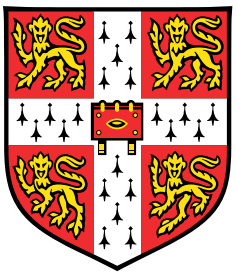
Interpreting noise in stem cell regulation in the shoot apical meristem of *Arabidopsis thaliana*



Henrik Åhl

Supervisors: Dr José Teles
Professor Henrik Jönsson

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

This dissertation is submitted for the degree of
Master of Philosophy in Computational Biology

Hughes Hall

August 2017

The stones we have thrown I hear
fall, glass-clear through the years. In the valley
the confused actions of the moment
fly howling from tree-top
to tree-top, quieting
in air thinner than now's, gliding
like swallows from mountain-top
to mountain-top till they
reach the furthest plateaus
along the edge of being. Where
all our deeds fall
glass-clear
to no end
but ourselves.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Henrik Åhl
August 2017

Acknowledgements

José, Henrik, Jérémy, Max, Julia, Yassin, Klara + Jönsson group

Abstract

What-Why-How-Results-Significance

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

AT	*Arabidopsis thaliana*
CLV3	CLAVATA-3
CZ	Central Zone. The region harboring stem cells in the SAM.
CZ	Central zone
GRN	Gene Regulatory Network
KAN1	KANADI-1
L1	Layer-1. The outermost cell layer of the SAM
RAM	Root Apical Meristem
SAM	Shoot Apical Meristem
SLCU	Sainsbury Laboratory at the University of Cambridge
WUS	WUSCHEL

Chapter 1

Introduction

1.1 The Shoot Apical Meristem of *Arabidopsis thaliana*

Plant stem cells are governed by two developing centra – the Shoot Apical Meristem (SAM) and the Root Apical Meristem (RAM). The SAM is the region responsible for development of all aerial organs in the plant, which includes aspects of cell proliferation and specification, as well as an ability of the plant to maintain and regulate the stem cell identity of the cells at the very apex of the shoot. **ADD CITATION**As opposed to the RAM, which has two stem cell pools in the inside of the root, the SAM maintains a single stem cell pool centered at the apex. It also lacks the root's cap, which protects the stem cells on the inside of the root, whereas these in the shoot are directly exposed to the plant's surroundings.

The stem cells at the SAM contribute to the construction of new organs and general tissue by dividing frequently at the top and subsequently being mechanically pushed out of the center in order to differentiate. The steady maintenance of the stem cell niche allows for a constant production and supply of cells that the plant utilises during both growth and repair of damaged tissue. **ADD CITATION**

In a simple outline of the SAM, it can be said to consist of three core regions: 1. The *central zone* (CZ), which harbors the aerial *stem cell niche* of the plant; 2. The *Rib Meristem*, which is located beneath the CZ and consists of the cells constructing the stem of the plant; 3. The *Peripheral Zone* (PZ), where cells form organs and new tissue through differentiation. In addition to these regions, the SAM is also often separated into the different layers of the dermis, denoted *L1* for the epidermal layer, *L2* for the sub-epidermal one, and *L3* for the inner ground and vascular tissues. For cells in both L1 and L2, proliferation happens orthogonally to the shoot surface, i.e. so that cell lineages are preserved within L1 or L2 correspondingly. In contrast, this is not the case for L3, where cells can divide in all directions. In addition, it has been shown that the epidermis is involved in both promoting and restricting shoot development, adding to the notion of coordination and regulation between the different cell layers in order to accurately direct plant growth.

TODO: Lead into modelling

1.2 Modelling Biological Systems

Due to the interaction of molecules in various ways, e.g. the activation or repression of transcription by certain proteins, organismal development can be considered in the framework of being a *complex system*. In a *systems biology* setting, molecular and mechanical interactions are treated as abstract entities, each representing some fundamental part of the whole system in question, much like how machinery can be explained by its separate cogs and gears working together. In a molecular setting, the typical descriptive approach is through *Gene Regulatory Networks* (GRNs), where each component represents some molecular aspect of the system that is involved in producing expression levels of mRNA, proteins and hormones. **ADD CITATION**

GRNs are commonly understood both through analytical and computational means, where in the latter computer-generated models provides as a modern tool for better understand the complex nature of many biological systems. Typically, reaction kinetics are modelled using various types of *Ordinary Differential Equations* (ODEs). However, due to the large supplies of computer power available in the modern day, many recent studies also utilise more computationally demanding resources such as *Stochastic Differential Equations* (SDEs), where also the inherently random nature of molecular motions, interactions, and processes are accounted for. In particular stochastic modelling of biological systems able to capture dynamical features that deterministic versions cannot. For example, cells often require in various ways to be able to switch between active and inactive states, e.g. when commiting to producing a certain protein or not. Utilising the inherent noise of microscopic systems, cells have been shown to probabilistically tune their responses; **ADD CITATION**modelling these types of phenomena using stochastic approaches has gained much insight into the decision-making process of cells both specifically and in general. **ADD CITATION**

In addition to stochasticity, the increase in computability has also allowed for the development of spatiotemporal modelling, where models are evaluated not only in a static context, but also in a changing setting. A straightforward example taken directly out of the context of plant development is how the distribution of gene expression varies during plant growth and organ formation, both spatially and over time. Typical modelling aspects at such a problem case involves the formulation of which genes and molecules are imporant for the problem of interest, as well as how the discretisation and representation of spatial elements is done. **ADD CITATION**

Computer models in general have two separate aims: exploration and verification. In the former case, computer simulations can be the core for designing experimental experiments, where observed theoretical phenomena can be experimentally tested. An example of this is the classic example of the *Repressilator* **ADD CITATION**, where researchers set up a model framework for how oscillations could occur due to cyclic repressive interactions between three genes. This was then verified to occur by synthetically implementing the system in a bacterium, showing how gene expression profiles could oscillate due to the system motif constructed.

In the latter case, which is the more prevalent in modern computational biology, computer models are established in order to verify or support the potential of an hypothesis due to experimental

observations. ADD CITATIONREWRITE

1.3 Regulatory Mechanics of Plant Stem Cells

1.3.1 Molecular tuning determines cell phenotype

The ultimate phenotype of a cell is to a large degree determined by the underlying expressed genes and proteins, which in turn are regulated by the core GRN. Cells that have not yet undergone the differentiation process are those which are broadly described as stem cells. In addition to not having a specialised phenotype, stem cells continuously proliferate in order to give rise to new cells that can be used for development or repair. ADD CITATIONSimilar to in animals, stem cells require an intricate network both specifying the pluripotency to the cell, and being able to maintain this both when the plant conformation or the environment changes. Effectively, this regulation causes the stem cell niches of the plant to be determined by various types of patterning, which also plays a role in specifying zones of initiation of primordia. A viable and robust network maintaining patterning is thus important for the plant in order undergo phyllotaxis in a functional manner, and to know when and how to commit to more structural changes. Patterning can consist of several types of molecular spatiotemporal expression, including hormones, proteins and RNA localisation, although also patterns of stress and strain have in recent studies been shown to play a role in determining both growth and cell identity. ADD CITATION(TWICE) Typically, whenever gene expression is the focus of a study, it is often used as a proxy for protein expression, as fluorescent tagging and tracking of proteins sometimes interfere with the function or transport of the molecule.

1.3.2 Role of the WUSCHEL-CLAVATA feedback loop

The GRN in the SAM is determined mainly by two core genes – *WUSCHEL* (WUS) and *CLAVATA* (CLV). Their corresponding network consists of the homeodomain protein WUS and a ligand-receptor complex made up by CLV1 (receptor), CLV3 (ligand) and an assumed accessory protein CLV2. In particular CLV3, which is expressed at the very apex and is known to diffuse out of the CZ, correlates strongly with stem cell identity of the cells. ADD FIGUREWUS, in turn, is expressed in the RM and activates the CLV pathway, making it necessary for maintaining the correct stem cell niche. In contrast, CLV3 has an antagonistic impact on WUS, repressing its expression pattern at the very apex. Together, the CLV3-WUS feedback interaction forms the core of the GRN regulating cell identity in the niche. In the PZ however, differentiating genes such as KAN1 ADD CITATIONare known to promote cell specification. In extension, also other homeobox genes such as *Shoot Meristemless* (STM) have roles in the growth network, but are as of today unknown for their role in specifying pluripotency. ADD CITATION

REWRITECLV3 / WUS mutants

HORMONES
STM / Cytokinin / Auxin

How much of the details in GRN is known?

1.4 Lead into what I've done (TODO)

An important question in developmenal biology is how organisms can have robust development despite consisting of many independetly variable parts. **ADD CITATION**(x3)

Chapter 2

Methodology

2.1 Data consists of *in vivo* confocal timelapses

Six plants, labelled *plant 1*, *2*,*4*,*13*,*15*, and *18*, were grown on a solution consisting of $10\mu M$ auxin transport inhibitor NPA to a depth of roughly 1 cm for 22-26 days. The inhibition of auxin prevents formation of new primordia, and this gives rise to a small and naked, organ-free meristem which is tractable for imaging.

The plantlets were marked with pUBQ10::acyl-YFP, which localises in the cell membrane **ADD CITATION**, as well as with pCLV3::dsRED-N7, which was used as a nuclear tracker for CLV3 mRNA expression. Also pPin1::PIN1-GFP was tracked, but not quantified in this study. In addition, *plant 1* did not express the nuclear marker for CLV3.

Using confocal microscopy, the six plantlings were tracked in intervals of 4 hours up to 76 (*plants 1, 2, 4*) or 84 hours (*plants 13, 15, 18*), using a 63x/1.0 N.A. water immersion objective. Due to the high resolution of the images, the acquisition of each z-stack took $\sim$10 minutes, which induced vertical stretching in the images due to stem elongation. Because of this, a second batch of z-stacks was acquired, using low-resolution imaging over $\sim$10 seconds. The original images were then corrected, using this second batch as reference. **ADD CITATION**

2.2 Image Pre-processing and Segmentation

In order to eliminate segmentation errors, the ImageJ plugin StackReg was used to perform a translation transformation for each stack. Individual slices which contained horizontal shifts because of vibrations or other types of system disturbances were identified and replaced with the nearest slice that contained no such shift. The z-directional stretching due to stem elongation was corrected for by mapping the low-resolution stacks to the high-resolution ones in order to attain stretching factors that the images were thereafter corrected for.

Add figure of segmented

For the membrane channel, noise removal was done by Gaussian and alternative-sequential filtering. The filtered z-stacks were then watershed in 3D using the algorithm implemented in the

segmentation software *MARS-ALT*. Segmentation and tracking was thereafter done using the same software. Cellular volumes were from this then calculated as the sum of voxel volumes belonging to the same cell. The tracking, also performed using *MARS-ALT*, was assessed for quality using an F1 score between the parent and corresponding daughter cell. For all analyses discussed in this report, a cutoff value of 0.30 was set for the tracking in order to account for likely incorrect mappings. These cells are included in the overall analysis, but excluded from all cell line related investigations.

A longer outline of this is presented in section A.3.

The nuclear data were deconvolved to account for the microscope's point-spread function using the *PSF distiller* tool from Huygens software 15.05 **ADD CITATION**. As in the membrane case, the nuclear channels were adjusted with the corresponding stretching factors and thereafter segmented using segmentation tool *Costanza* **ADD CITATION**.

In order to link nuclei to membranes, the coordinates of each variable were fitted using a least square approach. Duplicately mapping nuclei were then consolidated as described in section 2.3. Note that whenever spatial coordinates are referenced in relation to membrane or nuclei, they are done so as the centroid coordinates of the basin of attraction found in the segmentation process.

Measures of distances to the top were done in multiple ways. The three methods herein considered consist of a definition of the top based on 1) the spatial coordinates, 2) the expression value, and 3) a least-square fit of a paraboloid to raw meristem images. In the case of spatial coordinates, the average $x-y$ coordinates of n nuclei were chosen, complemented with the highest z value registered in the corresponding timeframe. For the second case, the apex was defined as the n average spatial coordinates of the highest expressing CLV3 nuclei. Lastly, the paraboloid fit to the meristem was used to define the apex by taking the coordinates of the region have a zero-valued derivative. For both the segmentation-dependent approaches, the data was set to exclude subepidermal layers in order to prevent biases.

ershooting?

In order to achieve a cell-resolution description of distances in the SAM, an auxilliary measure of cell distances was used in the form of a cell-wise grouping. In the cell value utilising definitions of the apex above, cells included in the definition was set to have a cell-wise distance of 0. The neighbours of these cells were in turn defined to have a distance value of 1, and so on recursively.

2.3 Data Filtering

Due to thresholding effects for cell nuclei during the segmentation, some individual nuclei are occasionally identified as two or more. In order to accound for this, nuclei were mapped to the corresponding membranes using a minimum euclidian distance measure between the respective centroids. The nuclear quantified metrics were then corrected using the functions found in table 2.1. In addition to this, all mentions of numbers of nuclei are with respect to the number of cell membranes containing at least one nuclear volume identified within them.

For the data analysis section, data was excluded due to apparent segmentation errors. This was done for each plant in isolation, with the outline of the filtering described below in table 2.2. The

Table 2.1 Consolidation methods applied for duplicate nuclei.

Metric	Summary function
Coordinates (x, y, z)	mean
Nuclear volume	sum
Nuclear expression	mean

Parameter	Value
Maximal membrane volume	$\mu + 3\sigma$
Minimal membrane volume	0
Maximal nuclear volume	$\mu + 5\sigma$
Minimal nuclear volume	0
Maximal apical distance	$\mu + 3\sigma$
Maximal neighbour distance	7

Table 2.2 Filtering settings

choice of allowed deviance was done based on the distribution shape , with particular consideration taken to the nuclear and membrane volumes, where no lower boundary was set. The maximal neighbour distance was chosen due to the typical lack of data for cells more than 7 cell distances from the apex. Lastly, division events where loss of nuclear signal took place, we filter out expression values for the parental cell in the cases where $expression_p < 0.7(expression_{c1} + expression_{c2})/2$.

figure of these?

For reference, a quality assessment of the data can be found in section A.3. In addition, descriptions of the software used in segmentation and tracking are briefly introduced in **??**.

2.4 Models of Gene Regulatory Networks

2.4.1 Mathematical Formulation of Biochemical Reactions

Mass-action Kinetics

The formulation of processes in GRNs focus primarily on two aspects: synthesis and degradation of matter, which usually takes the form of molecular concentrations or absolute abundance. As implied in **??** (modelling biological systems), we here work using an ODE or SDE description of our regulatory systems.

We in this thesis represent our molecular reactions using mass-action interactions and Michaelis-Menten kinetics; both here relying on the naïve assumption that our primary reagents here act in isolation of other possibly intervening molecules. As part of our formalism, we write

$$S \underset{k_f}{\overset{k_b}{\rightleftharpoons}} P \tag{2.1}$$

to express that some substrate S is turned into a product P by some given *forward affinity* k_f. Likewise, as the reaction is *reversible*, the product P is transformed back into S with the *backward affinity* k_b.

The *law of mass action* states that the rate of a reaction is proportional to its affinity, e.g. k_f, and the concentration of the reacting species, here S. The reaction rate of the production of P would thus be $r_f = k_f S$. However, the rate of change of reactant P also depends on the backward affinity, which would give the overall rate-of-change for P as

$$\Delta P = k_f S - k_b P. \tag{2.2}$$

In the infinitesimal limit, we analogously have

$$\frac{\mathrm{d}P}{\mathrm{d}t} = k_f S - k_b P, \tag{2.3}$$

i.e. on the form of a differential equation, which will be the baseline for our formulations. Similar to the formulation of rate-of-change of P, we can do the same for species S, and our system is then fully represented as a system of differential equations.

Expanding on this, we can easily solve for the steady-state concentrations of the system by assming that all rates average to zero. In our example above, this gives us

$$\frac{k_f}{k_b} = \frac{P}{S}, \tag{2.4}$$

which holds in general, regardless of the number of reacting species.

Michaelis-Menten Kinetics

A conceptual drawback of the mass-kinetics formulation is the possibility to have infinite reaction rates, whereas the molecular reactions in nature typically are restricted by some means. One way to account for this is through *Michaelis-Menten kinetics*, which describes enzymatic chemical reactions. In these, the trivial example introduced in eq. (2.1) is expanded to include an enzymatic agent such that

$$E + S \underset{k_f}{\overset{k_b}{\rightleftharpoons}} ES \xrightarrow{k_p} E + P. \tag{2.5}$$

In other words, an enzyme-like molecule binds to the substrate S such that the complex ES is formed. This complex is thereafter transformed into the product molecule P and again the enzyme E. Assuming a total enzyme concentration of E_{tot} and the assumption that the enzyme-substrate binding process is in equilibrium, the rate-of-change of the product can be rephrased to be on the form

$$\frac{\mathrm{d}P}{\mathrm{d}t} = V_{max}\frac{S}{K+S}, \tag{2.6}$$

where $K = K_f/K_b$ and $V_{max} = k_p E_{tot}$. This expression is said to be on *Michaelis-Menten* form, where V_{max} is the maximal activation rate of the protein, and K can be thought of as a saturation coefficient.

Extrapolating on this type of reaction, introducing n enzymatically acting molecules instad gives the standard *Hill equation* form, namely

$$\frac{\mathrm{d}P}{\mathrm{d}t} = V_{max}\frac{S^n}{K^n + S^n}, \quad \text{and} \tag{2.7}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = V_{max}\frac{K^n}{K^n + S^n} \tag{2.8}$$

for an activating and repressing reaction respectively.

2.4.2 Numerically solving stochastic systems

Gillespie Algorithm

The Gillespie algorithm is a discrete approach for simulating stochastic molecular dynamics. It first appeared in print by Dan Gillespie in 1977, and has since been widely used for stochastic simulations in multiple fields.

While being computationally expensive, the Gillespie algorithm compensates for its lack in tractability by producing a statistically exact trace of the molecular dynamics of a system.

The algorithm originates in the formulation of the *chemical master equation*, which specifies the rate of change of the transition probabilty between states in the form of

$$\frac{\partial P(x,t|x_0,t_0)}{\partial t} = \sum_{j=1}^{M} \left[a_j(x-v_j)\,P(x-v_j,t|x_0,t_0) - a_j(x)P(x,t|x_0,t_0)\right] \tag{2.9}$$

where a defines to reaction probability, or propensity, for each type of reaction, and v the stoichiometry, i.e. information of how the molecular species are changed due to the reaction. $P(x,t|x_0,t_0)$ on its own denotes the probability of $X(t) = x$, given that the initial value is x_0. Solving the master equation analytically is usually complicated, so simulating a complex biological system using the Gillespie approach can often be far more tractable.

Proceduraly, the algorithm can be formulated in four steps:

Initialisation Generation of number of molecules and reaction parameters.

Randomisation Generation of random numbers to determine 1) next interaction, and 2) the time increment.

Update of system Time and molecular numbers are update correspondingly to the determined event in step 2.

Repetition Step 2-4 are repeated until some stop condition is met.

In principle, the Gillespie algorithm is interested in two fundamental questions: 1) When does the next reaction happen? 2) Which is the next reaction? The time until the next reaction at time t

is denoted τ and can be shown to be an exponential distribution centered at $1/\sum_{j=1}^{M} a_j(x)$ for some molecular concentration x, i.e.

$$p(\tau = t') = \sum_{j=1}^{M} a_j(x) e^{-\sum_{j=1}^{M} a_j(x) t'} \tag{2.10}$$

with the reaction probability instead being described by the normalised propensity. Historically, due to the limitation of random number generators, the time update has been described as being drawn from

$$\tau = \frac{1}{\sum_{j=1}^{M} a_j(x)} \ln \frac{1}{r_1} \tag{2.11}$$

although modern high-level programming languages do in some cases perform better using the inherent random number generator for a specific type of distribution. **ADD CITATION**

Milstein's Method

If there is no requirement for exactness, less computationally intense alternatives to Gillespie's algorithm exists. One such example is the *Langevin* formulation of chemical systems, which utilises SDEs to attain an approximate solution to the system trajectory, and is particularly useful when the number of molecular reagents is high.

The Langevin formulation, like Gillespie's, utilises the chemical master equation to compute the behaviour of the system. In principle, the Langevin formulation can be said to reformulate a deterministic increment of the form

$$X_i(t + \mathrm{d}t) = X_i(t) + \sum_{j=1}^{M} v_{ji} a_j(X(t))\, \mathrm{d}t \tag{2.12}$$

to the stochastic form

$$X_i(t + \mathrm{d}t) = X_i(t) + \sum_{j=1}^{M} v_{ji} a_j(X(t))\, \mathrm{d}t + \sum_{j=1}^{M} v_{ji} a_j^{1/2} N_j(t)\, \mathrm{d}t^{1/2} \tag{2.13}$$

where X denotes the molecular number, v the stochiometric coefficient of the equation in question, and N_j are temporally uncorrelated and statistically independent, Gaussian random numbers with mean 0. From this stage, the equation is then easily extended to its multivariate form, namely

$$X_i(t + \mathrm{d}t) = \sum_{j=1}^{M} v_{ji} a_j(\bar{x})\, \mathrm{d}t + \sum_{j=1}^{M} v_{ji} {a_j}^{1/2}(\bar{x}) N_j(t)\, (\mathrm{d}t)^{1/2} \tag{2.14}$$

Milstein's approach to solving this equation numerically utilises eq. (2.13) on the differential form

$$\mathrm{d}X_t = a(X_t) + b(X_t)\mathrm{d}W_t \tag{2.15}$$

where W_t is a continuous-time stochastic process. The simulation interval $[t_0, T]$is then partitioned into parts of size $\Delta t = T/N$, where N is the number of partitions. We thereafter define the update

$$Y_{n+1} = Y_n + a(Y_n)\Delta t + b(Y_n)\Delta W_n + \frac{1}{2}b(Y_n)b'(Y_n)\left((\Delta W_n)^2 - \Delta t\right) \tag{2.16}$$

$$Y_{n+1} = Y_n + a(Y_n)\Delta t + b(Y_n)\Delta W_n + \frac{1}{2}b(Y_n)b'(Y_n)(\Delta W_n)^2 \tag{2.17}$$

on Itō and Stratonovich form respectively. Here b' denotes the spatial derivative of b, whereas $\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$. The difference between the Itō and Stratonovich form in turn is the interpretation of the integral of $\mathrm{d}W_t$. **ADD CITATION**In many cases, it is preferable to express the numerical update on a derivative-free form, which can be done through a Runge-Kutta like approach **ADD CITATION**and gives the final, multivariate expression as

$$Y_{i,n+1} = Y_{i,n} + a_i(Y_{i,n})\Delta t + b_{ii}(Y_{i,n})\sqrt{\Delta t}N_i + \frac{1}{2\sqrt{\Delta t}}\left[b_{ii}(\bar{x}, n) - b_{ii}\right]\Delta t(N_i)^2 \tag{2.18}$$

where a supporting predictory step is calculated in the form of

$$\bar{x_i} = x_i + a_i(Y_n)\Delta t + b_{ii}\sqrt{\Delta t}. \tag{2.19}$$

Algorithmically, the Milstein approach is of strong order of convergence $\mathscr{O}\left(\sqrt{\Delta t}\right)$ and weak order $\mathscr{O}(\Delta t)$. In this thesis, we utilise the Milstein approach under the Stratonovich interpretation.

Chapter 3

Results

3.1 NPA dilution shows feedback dynamics

When tracking the number of CLV3 nuclei identified by Costanza, the results observable in fig. 3.1 shows the number of observed CLV3 nuclei increasing for plants 2, 4, 13, and 15, corresponding to a visually observable enlargement of the CLV3 domain in **ADD FIGURE**(timelapse of example plant over 4 timepoints?). Similarly, fluctuations in this number can be seen not to correlate as described in table 3.1.

Table 3.1 TODO

Plant	p-value
2	0.48
4	0.84
13	0.91
15	0.11

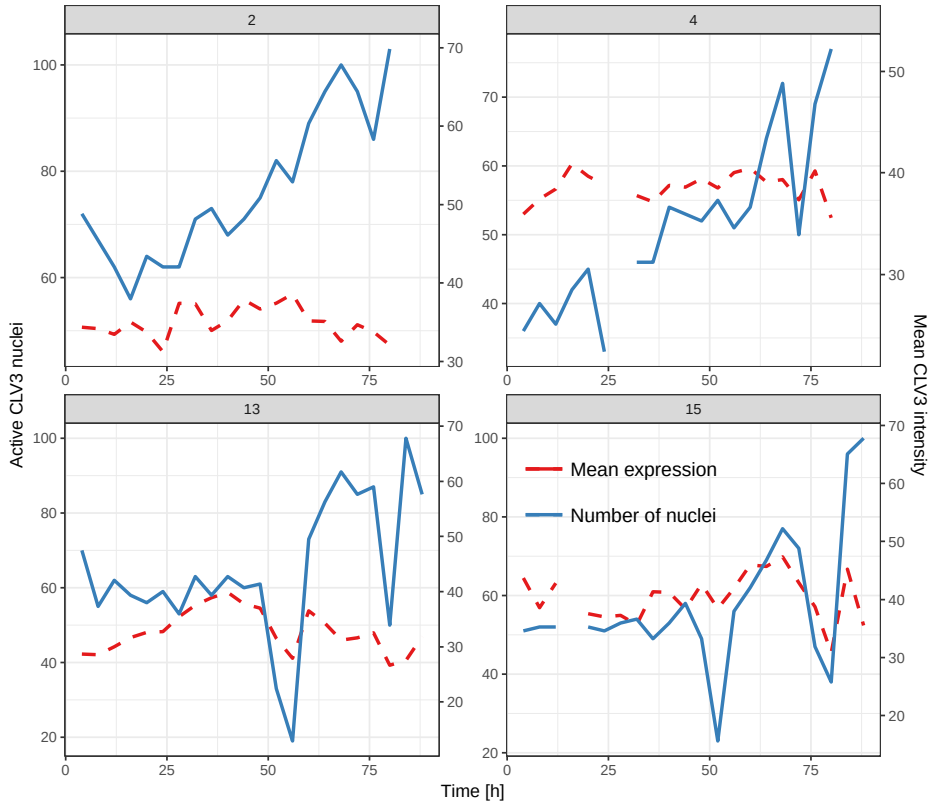


Fig. 3.1 yep

3.2 Epidermal regulation maintains low variance in CZ

3.2.1 Layer-wise separation in the distribution of nuclear volume suggests Epidermal activation

3.2.2 Radial decay in CLV3 expression shows robustness in apical expression

3.2.3 A simple model supports L1 activation hypothesis (hopefully)

3.3 Trajectorial analysis suggests CLV3 decay outside of apex

This goes in line with previous point. What do I want to say about this?

3.4 Longevity thingy (?)

3.5 Extended data analysis. Include this at all?

Chapter 4

Discussion

4.1 This does that. Why?

4.2 Because.

Chapter 5

Outlook

5.1 Model expansion

5.2 Data investigation

5.3 New experiments

References

Appendix A

Addition Information on Data Acquisition

A.1 Experimental procedure

The Yellow Fluorescent Protein (YFP) marker for the plasma membrane was amplified using PCR with primers attb1-mYfwd (5'-AGAAAGCTGGGTTTACTTGTACAGCTCGTCCATGCCGAGAGTG) and attb2-YFPrev (5'-AGAAAGCTGGGTTTACTTGTACAGCTCGTCCATGCCGAGAGTG), with the forward primer sequence containing a motif known to acetylate in plant cells CITEHERE. 50 μL solution was amplified in 96 °C for 1 minute, followed by 25 cycles of 96 °C for 30 seconds, and a final elongation for 30 seconds. 5 μL of the result was then used in a second reaction consisting of 40 μL solution in total, with primers B1 adapt (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCT) and B2 adapt (5'-GGGGACCACTTTGTACAAGAAAGCTGGGT) included. Similar to the first solution, the second mixture was amplified by PCR in 95 °C for 2 minutes, followed by 94 °C for 30 seconds, 48 °C for 30 seconds, and 72 °C for 1 minute, 20 cycles of 94 °C for 30 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute. Finally, elongation took place under 72 °C for 1 minute.
Membrane
Nuclei
Unnused data (PIN)

A.2 Segmentation

A.2.1 Settings and approach

A.3 Errors in data

Segmentation Errors

Tracking errors

Manual tracking in some timepoints...

A.4 Layer Quality Assessment

Appendix B

Extended data analysis

B.1 NPA dilution is revealed by primordia initiation

B.2 Just cram in the rest of the data

Appendix C

Software Descriptions

C.1 Costanza

Costanza (COnfocal STack ANalyZer Application) **ADD CITATION**is an ImageJ plugin for segmenting compartments in the form of cells and extract quantitative data, including intensities. Primarily, Costanza is used to segment nuclei marked cells in three dimensions and effectively extract information relating to the intensity of the used GFP markers.

The software utilises a steepest gradient ascent approach for segmentation, which initiates at each voxel in the stack and attempts to find local intensity maxima by ascension in the neighbourhood of the voxel. All paths leading to the maximum are then grouped and recorded as a Basin Of Attraction (BOA), i.e. cell. In the event that multiple neighbourhood voxels have a higher intensity than the current one, the path with the highest ratio of intensity difference over spatial difference is chosen, i.e. the path returning $max\left(\frac{\Delta I}{\Delta \vec{x}}\right)$.

Costanza performs preprocessing in the form of intensity inversion, background extraction and applied denoising filters. For postprocessing, it allows for Basin-Of-Attraction (BOA) removal and merging, in order to exclude misidentified cells in the background and avoid faulty separation of individual cells.

C.2 MARS-ALT

MARS-ALT is software developed for spatiotemportal tissue reconstruction and lineaging at cell resolution. The software consist of the two parts MARS (Multi-angle image Acquisition, 3D Reconstruction and cell Segmentation), and ALT (Automated Linage Tracking). It works by importing fluorescently stained confocal images, and (optionally) correcting these using low-resolution reference stacks.

For cell segmentation, MARS denoises the image using an alternate sequential filter in order to increase the signal / noise ratio. Seeds are then extracted by computing voxel minima, and merging those that have a lower largest valley between them than some defined value. The background is

then extracted from the largest connected component found in the image after thresholding. Cells are thereafter watershed using the seeds as references, with cell volumes less than some specified value are filtered. Markers are subsequently removed from the seeds, and the watershed algorithm is repeated until convergence. **ADD CITATION**

C.3 Organism

C.3.1 Main software

Organism is C++ software for simulating biological systems, in particular with multiple compartments primarily in the form of cells. It includes both biochemical and mechanical rules, including rules for proliferation, for numerically simulating the system. Organism relies on a representation of compartments such that all of them can be described by a fixed number of variables and parameters, such as spheres, cylinders and similar. It can also be used for simulations of natural tissues, extracted from confocal microscopy images. As part of the Organism toolset, software for visually inspecting simulated tissue from output data is accessible (Newman), as well tools for performing parameter optimisation of GRNs. **ADD CITATION**

For this thesis, an R wrapper for interacting with parts of the Organism software was constructed, allowing for R-based solver, model, parameter, and initialisation file setup. Code relating to this is hosted at the *Sainsbury Laboratory at the University of Cambridge* GitLab portal, and can be accessed at https://gitlab.com/slcu/teamHJ/henrik_aahl/organism_wrappeR.

C.4 extractoR

Should be tidied up first.