ Painter, and HeleneáNguyen Phong Chau. Spatial pattern formation in chemical and biological systems. *Journal of the Chemical Society, Faraday Transactions*, 93(20):3601–3610, 1997.

[241] Przemyslaw Prusinkiewicz. Art and science of life: designing and growing virtual plants with L-systems. In *XXVI International Horticultural Congress: Nursery Crops; Development, Evaluation, Production and Use 630*, pages 15–28, 2002. ISBN 9066054972.

[242] Christophe Pradal, Samuel Dufour-Kowalski, Frédéric Boudon, Christian Fournier, and Christophe Godin. OpenAlea: a visual programming and component-based software platform for plant modelling. *Functional plant biology*, 35(10):751–760, 2008. ISSN 1445-4416.

[243] Roeland M H Merks, Michael Guravage, Dirk Inzé, and Gerrit T S Beemster. VirtualLeaf: an open-source framework for cell-based modeling of plant tissue growth and development. *Plant physiology*, 155(2):656–666, 2011. ISSN 1532-2548.

[244] Ross Carter, Yara E Sánchez-Corrales, Matthew Hartley, Verônica A Grieneisen, and Athanasius F M Marée. Pavement cells and the topology puzzle. *Development*, 144(23):4386–4397, dec 2017. ISSN 0950-1991. doi: 10.1242/dev.157073. URL https://doi.org/10.1242/dev.157073.

[245] William T Gibson, James H Veldhuis, Boris Rubinstein, Heather N Cartwright, Norbert Perrimon, G Wayne Brodland, Radhika Nagpal, and Matthew C Gibson. Control of the Mitotic Cleavage Plane by Local Epithelial Topology. *Cell*, 144(3):427–438, 2011. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2010.12.035. URL https://www.sciencedirect.com/science/article/pii/S0092867410015278.

[246] Radek Erban, Jonathan Chapman, and Philip Maini. A practical guide to stochastic simulations of reaction-diffusion processes. *arXiv preprint arXiv:0704.1908*, 2007.

[247] Radek Erban and S Jonathan Chapman. Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions. *Physical biology*, 6(4):46001, 2009. ISSN 1478-3975.

[248] J Paul Knox and Yoselin Benitez-Alfonso. Roles and regulation of plant cell walls surrounding plasmodesmata. *Current Opinion in Plant Biology*, 22:93–100, 2014. ISSN 1369-5266. doi: https://doi.org/10.1016/

j.pbi.2014.09.009. URL https://www.sciencedirect.com/science/article/pii/S1369526614001307.

[249] Yoselin Benitez-Alfonso, Christine Faulkner, Ali Pendle, Shunsuke Miyashima, Ykä Helariutta, and Andrew Maule. Symplastic Intercellular Connectivity Regulates Lateral Root Patterning. *Developmental Cell*, 26(2):136–147, 2013. ISSN 1534-5807. doi: https://doi.org/10.1016/j.devcel.2013.06.010. URL https://www.sciencedirect.com/science/article/pii/S1534580713003493.

[250] Friedrich Kragler, William J Lucas, and J. A. N. Monzer. Plasmodesmata: dynamics, domains and patterning. *Annals of Botany*, 81(1):1–10, 1998. ISSN 1095-8290.

[251] Ilka Schultheiß Araújo, Jessica Magdalena Pietsch, Emma Mathilde Keizer, Bettina Greese, Rachappa Balkunde, Christian Fleck, and Martin Hülskamp. Stochastic gene expression in Arabidopsis thaliana. *Nature communications*, 8(1):1–9, 2017. ISSN 2041-1723.

[252] Heather M Meyer, José Teles, Pau Formosa-Jordan, Yassin Refahi, Rita San-Bento, Gwyneth Ingram, Henrik Jönsson, James C W Locke, and Adrienne H K Roeder. Fluctuations of the transcription factor ATML1 generate the pattern of giant cells in the Arabidopsis sepal. *Elife*, 6: e19131, 2017. ISSN 2050-084X.

[253] Adrienne H K Roeder. Use it or average it: stochasticity in plant development. *Current Opinion in Plant Biology*, 41:8–15, 2018. ISSN 1369-5266. doi: https://doi.org/10.1016/j.pbi.2017.07.010. URL https://www.sciencedirect.com/science/article/pii/S1369526617300985.

[254] Megan G Sawchuk, Alexander Edgar, and Enrico Scarpella. Patterning of leaf vein networks by convergent auxin transport pathways. *PLoS genetics*, 9(2):e1003294, 2013. ISSN 1553-7390.

[255] Georgette C Briggs, Karen S Osmont, Chikako Shindo, Richard Sibout, and Christian S Hardtke. Unequal genetic redundancies in Arabidopsis–a neglected phenomenon? *Trends in plant science*, 11(10):492–498, 2006. ISSN 1360-1385.

[256] Otho Mantegazza, Veronica Gregis, Matteo Chiara, Caterina Selva, Giulia Leo, David S Horner, and Martin M Kater. Gene coexpression patterns during early development of the native Arabidopsis reproductive meristem: novel candidate developmental regulators and patterns of

functional redundancy. *The Plant Journal*, 79(5):861–877, 2014. ISSN 0960-7412.

[257] David Sprinzak, Amit Lakhanpal, Lauren LeBon, Leah A Santat, Michelle E Fontes, Graham A Anderson, Jordi Garcia-Ojalvo, and Michael B Elowitz. Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature*, 465(7294):86–90, 2010. ISSN 1476-4687.

[258] Yamaguchi Motoomi, Yoshimoto Eiichi, and Kondo Shigeru. Pattern regulation in the stripe of zebrafish suggests an underlying dynamic and autonomous mechanism. *Proceedings of the National Academy of Sciences*, 104(12):4790–4793, mar 2007. doi: 10.1073/pnas.0607790104. URL https://doi.org/10.1073/pnas.0607790104.

[259] David J Hearn. Turing-like mechanism in a stochastic reaction-diffusion model recreates three dimensional vascular patterning of plant stems. *PLOS ONE*, 14(7):e0219055, jul 2019. URL https://doi.org/10.1371/journal.pone.0219055.

[260] Antonio Gonzalez, Mingzhe Zhao, John M Leavitt, and Alan M Lloyd. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *The Plant Journal*, 53(5):814–827, 2008. ISSN 0960-7412.

[261] Song Feng Li, Patrick J Allen, Ross S Napoli, Richard G Browne, Hanh Pham, and Roger W Parish. MYB–bHLH–TTG1 regulates Arabidopsis seed coat biosynthesis pathways directly and indirectly via multiple tiers of transcription factors. *Plant and Cell Physiology*, 61(5):1005–1018, 2020. ISSN 0032-0781.

[262] Noelia Arteaga, Marija Savic, Belén Méndez-Vigo, Alberto Fuster-Pons, Rafael Torres-Pérez, Juan Carlos Oliveros, F Xavier Picó, and Carlos Alonso-Blanco. MYB transcription factors drive evolutionary innovations in Arabidopsis fruit trichome patterning. *The Plant Cell*, 33(3):548–565, 2021. ISSN 1532-298X.

[263] Marie-Theres Hauser. Molecular basis of natural variation and environmental control of trichome patterning. *Frontiers in Plant Science*, 5:320, 2014. ISSN 1664-462X.

[264] Julia Hilscher, Christian Schlötterer, and Marie-Theres Hauser. A single amino acid replacement in ETC2 shapes trichome patterning in natural Arabidopsis populations. *Current Biology*, 19(20):1747–1751, 2009. ISSN 0960-9822.

Summary

Trichomes are unicellular hairs that originate from epidermal cells and show a regular distribution on aerial surfaces of the plant. They are thought to protect plants from adverse conditions, including UV light and herbivore attack, and are an important source of a number of phytochemicals. Experimental studies have revealed an intricate and extensive network of genes underlying the regulation of trichome development in the model plant *Arabidopsis thaliana*. The groups of proteins – and their interactions and regulations involved – have been found across different developmental processes and plant species. This makes the formation of trichomes an excellent model system for studying cell differentiation, development and pattern formation in plants. Trichome patterns are generated *de novo*, meaning they are independent of cell lineage or signalling from underlying cell layers, thus a dynamic self-regulatory mechanism is required. To elucidate how self-regulatory mechanisms operate in trichome patterning, studies on mutant phenotypes have been instrumental. This is one of the advantages of the trichome system: genetic perturbations often result in observable patterning defects, which has revealed a wide array of mutant phenotypes in *Arabidopsis*. Intriguingly, many of these phenotypes are contradictory to current understanding or are hard to explain merely by intuition and the currently available data. The use of mathematical modelling approaches has greatly aided the elucidation of mechanisms that form the core of trichome patterning. In this thesis, we have treated several models which have been formulated based on experimental observations, with the aim of describing the genetic and molecular basis of trichome patterning.

In **Chapter 1** a general introduction is given on pattern formation in biology and how this is related to trichomes. At the basis of the mathematical theory behind pattern formation are the reaction-diffusion models first developed by Alan Turing in 1952. In this chapter, we have reviewed the basic principles of Turing's theory and have considered extensions to the reaction-diffusion models, namely the activator-inhibitor (AD) and activator depletion (AD) models. These two principles play an important role in trichome patterning, as is evident from the genetic analysis of the trichome gene regulatory network that was summarized in this chapter. Finally, we introduced the

theory behind dealing with uncertainty that is one of the most challenging aspects of developing models on trichome pattern formation.

In **Chapter 2** we consider the use of spectral methods for the purpose of uncertainty quantification, specifically in the context of biological systems. We show how these methods can be used as an alternative to commonly used Monte Carlo (MC) methods. Where MC methods typically rely on a high number of samples, the scheme proposed in this chapter provides a way of choosing these samples in an efficient way. As such, it is especially suitable for cases where models are computationally expensive to solve and where brute-force approaches like MC would be infeasible. We show how the method can be adapted to deal with difficulties often encountered in biological systems, such as high-dimensionality and bifurcations.

The trichome mutant studied in **Chapter 3** shows a complex change in pattern as the result of a point mutation that strongly reduces the interaction between two of the core patterning protein. In this chapter, we have seen how the mechanism and phenotypic consequences of modifying the active complex pool as a consequence of a point-mutation in TTG1 leads to increased trichome cluster frequency and TTG1 depletion. We analyzed weak *ttg1* alleles, particularly *ttg1-9*, which produces few but clustered trichomes (in contrast to the glabrous *ttg1-1* allele). We find that TTG1-9 retains minimal interaction with GL3, which is not sufficiently strong to ensure nuclear retention. *ttg1-9* mutants have no TRY expression and less CPC than wild-type. This suggests a scenario in which a weak interaction of TTG1-GL3 is sufficient to produce some antagonism to GL1-GL3 – based on the observation that CPC expression is reduced compared to wild-type – but not enough that it results in productive TTG1-GL3 mediated activation of TRY. Given that trichome patterning depends on two parallel principles (activator-inhibitor and activator depletion), both these models were tested individually for the consequence of reduced binding between TTG1 and GL3. Whereas in the AI model, reduced binding strength almost never led to the formation of clusters, in the AD model cluster frequency was almost always higher than observed in *ttg1-9*. A combined model (AIAD) produces more reasonable cluster estimates and shows loss of TTG1 depletion as is observed in mutant plants.

The **M**YB-**b**HLH-**W**D (MBW) protein complex drives trichome initiation. In **Chapter 4** we combinded LUMIER binding assays with mathematical models to quantitatively characterize the MBW complex formation. We follow up on the reported competitive binding of TTG1-GL1-GL3 reported in Pesch *et*

al., 2015, and use our newly obtained insights to make predictions on the expected composition of the various possible combinations of the MBW complexes found *in planta*, as well as the expected consequences thereof on pattern formation. A challenge here was to derive a method that would allow the quantification of dissociation constants of insoluble proteins. Through a ratiometric approach where we quantified all interactions relative to the core component of the MBW complex (the bHLH protein, GL3), we show that the activator proteins GL1 and TTG1 have similar binding strengths to each other, and bind stronger than the inhibitor proteins TRY and CPC. Furthermore, we found a model with negative cooperativity between GL1 and TTG1 to be the most parsimonious with the data, defining a possible scenario of the competitive behaviour shown by Pesch *et al.* in more detail.

Both TRY and CPC negatively regulate trichome initiation through binding to GL3 and counteracting GL3-complex formation with other activators. Constitutive overexpression of TRY or CPC results in loss of trichomes; the *try cpc* double mutants show large clusters of trichomes. Although TRY and CPC proteins are 55% identical, *try* single mutants show increased number of clusters but decreased number of trichomes, whereas *cpc* single mutants show no clusters but increased number of trichomes (higher density). In **Chapter 5**, we developed a model that provides a possible explanation for these puzzling phenotypes. The pre-existing hypothesis on these mutant phenotypes was that TRY is acting on short ranges of effect, whereas CPC is acting on longer ranges. While testing this hypothesis with the model in **Chapter 5**, we found that the differing inhibitor mobility could explain the single mutant phenotypes, but that it was not sufficient to reproduce the clustering behaviour observed in the double mutant. This indicated that the mutation effects were not simply additive as previously assumed. We therefore looked to other explanations and developed a model that simulated a simplified form of cell-differentiation, in which patterning genes were down-regulated upon differentiation. This addition led to the observed cluster formation in the double mutant and is predicted to be an important mechanism in explaining this mutant.

In Chapters **6** and **7** we have studied trichome patterning from an evolutionary perspective. By comparing closely related Brassicaceae species, we have seen how the same underlying network can adapt to the varying expression levels observed between *Arabidopsis thaliana*, *Arabis alpina* and *Cardamine hirsuta*. In **Chapter 6** we have used a modelling approach to explain two opposite phenotypes following from the same mutation between *A. thaliana* and

A. alpina. Where overexpression of GL3 in *A. thaliana* leads to an increased number of trichomes, in *A. alpina* the density is decreased. Guided by the observation that GL1 expression levels are lower in the wild-type of *A. alpina* compared to *A. thaliana*, our model predicted that this change would be sufficient to explain the opposite effect of GL3 overexpression. In **Chapter 7** we additionally compared expression levels of patterning genes in *Cardamine hirsuta*. We used a mathematical model to identify which of the possible adaptations in the network may lead to the observed differences between species. Furthermore, a sensitivity analysis predicted that the density of the trichome pattern was most sensitive to the stability of the MBW complexes in all three species.

Finally, **Chapter 8** provides a general discussion of the work in the preceding chapters. Here, we review the challenges we have faced when modelling trichome patterning, novel findings and assumptions are discussed, and concluding remarks are given – specifically for the trichome system, but also in the context of developmental and patterning studies in general – on the relevance of our findings

Samenvatting

Trichomen zijn eencellige haren die ontstaan op de epidermis en een regelmatige distributie vertonen op de uitwendige oppervlakten van de plant. De veronderstelling is dat trichomen planten beschermen tegen ongunstige omstandigheden, zoals blootstelling aan Uv-licht en planteneters, en zijn een belangrijke bron van verscheidene fytochemische substanties. Experimenten hebben een ingewikkeld en uitgebreid netwerk van genen onthuld wat ten grondslag ligt van de regulatie van trichoom ontwikkeling in de model-plant *Arabidopsis*. De groepen eiwitten – en bijbehorende interacties en regulaties daarvan – zijn gevonden in meerdere ontwikkelingsprocessen en plantensoorten. Dit maakt de vorming van patronen van trichomen een uitstekend model systeem voor het onderzoeken van cel differentiatie, ontwikkeling en patroonvorming in planten. Patronen van trichomen worden *de novo* gevormd, wat betekent dat de vorming onafhankelijk is van cellulaire afstamming en signalen van onderliggende cellagen, en er dus een dynamisch, zelfregulerend proces nodig is. Bij de verheldering van hoe dergelijke zelfregulerende processen functioneren in patroon-vorming van trichomen is het bestuderen van gemuteerde fenotypes van groot nut geweest. Dit is een van de voordelen van het trichoom systeem: genetische verstoringen resulteren vaak in een waarneembaar effect in patroon-afwijkingen, wat een breed scala aan gemuteerde fenotypes in Arabidopsis heeft voortgebracht. Veel van deze fenotypes zijn moeilijk uit te leggen aan de hand van uitsluitend intuïtieve interpretaties en de beschikbare data. Het gebruik van wiskundige modellen heeft een grote bijdrage geleverd in het ophelderen van de mechanismen die de kern vormen van trichoom patroonformatie. In deze thesis hebben we verschillende modellen behandeld die geformuleerd zijn op basis van experimentele observaties, met het doel om de genetische en moleculaire basis van trichoom patroonformatie te beschrijven.

Hoofdstuk 1 bestaat uit een algemene introductie op het gebied van patroonformatie in de biologie en hoe dit gerelateerd is aan trichomen. Aan de basis van de wiskundige theorie achter patroonformatie liggen de reactie-diffusie modellen die voor het eerst ontwikkeld zijn door Alan Turing in 1952. In dit hoofdstuk hebben we een overzicht gegeven van de basis principes van

Turing's theorie en hebben we extensies op dit model in beschouwing genomen, namelijk de activator-inhibitor (AI) en activator depletion (AD) modellen. Deze twee principes spelen een belangrijke rol in trichoom patroonvorming, zoals blijkt uit de analyse van regulerende netwerk van genen zoals die is samengevat in dit hoofdstuk. Ten slotte introduceren we de theorie achter hoe we omgaan met onzekerheid, wat een van de grootste uitdagingen is in het ontwikkelen van modellen voor trichoom patroonvorming.

In **Hoofdstuk 2** beschouwen we het gebruik van spectrale methoden ten behoeve van onzekerheids kwantificatie, specifiek in het geval van biologische systemen. We laten zien hoe deze methoden ingezet kunnen worden als een alternatief voor de veelgebruikte Monte Carlo (MC) methoden. Terwijl MC methoden gebaseerd zijn een groot aantal toetsingen, is het schema wat voorgesteld wordt in dit hoofdstuk een manier om deze toetsingen op een efficiënte manier te kiezen. Als zodanig is deze methode vooral geschikt in het geval waarbij het model rekenkundig gezien kostbaar is om op te lossen en waar brute-kracht methoden als MC onhaalbaar zouden zijn. We laten zien hoe de methode kan worden aangepast om om te kunnen gaan met moeilijkheden die vaak voorkomen in het geval van biologische systemen, zoals hoge dimensionaliteit en bifurcaties.

De trichoom mutant die bestudeerd is in **Hoofdstuk 3** laat een complexe verandering in het patroon zien als gevolg van een puntmutatie waarbij de interactie tussen twee relevante eiwitten is verzwakt. In dit hoofdstuk zien we hoe het mechanisme en de fenotypische gevolgen van het veranderen van de actieve eiwitcomplexen – tot stand gebracht door een puntmutatie in TTG1 – leidt tot een toename in clusters van trichomen en TTG1 uitputting. We hebben zwakke *ttg1* allelen onderzocht, met nadruk op *ttg1-9* die minder maar geclusterde trichomen vormt (in tegenstelling tot het kale *ttg1-1* allel). We zien dat TTG1-9 minimale interactie behoudt met GL3, wat niet sterk genoeg is om te verzekeren dat de eiwitten in de nucleus blijven. *ttg1-9* mutanten hebben geen expressie van TRY en minder CPC dan wild-type. Dit suggereert een scenario waarbij de zwakke interactie tussen TTG1-GL3 genoeg is om antagonisme te vertonen met GL1-GL3 – gebaseerd op de waarneming dat CPC expressie verminderd is vergeleken met wild-type – maar niet genoeg dat het resulteert in TTG1-GL3 gereguleerde activatie van TRY. Aangezien trichoom patronen afhankelijk zijn van twee parallelle principes (AI en AD), zijn beide modellen getest voor het gevolg van de verzwakte binding tussen TTG1 en GL3. Terwijl in het AI model de verzwakte binding bijna nooit leidde tot formatie van clusters, was in het AD model de cluster frequentie consistent hoger

dan waargenomen voor *ttg1-9*. Een gecombineerd model (AIAD) produceerde redelijkere cluster schattingen en vertoonde een gebrek aan TTG1 uitputting zoals ook waargenomen voor de gemuteerde planten.

Het MYB-bHLH-WD (MBW) eiwit complex stimuleert de initialisatie van trichomen. In **Hoofdstuk 4** hebben we LUMIER proeven gecombineerd met wiskundige modellen om de formatie van het MBW complex kwantitatief te karakteriseren. We volgen onderzoek over de competitieve binding van TTG1-GL1-GL3 door Pesch *et al.*, 2015, op en maken gebruik van onze nieuwe inzichten door voorspellingen te maken over de verwachte samenstelling van de mogelijke combinaties van MBW complexen in planten. Een uitdaging hierbij was om een methode te ontwikkelen die de kwantificatie van dissociatie constanten van onoplosbare eiwitten mogelijk kon maken. Met behulp van een ratiometrische aanpak waarbij we alle interacties relatief nemen aan de kern van het MBW complex (het bHLH eiwit, GL3), laten we zien dat de activator eiwitten GL1 en TTG1 gelijkwaardige bindingssterkten hebben, en sterker binden dan de remmende eiwitten TRY en CPC. Daarnaast zagen we dat een model met negatieve coöperativiteit het beste overeenkomt met de data, en daarmee in meer detail een mogelijk scenario definieert van de competitieve binding zoals getoond door Pesch *et al*.

TRY en CPC hebben een negatief regulerend effect op initialisatie van trichomen via hun binding met GL3 waarbij ze GL3-complex formatie met activators tegenwerken. Constitutieve over-expressie van TRY of CPC resulteert in afwezigheid van trichomen; de *try cpc* dubbel mutant vertoont grote clusters van trichomen. Hoewel TRY en CPC eiwitten zijn die 55% identiek zijn, vertonen de *try* mutant een toename in clusters en afname in aan aantal trichomes, terwijl de *cpc* mutant geen clusters maar een toename in aantal trichomen vertoont. In **Hoofdstuk 5** hebben we een model ontwikkeld wat een mogelijke verklaring kan geven voor deze raadselachtige fenotypes. De reeds bestaande hypothese over de gemuteerde fenotypes was dat TRY op korte afstanden effectief is en CPC op lange afstanden. Tijdens het testen van deze hypothese met het model in **Hoofdstuk 4** zagen we dat het verschil in mobiliteit van de remmers wel de enkele mutanten kon verklaren, maar niet de dubbel mutant. Dit gaf aan dat de effecten van deze mutaties niet simpelweg additief waren zoals eerder aangenomen. Daarom keken we naar andere verklaringen en ontwikkelden een model dat een gesimplificeerde vorm van cel differentiatie simuleerde waarbij patroonvormende genen minder tot expressie kwamen. Deze toevoeging leidde tot de geobserveerde cluster formatie in de dubbel mutant en de voorspelling is dat dit een belangrijk mechanisme is

in de verklaring van deze mutant.

In **Hoofdstukken 6** en **7** hebben we trichoom patroonvorming bestudeerd vanuit een evolutionair perspectief. Door nauw verwante Brassicaceae soorten te vergelijken hebben we gezien hoe het onderliggende netwerk zich kan aanpassen aan de verscheidene expressie-niveaus die waargenomen zijn voor *Arabidopsis thaliana*, *Arabis alpina* en *Cardamine hirsuta*. In **Hoofdstuk 6** hebben we een modelleerbenadering gebruikt om de twee tegenovergestelde fenotypen as gevolg van dezelfde mutatie tussen *A. thaliana* en *A.alpina* te verklaren. Terwijl over-expressie van GL3 in *A. thaliana* leidt tot een toename in trichoom aantallen, leidt dezeflde mutatie in *A. alpina* tot lagere trichoom aantallen. Geleid door de observatie dat GL1 expressie niveaus in de wild-types van *A. alpina* lager waren dan die van *A. thaliana*, voorspelde ons model dat deze verandering voldoende is om de tegenovergestelde fenotypes van de soorten te verklaren. In **Hoofdstuk 7** hebben we daarnaast ook de expressie niveaus van de patroonvormende genen in *Cardamine hirsuta* bepaald. We hebben een wiskundig model gebruik om te identificeren welke mogelijk adaptaties in het netwerk kunnen leiden tot de waargenomen verschillen tussen de soorten. Bovendien voorspelde een gevoeligheidsanalyse dat de dichtheid van het trichoom patroon het meest gevoelig was voor veranderingen in de stabiliteit van het MBW complex in alle drie de soorten.

Ten slotte verschaft **Hoofdstuk 8** een algemene discussie van het werk uit de voorgaande hoofdstukken. Hier kijken we terug naar de uitdagingen die we zijn tegenkomen bij het modelleren van trichoom patronen, nieuwe bevindingen en assumpties worden bediscussieerd, en afsluitende opmerkingen over de relevantie van onze bevindingen – specifiek voor het trichoom systeem, maar ook in de context van ontwikkelings- en patroon-onderzoek in het algemeen – worden besproken.

List of publications

Divykriti Chopra, Mona Mapar, Lisa Stephan, Maria C. Albani, **Anna Deneer**, George Coupland, Eva-Maria Willing, Swen Schellmann, Korbinian Schneeberger, Christian Fleck, Andrea Schrader, Martin Hülskamp (2019). "Genetic and molecular analysis of trichome development in *Arabis alpina*." *PNAS 116, 24.*

Rachappa Balkunde*, **Anna Deneer***, Hanna Bechtel, Bipei Zhang, Stefanie Herberth, Martina Pesch, Benjamin Jägle, Christian Fleck and Martin Hülskamp (2020). "Identification of the trichome patterning core network using data from weak *ttg1* alleles to constrain the model space." *Cell Reports 33, 108497.*

Anna Deneer, Christian Fleck (2022). "Mathematical Modelling in Plant Synthetic Biology." *Methods in Molecular Biology, 2379.*

Anna Deneer, Christian Fleck, Jaap Molenaar. "Spectral methods for prediction uncertainty quantification in Systems Biology." *Manuscript prepared for submission.*

Anna Deneer*, Bipei Zhang*, Christian Fleck and Martin Hülskamp. "Quantitative identification of the MBW-complex binding behaviour in trichome patterning through a ratiometric approach." *Manuscript prepared for submission.*

Jessica Pietsch*, **Anna Deneer***, Christian Fleck and Martin Hülskamp. "Comparative expression analysis in three Brassicaceae species revealed compensatory changes of the underlying gene regulatory network" *Manuscript prepared for submission.*

** Contributed equally*

Acknowledgements

Although there is only one lonely name on the cover of this thesis, this has not been a lonely journey. It was an endeavour that would never have been possible without the collaboration, support, patience and hard work of a group of people. I feel I have been very fortunate with my particular group of people, where each had different perspectives, tastes and wisdoms, but were all equally kind-hearted and supportive.

Christian, my adventure into the world of patterning started with you. If I look back on my first project under your supervision (Bachelor thesis, 2014) it's clear to see I still had a lot to learn. At that point I remember thinking I would be doomed as a researcher for I would surely make a mess of things. What I didn't realize at that time was how much I could learn from you and how much fun that process would be. Racing against an imaginary clock, racking our brains over conflicting results or amusing mistakes in my text/calculations/code – thinking back to these moments never fails to put a smile on my face. Thank you for all your effort, time, patience and wisdom.

My sincerest gratitude goes out to **Martin** for giving me the opportunity to work on a great project and for making the puzzles behind trichome patterning a joy to work on, in large part thanks to your enthusiasm and creative thinking. It must have been almost weekly that your insights helped make a difficult problem seem simple. I will fondly remember my visits to your group in Cologne where I was always warmly welcomed and got to meet the wonderful people I was lucky enough to work with. Thank you for all your support and kindness.

Jaap, bedankt voor je warme welkom van Emma en mij binnen Biometris. Het was altijd leuk om vooruitgang en voorgaande resultaten met jou te bespreken dankzij je enthousiasme en optimisme, wat ook weer een aanstekelijk effect had op momenten wanneer dat erg welkom was. Ik heb ook het geluk gehad om samen met jou aan het PCE project te mogen werken, iets wat met mede dankzij jouw hulp en heldere uitleg iets moois bleek te zijn in plaats van iets om verward over te zijn.

Vitor, thank you for giving me the opportunity to start my PhD at SSB. It was a joy to work in your group and thank you for fulfilling the important role of being my promotor.

Peter, bedankt dat je me ook tijdens het laatste deel van mijn PhD nog steunde en goed in de gaten hield of ik niet gek werd van thuis werken en schrijven.

Rik en **Bob** bedankt dat jullie mijn paranimfen willen zijn. Kantoren delen met jullie, naar Aussios gaan, serieuze en minder serieuze gesprekken met een kop koffie erbij, was altijd leuk (zelfs wanneer ik vaak moest zuchten om je verhalen, Bob, of honger kreeg van jouw verhalen, Rik).

Emma and **Rob**, it's been great sharing an office with you. When I started my PhD I looked at you two to see how it's done. Emma, ik hoop dat je het in Nijmegen naar je zin hebt, ik twijfel er niet aan dat je daar geweldig werk zal leveren. Bedankt voor de gezelligheid en voor het delen van je liefde van o.a. katten en fietsen, ik hoop dat ik ooit zo snel zal zijn als jij op de fiets (waarschijnlijk niet...). Rob, starting out at SSB was made easier by you, thanks to all your efforts in organization and guidance. It's a long time ago, but I still remember you had answers to every question (even the poorly formulated ones) when you were supervising my master thesis, thank you for all your help and support.

Special thanks to my collaborators in Cologne, **Jessica, Hanna, Bipei** and **Mona**. It was great to see how fast you all picked up the abstract notions behind the models and how you helped to solve the puzzles. I am very grateful for your support in helping me understand the experiments behind the patterns and for your hard work.

Thanks to the **CyGenTig** team for giving me the opportunity to work on the interesting optogenetics projects and for inviting me to the wonderful cities of Düsseldorf and Edinburgh, exploring the different labs and cities with the group has been amazing.

I would like to thank my colleagues at **Biometris** and **SSB**. Both departments were great workplaces, in large part thanks to the great people each department consisted of.

Finally, I would like to thank my family and friends who supported me throughout my PhD. **Ans** en **Lei**, bedanktj dat ich altied thoes kos kommen bie uch in 't zuiden. **Ruben** en **Imre**, al sinds ik jong ben zijn jullie mijn grote voorbeelden, **Rachna, Djur, Géza** en **Alek** maken dat plaatje compleet.

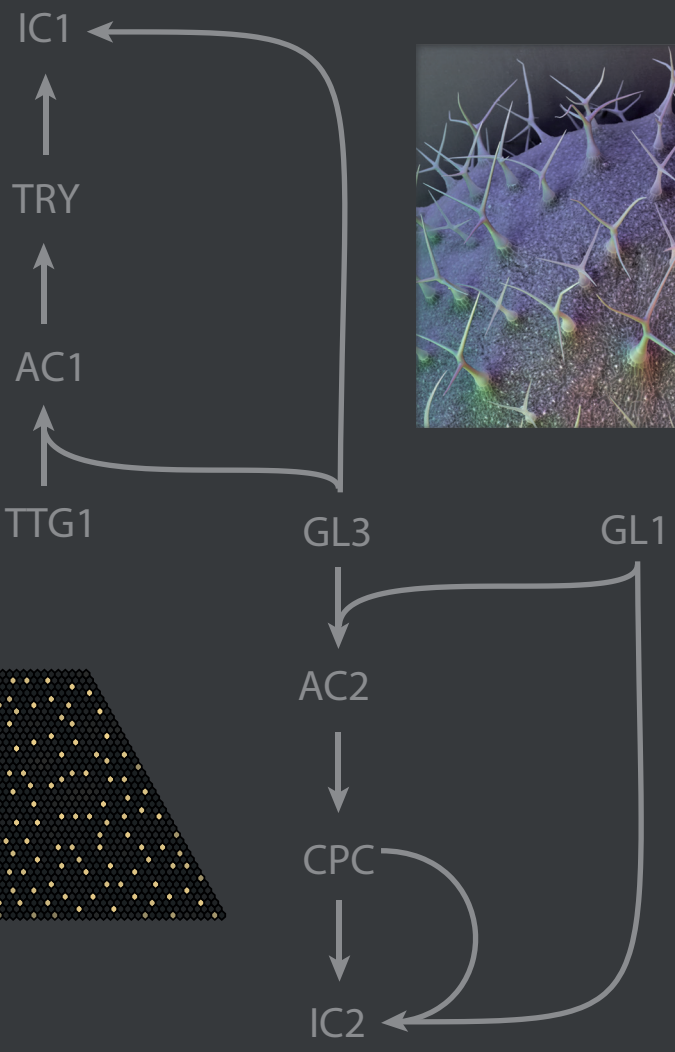
Carlo, maat, onze eeuwenoude vriendschap heeft ook mijn PhD tijd overleefd. Zolang wij niet teveel doolhoven samen proberen op te lossen denk ik dat we alles overleven, zelfs **Brian**, die ik wil bedanken voor zijn vakkundige hulp met schrijven en de gezellige potjes AoM. **Tessa** en **Astrid**, jullie maken mij al slimmer sinds de middelbare school, bedankt dat jullie altijd zo lief zijn voor mij. **Nikolas**, thanks for carrying me to diamond rank in OW, those games and your friendship meant a lot to me. **Iris**, helaas heb ik je niet weer

kunnen volgen naar hetzelfde appartementen-complex, maar gelukkig heb ik toch nog kunnen genieten van jouw wijsheden, handigheid en gedrevenheid – je studenten hebben geluk met jou.

Lieve **Jona**, bedankt voor alle memes en je voortdurende steun. Je wist altijd het goede te zeggen als ik me druk maakte om niks (based). Ook jouw lieve familie bedankt voor alle steun en liefde, **Jan, Marie-Joze, Oma, Opa, Anja, Karo, Elsbert** en **Koko**.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover images supplied by Martin Hülskamp
Cover design by Anna Deneer
Printed by ProefschriftMaken || www.proefschriftmaken.nl


IC1
TRY
AC1
TTG1
GL3
GL1
AC2
CPC
IC2