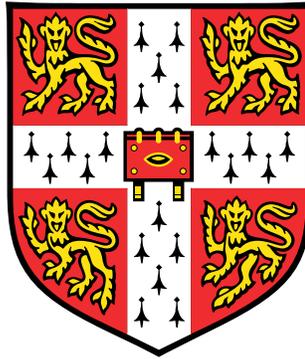


Engineering morphogenesis of *Marchantia*
polymorpha gemmae



Mihails Delmans

Department of Plant Sciences
University of Cambridge

*This thesis is submitted for the degree of
Doctor of Philosophy*

Peterhouse
December 2018

Summary

Morphogenesis is an apparent yet complex process: emergence of a plant shape is the result of an intricate interplay between genetic regulation, cell physiology and mechanical processes at the tissue scale. Morphogenesis spans three levels of biological organisation, forming a nested complex system. Genes, which form the lowest level of the system, are arranged into networks that control properties of cells. Cells, which form the middle level of the system, are arranged into geometric networks, and their mechanical and chemical interactions give rise to the morphology of a whole organism. Therefore, the study of plant morphogenesis relies on understanding how genetically-driven cellular interactions influence the formation of a plant shape.

Clonal propagules of *Marchantia polymorpha* (Marchantia), also known as gemmae, are an attractive system for studying these interactions. Gemmae are small, have a simple disk-like shape and are resilient to environmental conditions. As such, they are well-suited for fluorescent microscopy and the collection of gene expression data. These features create an opportunity to study the processes of gemma morphogenesis at tissue, cellular and genetic levels through fluorescent microscopy in a single assay.

In order to enable the engineering of morphogenesis in Marchantia gemmae, tools and frameworks for obtaining, storing and analysing the observations from the three levels of biological organisation should be put into place. The work presented in this thesis focuses on the development of such tools and their application in studying the role of phytohormone auxin in Marchantia development.

I introduced novel sample preparation and time-lapse imaging assays for Marchantia gemmae along with image-processing methods for the estimation of tissue expansion rates with sub-cellular resolution. These methods allowed me to hypothesise a mechanism behind the regulation of cell proliferation in Marchantia and the role auxin plays in controlling this process.

Together with Bernardo Pollak, I developed MarpoDB, a gene-centric representation of the Marchantia genome that enables the search and preparation of Marchantia genetic parts for assembly into synthetic DNA constructs.

I used MarpoDB to extract parts and build fluorescent reporters, providing proxies for auxin biosynthesis and signalling. The reporters were then introduced into the Marchantia genome, and the gemmae of the transgenic lines were imaged to estimate the average patterns of the reporters' expression.

The collected patterns were then overlaid on the patterns of relative tissue expansion to validate the proposed role of auxin as an inhibitor of cell proliferation and the mechanism behind its transport in Marchantia.

Acknowledgements

First of all I would like to thank my supervisor, Jim Haseloff, for giving me the opportunity to journey into the exciting world of academic research and explore scientific ideas bound only by my limited imagination. I would also like to thank the Biotechnology and Biological Sciences Research Council for financing my daily meals and a cosy flat at 33 Cosin Court.

I am thankful that during my time at Haseloff Lab I had the support and friendship of like-minded peers, Tim Rudge, Fernan Federici, Anton Kan, Paul Grant, Lukas Müller and Limin Wang, whose drive for discovery and appreciation of ideas inspired me in my own research. In particular, I would like to thank Bernardo Pollak who, during my PhD years and beyond, has been a collaborator, a dear friend and the host of extravagant barbecue parties, during which the wildest scientific discussions took place. Thanks to the other members of the Haseloff Lab and Department of Plant Sciences, Marta Tomaselli, Alexandra Ting, Susana Sauret-Gueto, Eftychis Frangedakis, Sean Stevenson and Gregory Reeves, for humanly chats, which kept me cognisant that there is life after PhD. Special thanks to Linda Silvestri, who shared with me the secrets of artificial *Marchantia* insemination.

I would like to thank Helena Steiner, David Sweeney, Anja Thieme and Richard Banks for hosting me at Microsoft Research for the duration of two internships and allowing me a glimpse into the corporate side of scientific research.

Special thanks to Mr Gaius and his grandmother Esmé for keeping me company during six long months of thesis corrections.

Thanks to my parents, who gave me support and advice through the hard times; to my brother Andrejs, who apart from moral support offered late-night lectures on statistics and data analysis.

I would like to thank my dearest Kate, whose care and love kept my thoughts away from the horrors of writing up, and who helped translate this thesis from my version of English to one that a regular person can understand.

Finally, I would like to thank my examiners, Timothy O'Leary and Lionel Dupuy, for their sincere interest in the work and constructive feedback during my viva.

Contents

1	Introduction	15
1.1	Plant growth and engineering	15
1.2	Hierarchical view of morphogenesis	16
1.2.1	Mathematical notations	16
1.2.2	Tissue level	16
1.2.3	Cell level	18
1.2.4	Gene level	21
1.2.5	Cell communication and auxin transport	23
1.2.6	Summary	24
1.3	Tools	25
1.3.1	Synthetic biology	25
1.3.2	Genomic resources	26
1.3.3	Fluorescence microscopy and image analysis	27
1.3.4	Computational models of plant development	28
1.3.5	Marchantia as a model organism	29
1.4	Aims	29
2	Materials and Methods	33
2.1	Marchantia methods	33
2.1.1	Chassis	33
2.1.2	Tissue culture	33
2.1.3	Production of spores	34
2.1.4	Agrobacterium-mediated transformation	34
2.1.5	Surgical assay	35
2.2	Construct Design	35
2.2.1	Domestication of parts amplified from the Marchantia genome	36
2.2.2	Assembly of the Level 1 and Level 2 constructs	36

2.3	Microscopy	38
2.3.1	Sample preparation	38
2.3.2	Fluorescent stereo microscopy	39
2.3.3	Laser-scanning confocal microscopy	39
2.4	Image processing	40
2.4.1	Image-processing pipeline	40
2.4.2	Image-processing algorithms	42
2.5	MarpoDB	46
2.6	Simulating auxin transport	47
2.6.1	Generation of the gemma meshTo simulate the geometry of a day 0 gemma, a polygon with an outline governed by the following equation was extruded by 0.1 arbitrary units:	47
2.6.2	Numerical solution of the reaction–diffusion equation	47
3	Growth dynamics of <i>Marchantia gemmae</i>	51
3.1	Introduction	51
3.2	Image acquisition	53
3.3	Image processing	54
3.3.1	Preprocessing	54
3.3.2	Quantifying tissue expansion	54
3.4	Normalisation and averaging	54
3.5	Discussion	61
4	Maintenance and reestablishment of the apical regions	65
4.1	Introduction	65
4.2	Analysis of the transport model	66
4.3	Growth of the isolated apical and central regions	69
4.4	Role of auxin in cell proliferation	73
4.5	Discussion	74
5	MarpoDB	77
5.1	Introduction	77
5.2	Software architecture	78
5.2.1	Data model	78
5.2.2	Core module	82
5.2.3	Frontend module	82
5.2.4	Server module	84
5.3	Datasets and annotations	84
5.3.1	Genome and transcriptome	84

5.3.2	Annotation pipeline	86
5.4	Applications	88
5.5	Discussion	88
6	Patterns of auxin production and accumulation	91
6.1	Introduction	91
6.2	Reporter design	92
6.2.1	Auxin-biosynthesis reporter	92
6.2.2	Auxin-signalling reporter	94
6.2.3	Transgenic reporter lines	94
6.3	Average patterns of auxin biosynthesis and signalling	98
6.4	Reestablishment of auxin signalling and biosynthesis in the isolated central region	105
6.5	Continuous model of auxin transport	112
6.6	Discussion	115
7	Discussion and Conclusions	117
7.1	Summary of results	117
7.1.1	Genome representation for synthetic biology	117
7.1.2	Methods for recording development of a gemma	118
7.1.3	Role of auxin in gemma development	119
7.2	Future work	120
7.2.1	Model development	120
7.3	General discussion	122
7.4	Conclusion	125
	Appendices	145
A	Marchantia protocols	147
B	Effect of kinetin on Marchantia gemma growth and regeneration	153
C	Continuous model of pressure-driven growth	157

List of Figures

1.1	Expansion at the tissue and cellular level	20
1.2	Regulation of tissue expansion at the gene level	22
1.3	<i>M. polymorpha</i> gemma development.	30
2.1	Assembly schema	36
2.2	Summary of the image-processing pipeline	41
3.1	Image acquisition and preprocessing	52
3.2	Quantifying tissue expansion	55
3.3	Alignment, normalisation and averaging of relative expansion rate patterns	57
3.4	Average velocity and tissue expansion rates	59
3.5	Sample variability of tissue expansion rate	63
3.6	Time variability of tissue expansion rate	64
4.1	Possible mechanisms explaining the maintenance of the apical regions	68
4.2	Surgical assay	70
4.3	Tracking cell divisions in the isolated central regions	71
4.4	Regeneration of the isolated apical regions	72
4.5	Growth of a gemma split along the line connecting the two apical notches	73
4.6	Cellular response to exogenous auxin	75
5.1	MarpoDB software architecture	79
5.2	MarpoDB schema	81
5.3	MarpoDB frontend	83
5.4	MarpoDB population and sequence annotation pipeline	85
5.5	Comparative analysis of Tak and Cam datasets	87
6.1	Design of synthetic parts	95
6.2	Design of auxin signalling and biosynthesis reporters	97

6.3	Auxin-biosynthesis reporter	99
6.4	Auxin-signalling reporter	100
6.5	Induction of the auxin-signalling reporter	101
6.6	Extraction of nuclear signal	103
6.7	Average patterns of estimated auxin biosynthesis and signalling	104
6.8	Correlation between growth parameters	105
6.9	Average auxin biosynthesis reporter signal per sample	106
6.10	Average auxin biosynthesis reporter signal per time frame.	107
6.11	Average auxin signalling reporter signal per sample	108
6.12	Average auxin signalling reporter signal per time frame.	109
6.13	Reestablishment of the auxin-signalling pattern	110
6.14	Reestablishment of the auxin-biosynthesis pattern	111
6.15	Simulation of the auxin transport model	114
7.1	Summary of the single-cell RNA-sequencing data	123
B.1	Dose-response to kinetin	154
B.2	Effect of kinetin and auxin on regeneration of <i>Marchantia gemma</i>	155

Chapter 1

Introduction

1.1 Plant growth and engineering

Members of the plant kingdom demonstrate great diversity in the shape of their organs: leaves, flowers, fruits, seeds, etc. What makes them stand out from the members of other kingdoms is a rigid cell wall, which provides stiffness to the plant tissue and resilience to external forces. The cellulose, deposited into the cell wall, allows trees to reach heights of tens of metres without collapsing under their own weight and withstand the forces generated by the wind. Being tough yet mouldable, plant tissue has been used as a construction material for many thousands of years.

From the developmental perspective, the rigid cell wall prevents cell motility and so the process of shape formation relies on the precise coordination of cell expansion and division. This feature makes plants attractive candidates for the creation of programmable multicellular organisms. Developing tools for predicting the plant shape based on its DNA sequence opens up possibilities for the precise engineering of crop traits as well as a new manufacturing paradigm, where wooden structures can be naturally grown without the need for traditional processing methods.

The state of the art for artificial shape design is computer graphics programs^[1-3], where three-dimensional (3D) shapes can be modelled through the iterative subdivision and deformation of polyhedral meshes. These systems allow for the precise manipulation of shape through specifications of the topology and geometry of the meshes, but also include artistic sculpting tools like 3D brushes^[4]. In essence, these brushes apply local pre-defined deformations to a user-specified location on a mesh and are analogous to using physical sculpting tools on clay or other modelling substrate. The basic principle behind plant morphogenesis is a differential expansion of tissue, which leads to its deformation. On a fundamental level, the differential expansion that drives plant growth can be thought of as a deformation applied by a 3D brush, where the sculpting logic is driven by the biological processes inside the living plant.

Processes that control tissue expansion, and thereby contribute to plant morphogenesis, span three levels of biological organisation: genetic, cellular and tissue-wide. Being such a broad topic, plant morphogenesis has attracted scientists from various branches of biology, e.g., theoretical biology, systems biology, biophysics, biomechanics, cell

biology, genetics, physiology, molecular biology. Over the past hundred years, the scientific community has aggregated a large amount of data and many models, describing the details of plant development at different scales and from the perspective of different scientific disciplines.

Bringing details from all of these fields into a single multi-scale model of plant morphogenesis for the purpose of engineering plant shape is a challenging task. An alternative approach is to aim to produce a minimal viable model that would not explain morphogenesis to its finest details, but would be sufficient for the practical modification of the plant shape. Then the main challenge is to choose the right abstraction with the right subset of details that would result in a satisfactory model.

The approach taken in this thesis is not aimed at any particular question in the world of plant morphogenesis. It rather presents an attempt to create a certain abstraction of the knowledge in the field, and to develop the methods, that would be suitable for engineering the plant shape.

Section 1.2 of the Introduction reviews the main aspects of plant morphogenesis at various scales, while section 1.3 reviews the main tools and techniques that were used in this thesis.

1.2 Hierarchical view of morphogenesis

1.2.1 Mathematical notations

scalars	not bold font, lower case	a
vectors	bold font, lower case	\mathbf{v}
tensors	bold font, upper case	\mathbf{A}
inner product	\cdot	$\mathbf{v} \cdot \mathbf{u} = v_i u_i$
outer product	\otimes	$(\mathbf{u} \otimes \mathbf{v})\mathbf{a} = (\mathbf{v} \cdot \mathbf{a})\mathbf{u}$
divergence	$\nabla \cdot$	$\nabla \cdot \mathbf{v} = \frac{\partial v_i}{\partial x_i}$
trace	tr	$tr \mathbf{L} = L_{ii}$
gradient	$\frac{\partial}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \frac{\partial u_i}{\partial x_j}$

1.2.2 Tissue level

Morphogenesis is a dynamic process that results in the formation of the characteristic shape of an organism. The ultimate goal of engineering morphogenesis is to be able to control the geometry of plants at the tissue level in a predictable way. From this perspective, growth can be represented as a time-dependent transformation function $\chi : B_0 \rightarrow B_t$ that maps coordinates of the tissue points (\mathbf{x}) at an initial configuration B_0 to coordinates of the same points (\mathbf{X}) at some later configuration B_t at time t , such that:

$$\mathbf{x} = \chi(\mathbf{X}, t), \forall \mathbf{X} \in B_0$$

This parallel between the development of biological shape and geometric transformations was laid out more than a hundred years ago by D'Arcy Thompson^[5]. Since then, the mechanics and mathematics behind tissue deformation under differential expansion have been formalised by a theory of morphoelasticity^[6-8]. Many aspects of plant tissue kinetics are analogous to those in compressible fluid flows^[9], and hence some analytical and experimental methods of fluid dynamics can be applied directly to the study of morphoelasticity. The key difference between the problems in fluid dynamics and those of morphoelasticity is that the law of mass conservation does not hold in a classical sense. Without taking into account plant physiology and metabolism, and focusing on sheer observation, it may appear that growth goes in violation of the conservation of mass as, by definition, growth is an increase of the mass of the organism. Therefore, instead of a classical equation of mass conservation

$$\frac{\partial \rho}{\partial t} + \rho \nabla \cdot \mathbf{v} = 0$$

where ρ is tissue density and \mathbf{v} is velocity field in the Eulerian frame of reference, morphoelasticity deals with a more general problem

$$\frac{\partial \rho}{\partial t} + \rho \nabla \cdot \mathbf{v} = \gamma$$

where γ is tissue growth rate. The physical and biological meaning of γ is that it shows how fast a small volume element of tissue expands over time.

The gradient of the velocity field $\nabla \mathbf{v}$ has a special meaning in morphodynamics. It is called a *growth tensor* (\mathbf{L})^[10] and under the condition of incompressible material its trace is equal to the tissue growth rate (γ):

$$tr \mathbf{L} = \gamma$$

Apart from information about the rate of growth, the growth tensor contains information about anisotropy of the growth and rotation of the tissue. Hence, to have a complete description of tissue expansion it is sufficient to calculate its velocity field.

Estimation of the velocity field of a growing tissue can be achieved by tracking natural landmarks on its surface, e.g. the intersection of the cell walls in apical meristems^[11], veinal junctions on the leaf surface^[12] or repetitive organs on the scale of the whole organism. In addition, artificial landmarks like fluorescent particles^[13] or ink dots^[14] have been previously used for this purpose.

The complication of the morphoelastic models comes from the fact that the resultant deformation tensor ($\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}}$) of the tissues comes from two processes: the pure growth of individual tissue elements or cells, and the onset of mechanical consensus. The difference between the two can be visualised in the following thought experiment. An imaginary piece of tissue consisting of several joined cells grows due to the expansion of its individual cells. At the same time, the physical constraints the cells impose on each other deform the geometry of the tissue as a whole. The same resultant deformation could have been achieved by first separating the cells and allowing them to grow

independently from each other, and then stitching them back together and allowing mechanics to deform them into the final shapes. The details of this independent growth can be summarised in the tissue expansion tensor \mathbf{G} , while the onset of the mechanical consensus during stitching in the elastic response tensor \mathbf{A} (Figure 1.1A). The central postulate of morphoelasticity is that the resultant deformation can be decomposed as

$$\mathbf{F} = \mathbf{A}\mathbf{G} \quad (1.1)$$

while the connection between \mathbf{F} and \mathbf{L} is as follows:

$$\frac{\partial \mathbf{F}}{\partial t} = \mathbf{L}\mathbf{F}$$

In practice, this means that although it is possible to estimate the growth rate of the tissue by recording its velocity field, it is hard to distinguish between the effects of pure tissue expansion and tissue mechanics on the resultant tissue deformation. In order to make an educated guess about this distinction, it is important to understand how both \mathbf{A} and \mathbf{G} are controlled physiologically within the plant tissue on the scale of individual cells.

1.2.3 Cell level

Cells are discrete units of plant tissue; it is their rate and anisotropy of division and volumetric expansion that defines the resultant growth of tissue^[15]. Therefore, it is essential to consider the mechanics and physiology behind cell expansion in order to build realistic and predictive plant tissue models.

The basic physiological principle behind cell expansion is well understood^[16]. A cell's volume increases as a result of the net influx of water into the cell protoplast. The water fluxes going in and out of the cell are driven by the differences in water potential (ΔW) inside and outside of the cell:

$$\Delta W = W_{out} - W_{in} = \Delta p - \Delta \Pi$$

where $\Delta \Pi = \Pi_{out} - \Pi_{in}$ is the difference of osmotic pressure outside and inside of the cell and $\Delta p = p_{out} - \Delta p_{in}$ is the difference in hydrostatic pressure outside and inside of the cell. The former results from an increased concentration of cell-membrane-impermeable solutes inside the cell and can be considered as a driver of water influx. The latter is a result of mechanical constraints imposed by a cell wall on a cell membrane during cell expansion and can be considered as a suppressor of water influx.

Water always flows through the tissue from high to low water potential, so the positive ΔW values correspond to a net influx of water into the cells and resultant expansion. The process of osmotic-pressure-driven expansion is depicted in Figure 1.1B. The difference in osmotic potential between neighbouring cells, or between a cell and the environment, creates water fluxes, which lead to the accumulation of water inside cells with local water potential minima. The influx of water into a protoplast makes the cell membrane push upon the cell wall. The reaction force in the cell wall generates a hydrostatic turgor pressure, which increases the water potential, slowing down the water

influx. The water keeps flowing into the cell until its water potential equilibrates with that of its neighbours. If the cell wall stress exceeds the yield stress, stress-relaxation and loosening of the cell wall^[16,17] take place, leading to cell expansion. As a result, the water potential drops, promoting further water influx^[18–21] and restarting the cycle anew.

An important thing to note is that although hydrostatic pressure inside a cell suppresses water influx, it is the driver of cell-wall loosening responsible for the non-reversible expansion of the cell. A resultant model combining these two opposing effects of hydrostatic pressure on cell expansion was proposed by Lockhart^[16]. It can be summarised as

$$\frac{dV}{dt} = \frac{K_w \phi}{K_w + \phi} \Delta W \quad (1.2)$$

where $\frac{dV}{dt}$ is a rate of volumetric cell expansion. K_w is the water permeability of the membrane determining how fast the water crosses the cell membrane. Higher values of K_w would result in higher rates of water fluxes under the given difference in water potential. ϕ is the cell wall expansibility rate determining how fast a cell expands, where higher values of ϕ would result in higher expansion rates under the given hydrostatic pressure. As seen from the above equation, the cell expansion rate is driven entirely by the differences in the osmotic potential and does not depend on the hydrostatic pressure implicitly.

The physical and biological meanings the equation 1.2 are very similar to that of equation 1.1. The meaning of $\frac{1}{V} \frac{dV}{dt}$ is similar to that of the growth rate γ introduced in the previous subsection. Furthermore, the difference in water potential (ΔW) and water permeability (K_w) accounts for the pure growth of the isolated cells, similar to the tissue expansion tensor \mathbf{G} , while the cell wall expansibility (ϕ) accounts for the mechanical constraints, similar to the elastic response tensor (\mathbf{A}). Another important conclusion from equation 1.2 is that the biological processes responsible for controlling the rate of cell wall stress-relaxation and water fluxes across the cell membrane are key to understanding and engineering plant morphogenesis.

The loosening of the cell wall cannot be entirely accounted for by its viscoelastic properties^[22], which can only explain a short-term response to the applied stress. Early studies on the mechanical properties of plant cell walls uncovered that the irreversible cell wall extension can be induced by soaking plant tissue in a buffer with a low pH^[23,24]. This phenomenon, termed ‘acidic growth’, was shown to be absent in heat-treated plant tissue, suggesting that the loosening of the cell wall might be a result of enzymatic activity^[25]. Two enzymes able to recover the acidic growth of a heat-treated plant tissue were first identified in cucumber seedlings. They were named expansins (EXP)^[26] and were shown to bind to both the plant cell wall and purified cellulose^[27]. Furthermore, expansins have been shown to be expressed in the growing tissue^[28,29] and promote cell expansion when applied exogenously^[30,31].

In order to grow, cells should not only expand but also synthesise new cell wall material. The cellulose for the new cell wall is synthesised by membrane-bound cellulose synthase (CesA) complexes, arranged in hexameric arrays, or rosettes^[32]. The local arrangement of the cellulose microfibrils in the cell wall is important for defining the anisotropy of cell expansion. By aligning the microfibrils, cells can promote expansion in a particular direction^[33,34]. Thus, while the turgor pressure acts isotropically on the cell wall, the anisotropy of the cell expansion is controlled by the orientation of the microfibrils. The orientation of the cellulose microfibrils deposition was shown to be controlled

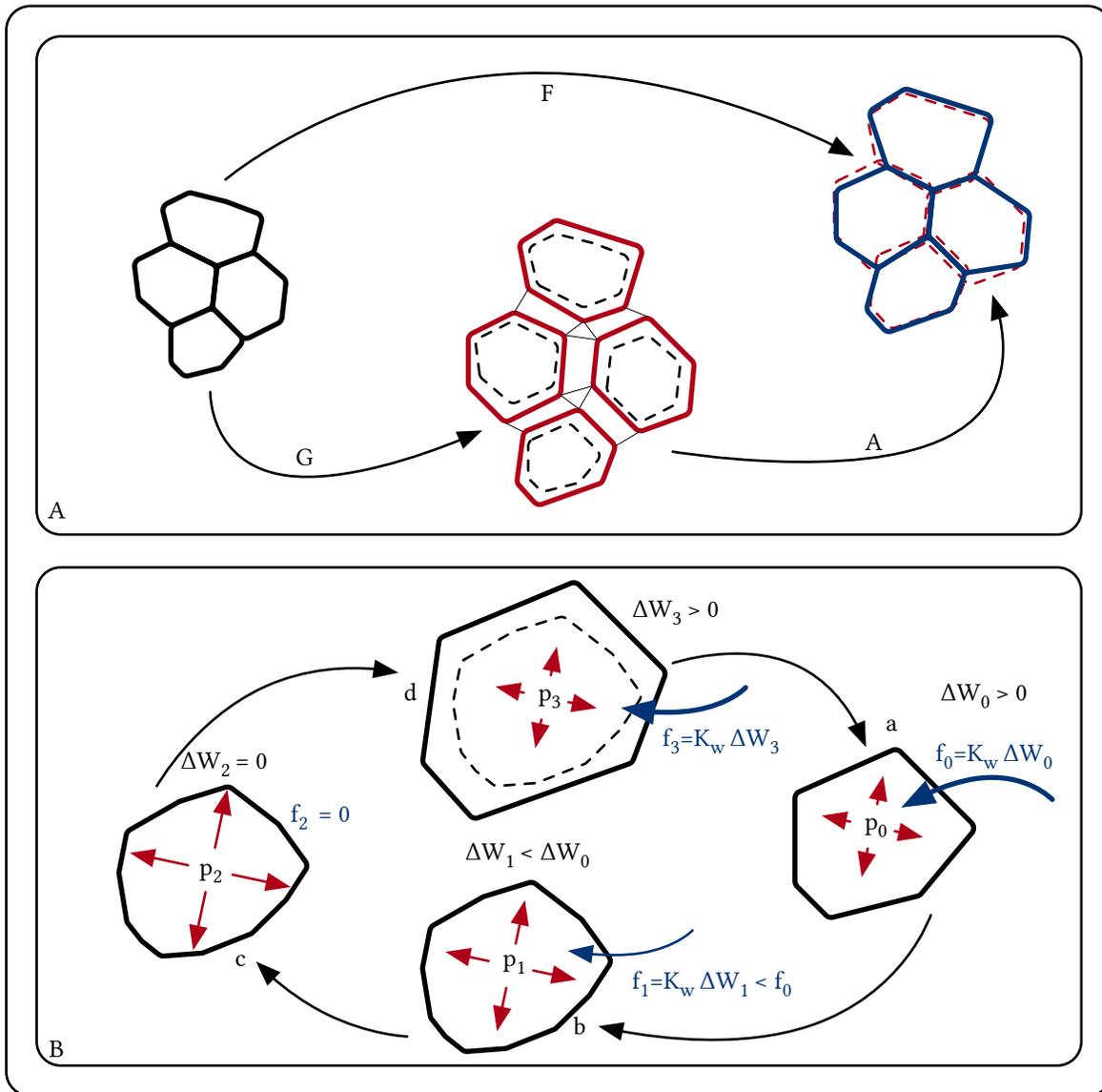


Figure 1.1: **Expansion at the tissue and cellular level.** Panel A depicts the central dogma of morphoelasticity. A deformation of the tissue as a result of growth (F) can be decomposed into two components: imaginary unconstrained growth of tissue elements or cells (G) and mechanical consensus between the tissue elements (A). The red lines represent geometry of the cells when allowed to grow unconstrained. Black lines represent the original geometry of cells before growth. Blue lines represent the geometry of cells after growth. Panel B depicts the osmosis-driven expansion of a plant cell. a: Water (blue arrows) flows into the cell with rate f_0 following the difference in water potential outside (W) and inside (W_0) of the cell. b: As a result of water influx, hydrostatic pressure (p) increases, lowering the water potential inside the cell and, therefore, the rate of water influx (f_1). c: Subsequent water influx increases the hydrostatic pressure until the water potential inside and outside of the cell equilibrates. d: Cell wall loosens while releasing the hydrostatic pressure and lowering the water potential inside the cell below that outside the cell.

by the cortical microtubules, such that the orientation of the microtubules is parallel to the direction of the microfibril deposition^[35]. The orientation of the cortical microtubules has in turn been shown to respond to tissue stress, aligning parallel to the direction of the principal stress both within the tissue and externally applied^[36,37], thus creating a feedback loop between growth dynamics at the tissue and cellular levels.

Water permeability of plant cell membranes is known to be controlled by a family of membrane channels called aquaporins. They are the only known membrane proteins that allow transport of water molecules and their expression level has been shown to be higher in expanding cells^[38-41].

The challenge of plant morphogenesis at the cellular level is to identify the fundamental principles behind the genetic and physiological control of the enzymes involved in cell wall expansion and water permeability. Successful engineering of plant morphogenesis will rely on the fine control of the temporal and spacial activities of these enzymes by rewiring the endogenous and introducing novel gene circuits.

1.2.4 Gene level

All processes inside a plant cell are dependent on and regulated by its genotype. As discussed in the previous section, control of the cell expansion rate and anisotropy as well as water transport is achieved via the action of particular proteins. Proteins are synthesised using ribonucleic acid (RNA) templates in a process known as translation. These messenger RNA (mRNA) molecules are polymers that consist of four types of monomers: adenine (A), thymine (T), guanine (G) and cytosine (C). Each consecutive triplet (codon) in the stretch of an mRNA molecule encodes one of 20 amino acids that are the building blocks of proteins. The mRNA molecules are in turn synthesised using deoxyribonucleic acid (DNA) templates, which are long polymers carefully packed into several chromosomes inside each cell. This synthesis of proteins from DNA templates through the processes of transcription and translation is a central dogma of molecular biology (Figure 1.2A).

Every plant cell has a limited amount of resources and several thousand genes, encoding proteins and enzymes of different functions. Synthesising all of the proteins constantly might not be efficient, since every protein is only useful at a particular stage of plant development. Therefore, cells should be able to regulate precisely which proteins to synthesise at any given moment based on local environmental cues. Proteins that control the expression of genes fall into a special class and are called *transcription factors*. They can be thought of as master regulators that orchestrate cell function and response.

The way transcription factors mediate the expression of their target genes is by binding to DNA downstream of the transcription start site, inside the promoter regions of the target genes. By doing so, transcription factors can either promote or repress the initiation of the transcription of the corresponding genes. The interaction between a transcription factor and DNA happens through the transcription-factor-specific binding DNA motifs; i.e. each transcription factor prefers to bind to its own set of small DNA motifs inside the promoter region (Figure 1.2A).

A functional separation between the promoter regions and the protein-coding regions decouples the regulation of a gene expression from its function. This means that the promoters and the protein-coding regions can be treated as modular building blocks that can be mixed and matched to program the desired behaviour of a cell.

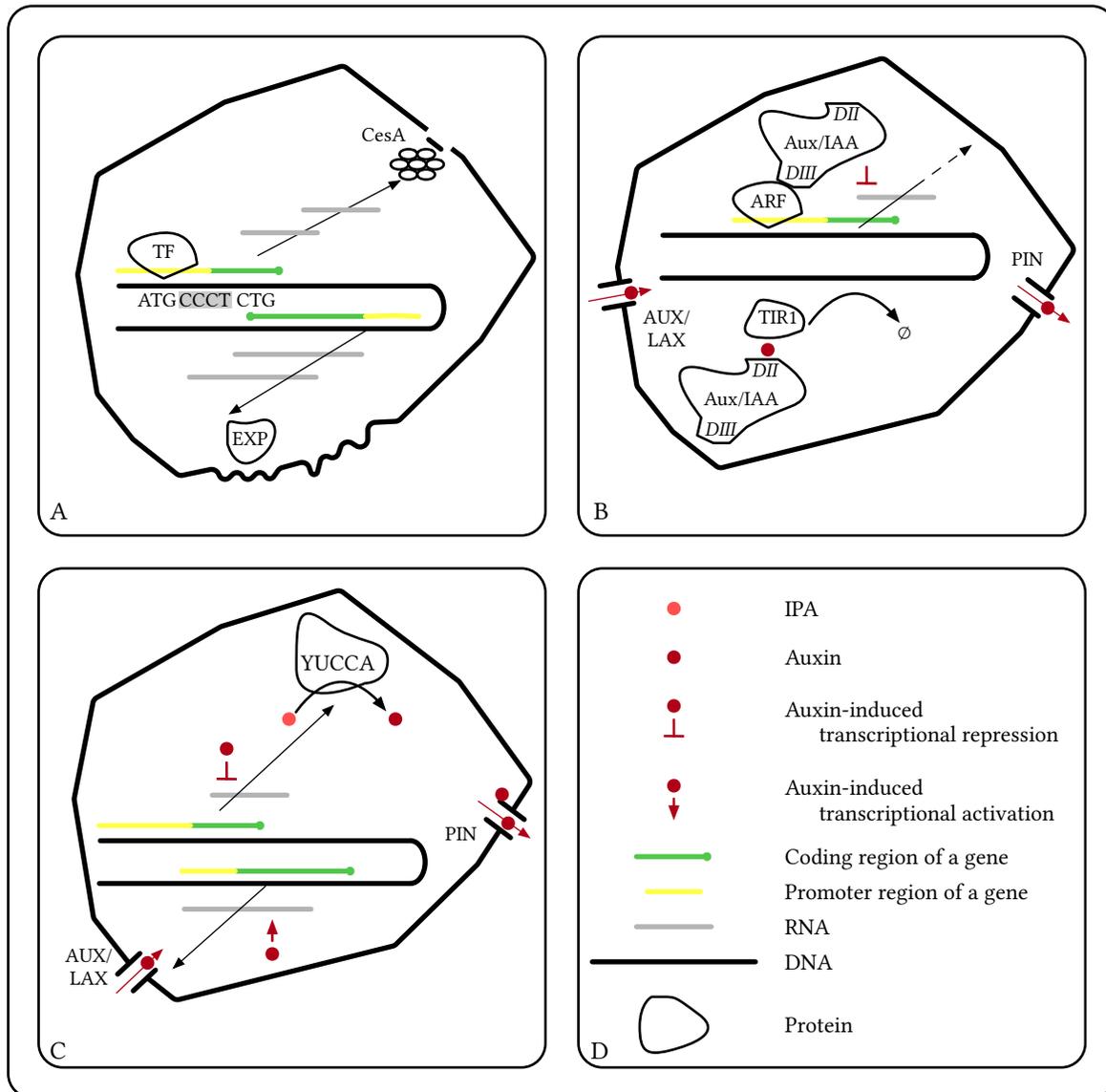


Figure 1.2: **Regulation of tissue expansion at the gene level.** Panel A shows the central dogma of molecular biology with arrows representing a process of transcription and translation on the way from a DNA template to a protein on an example of cellulose synthase (*CesA*), which is involved in the synthesis of new cell-wall material and expansin (*EXP*), which is responsible for cell-wall loosening during expansion. The panel also shows a transcription factor (*TF*), which binds to the promoter region of *CesA* via an interaction with the ‘CCCT’ motif to regulate its transcription rate. Panel B depicts the details of the auxin signalling pathway. Auxin response factor (*ARF*) is a transcription factor with a wide range of genes. In the absence of auxin it is bound to *Aux/IAA*, which represses the initiation of transcription of the target gene, via *DIII* domain. In the presence of auxin, *Aux/IAA* binds to *TIR1*, which tags *Aux/IAA* for degradation (\emptyset). Auxin can be actively imported into the cell by an action of *AUX/LAX* proteins or actively exported out of the cell by *PIN* proteins. Panel C shows the auxin-dependent regulation of auxin concentration inside a cell. *YUCCA* is an auxin-converting enzyme that converts auxin to *IPA*, which is then exported out of the cell by *AUX/LAX* proteins. *PIN* proteins are involved in auxin transport.

1.2.5 Cell communication and auxin transport

The coordination of processes within cells is orchestrated by gene interaction, while the coordination of the processes between cells happens via hormonal regulation. The classical model of hormonal regulation, known as the French flag model^[42], states that the difference in cell behaviour can be explained by the response of the cells to different concentrations of hormones, or *morphogens*. For example, a particular morphogen might signal cells to promote expansion. Then the cells exposed to high concentrations of such a morphogen would expand more actively than those exposed to low concentrations. According to this model, the spatial patterns of cell phenotypes must be correlated to the spatial pattern of the morphogens.

A theoretical mechanism on how patterns of morphogens might be self-organising in living organisms was proposed by Turing^[43]. The model, known as the chemical basis of morphogenesis, relies on two diffusible chemicals that are allowed to react with each other.

The role of phytohormone auxin in cell elongation is well known. In relation to the theory of acidic growth, it has been demonstrated that auxin triggers the activation and trafficking of the H^+ -ATPase to the plasma membrane in *Zea mays* coleoptiles^[44], lowering the apoplastic pH to levels favourable to the expansins. In *Arabidopsis thaliana* (*Arabidopsis*) roots, however, endogenous application of auxin has been shown to increase the pH of the apoplast, inhibiting cell expansion^[45]. Therefore, although the role of auxin in controlling cell expansion is well-established, the precise action of auxin varies depending on tissue type.

Apart from controlling apoplastic pH, the endogenous application of auxin has been shown to cause upregulation of expansins and cellulose synthases in seedlings of *Arabidopsis*^[46,47], thus not only promoting cell wall loosening but also activating machinery for the reorganisation and synthesis of cellulose microfibrils.

Finally, it has been demonstrated that the polarisation of the membrane localisation of the auxin efflux transporter PIN1 is aligned with the orientation of cortical microtubules in the shoot apical meristem of *Arabidopsis*^[48]. The analysis of this alignment led to the conclusion that the fluxes of auxin might also be controlled by tissue stress patterns.

Together, these pieces of evidence suggest that auxin plays a vital role in controlling cell expansion and creating a feedback loop between growth at the tissue and cellular levels in plant shoots. Therefore, auxin transport might be a good target for achieving control of plant morphogenesis.

Auxin response includes a well-characterised signal transduction pathway (Figure 1.2B). The interaction between auxin, Aux/IAA, TIR1/ABF and protein degradation machinery takes the central role in auxin signalling^[49]. Aux/IAA is a transcriptional repressor, the sequence of which features three functional domains. Domain I contains an EAR motif, responsible for transcriptional repression and the recruitment of co-repressors^[50]; domain II contains a DII degron motif, responsible for the auxin-sensitive degradation of the Aux/IAA^[49]; domain III facilitates interaction with auxin response factors (ARFs) and self-dimerisation^[51].

The auxin response factors bind to the AuxRE motifs^[52,53] in the promoters of auxin-responsive genes. When the concentration of auxin is low, the Aux/IAA oligomerises with itself and dimerises with the ARFs, repressing the transcription of auxin-responsive genes. When present, auxin facilitates the binding of the Aux/IAA to the F-box TIR1/ABF auxin co-receptor through the DII motif of the Aux/IAA. The binding of the TIR1/ABF mediates the

ubiquitination of the Aux/IAA-TIR1/ABF complex and its further degradation. The resultant drop of the Aux/IAA concentration liberates the ARFs, allowing for the transcriptional activation of auxin-responsive genes.

It has been shown that intracellular auxin controls its own transport by inhibiting the endosomal compartmentalisation of auxin efflux proteins (PINs) in *Arabidopsis*^[54]. This leads to the localisation of PINs at the plasma membrane and an increase in the auxin export rate. The distribution of the PINs on the plasma membrane is not homogeneous and is responsible for the polar auxin transport, which has been a basis for plant development models^[55].

Two mechanisms for the polarisation of the PIN proteins have been proposed. According to the concentration-based models, the PIN proteins predominantly localise to the section of the plasma membrane that faces a cell with the highest auxin concentration in its neighbourhood (Figure 1.2C). Such models have been used to explain the emergence of auxin maxima and the corresponding organ positioning during phyllotaxis^[56,57]. In contrast, flux-based models propose that the distribution of PIN proteins on the plasma membrane is proportional to the distribution of the total auxin transported through a particular region of the cell membrane. These models are used to explain the self-organisation of leaf venation^[58,59].

Apart from efflux, auxin has been shown to be involved in controlling its own influx, e.g. during lateral root emergence in *Arabidopsis*^[60], by activating transcription of the LAX gene, encoding auxin importers.

Another known feedback is between auxin and its biosynthesis. It has been demonstrated that the activity of some *YUCCA* genes, which encode an enzyme involved in tryptophan-dependent auxin biosynthesis, is transcriptionally down-regulated upon the external application of auxin in *Arabidopsis*^[61] as well as other species^[62].

These forms of feedback in auxin transport demonstrate the complexity of the regulation behind cell expansion, which should be taken into account when designing systems for the control of plant morphogenesis.

1.2.6 Summary

Plant tissue expansion can be studied at different scales and hierarchical levels. At the tissue scale, tissue expansion can be represented by a physical model offered by the theory of morphoelasticity. At this scale, no consideration is given to the physiological or genetic origin of plant growth and the purpose of such a model is to describe the physical aspect of growth derived from the laws of mass and momentum balance.

At the cell scale, physiological effects start to matter. The Lockhart equation (1.2) shows a physiological origin of cell expansion, i.e. interplay between water transport between the cells and extensibility of the cell walls. The model of morphoelasticity and the principle behind the Lockhart equation can be merged by acknowledging that the tissue growth rate (γ) and rate of cell volume change ($\frac{dV}{dt}$) share similar meaning. For example, as shown in Appendix C, these models can be combined through the following relationship:

$$\gamma = -\frac{K_w\phi}{K_w + \phi}\nabla^2\Pi \quad (1.3)$$

where all variables can be, in principal, functions of time and space. This combined model boils down the regulation of tissue growth (in an unrestricted imaginary sense of morphoelasticity) to two parameters: K_w – water permeability of the membrane and ϕ – extensibility of the tissue, i.e. its ability to grow under hydrostatic pressure.

At the genetic level, the effects of individual genes become apparent. The most relevant set of genes for the model in equation 1.3 are those related to modification of cell-wall properties, and hence those that influence ϕ and those that are related to water transport, and hence influence K_w . Among the former group are expansins, which regulate the loosening of the cell wall and cellulose synthases, which deposit new cell-wall material. The genes of both proteins have been shown to be controlled by phytohormone auxin, which itself has a sophisticated regulatory mechanism that has been attributed as being central to plant development at genetic, cell and tissue scales. The work presented in this thesis has a particular focus on auxin transport and its involvement in the regulation of plant tissue expansion.

1.3 Tools

1.3.1 Synthetic biology

As mentioned in the previous section, the genes and their regulation are the intrinsic drivers of shape development in plants. Over millions of years of evolution, plant shape has changed under the influence of random mutations and natural selection^[63]. Later, the domestication of plant species led to the artificial selection of favourable traits^[64], demonstrating that small genetic perturbations can lead to significant changes in plant shape. For example, there have been reports of a plant with a teosinte-like phenotype acquiring a maize-like phenotype in only 18 years of selective breeding.

As science has progressed, new tools for more direct modification of plant DNA have become available. The discovery of the restriction enzymes^[65,66], which are capable of cutting DNA at a precise recognition sequence, and DNA ligases^[67], which can repair the DNA molecule after being cut, in the late 1970s enabled the modification of DNA sequences *in vitro*. Around the same time, it was demonstrated that the species of soil bacteria *Agrobacterium tumefaciens* can be exploited to deliver foreign DNA into plant cells^[68]. Other methods for transferring DNA into plant cells were also invented, including electroporation^[69,70], microinjection^[71] and biolistics^[72], which involves the bombardment of plant cells with gold nanoparticles covered in DNA. Together, these methods allowed the artificial deletion and introduction of new genes into plant tissue and produced a series of genetically modified crop species with novel traits, like resistance to herbicides^[73], resistance to insects^[74] or longer shelf life^[75].

All of these early-days genetic modifications were introducing constitutively expressed genes, i.e. they were always ‘on’ in all the cells of the plant. Such strategy is arguably insufficient for engineering plant-shape development, as the process relies on orchestrated and differential behaviour in different cells. For the engineering of morphogenesis, the temporal and spatial effects of the introduced genes should be considered in a predictable fashion.

Synthetic biology builds on classical genetics with the addition of engineering principles, creating a heavily model-driven approach to modification of living systems. The implementation of analogs of simple electric circuits like a latch^[76] or an oscillator^[77] showed that it is feasible to build predictive models of genetic interactions. Since then, more advanced electronics-inspired biological systems have been implemented, e.g. logic gates^{[78][79]}, memory elements^[80], analog computers^[81], an oscillator with independent control for the modulation of frequency and amp-

litude^[82]. In addition, genetic circuits for the generation of spatial and temporal patterns have been built^[83,84]. The prospect of automating biological logic circuitry design has also been explored, e.g. by building a compiler that transforms digital logic defined using hardware description language Verilog into the corresponding DNA sequence^[85].

The majority of the genetic circuits described above rely on the modularity of promoter and coding regions, as described in subsection 1.2.4. This modularity allows the linking together of two genes by making the product of the coding region of one gene regulate the promoter activity of the other. The product of the first gene can either be a transcriptional regulator that activates or represses the activity of a particular promoter, or a protein that catalyses the production of a chemical to which the promoter of the second gene is sensitive, whether directly or with the help of additional genetic machinery.

Together with a growing repertoire of genetic parts, the methods of DNA assembly have become more efficient and now allow the stitching together of DNA fragments of increasing lengths and numbers of components. Early strategies for DNA assembly relied on restriction enzymes and DNA ligases^[86]. The next wave of strategies allowed DNA to be assembled without restriction enzymes by using either the overlap assembly method^[87] or DNA recombination^[88]. The state of the art method for DNA assembly uses the next generation of restriction enzymes Type IIS^[89,90]. The benefit of the Type IIS assembly is that a single restriction enzyme can be used in one reaction to combine multiple DNA parts. However, the ability to do this relies on establishing standards. In particular, the definition of the prefix and suffix sequences for each different type of DNA part – e.g. promoter, coding region – is required. Attempts to establish such standards have been made in both bacterial^[91] and plant synthetic biology communities^[92].

1.3.2 Genomic resources

The ability to introduce new traits into plants, in particular those related to morphogenesis, will depend on the comprehensive characterisation and isolation of plant genetic parts, i.e. the coding regions and the corresponding regulatory elements. Genomes of an ever-increasing number of plant species are being sequenced and stored at public resources like Phytozome^[93]. A recent initiative aims to sequence 10,000 plant genomes^[94]. These genomic data create a valuable resource for the mining of relevant DNA parts and the transfer of knowledge for the engineering of new species from well-established plant systems^[95].

The process of discovery of genetic parts relies on the data generated from genomic (DNA) and transcriptomic (mRNA) sequencing. The mRNA sequence can be used to predict the protein sequence it encodes. Protein-coding regions always start with an ‘ATG’ codon and end with a ‘TAA’, ‘TAG’ or ‘TGA’ codon. One approach to estimate the protein-coding regions within a given mRNA sequence is to find the longest subsequence that starts with an ‘ATG’ codon, ends with a stop codon and has only one stop codon, i.e. has no intermediate stop codons^[96]. Selection of the longest sequence is justified by the assumption that if one of the shorter potential coding sequences was the true one, the longer one would have disappeared in the process of evolution, i.e. that the shorter sequences with the same function are advantageous and would be naturally selected.

This direct approach can be improved upon by leveraging information from the comparative analysis of the tran-

scriptomes from various species. During the process of evolution, the sequence of the mRNA may change, however, certain functional motifs of the corresponding amino-acid sequence remain highly conserved. Many databases of these conserved domains and their corresponding functions exist, e.g. Pfam^[97], SUPERFAMILY^[98,99], PROSITE^[100]. The central resource InterPro^[101,102] combines protein motifs from a total of 13 databases and provides tools for annotating potential coding sequences with known protein motifs. Not only does this allow an insight into the function of a coding sequence, it also helps identify the coding regions of previously uncharacterised mRNA sequences. For example, a long potential coding region, identified by the direct approach described above, but containing no known protein motifs is unlikely to be correctly identified as being truly coding.

The non-coding regions, also known as untranslated regions (UTRs), upstream (3' UTRs) and downstream (5' UTRs) of the coding regions have an important role in the regulation of translation and hence are an important addition to the toolbox of plant genetic parts.

Since the sequence of the transcribed DNA region and the resultant mRNA molecule are the same, it is possible to predict from which stretch of the DNA it originated. In eukaryotes, like plants, this task is complicated by the fact that the primary transcripts are spliced, i.e. some of its regions (introns) are cut out, producing the mRNA molecule composed of the remaining regions (exons). Several computational algorithms exist that can map a given mRNA sequence to the genome of interest^[103-106].

Information about the origin of the mRNA on the genome allows the promoter and terminator regions of the corresponding gene to be defined as the sequences upstream and downstream from the mapped mRNA. Promoter regions are responsible for regulating the initiation of the gene transcription, while terminator regions signal the transcription machinery to stop the process.

The information about the coding region prediction and annotation with functional protein motifs, as well as information about the mapping of the transcripts to the genome, is stored in public genomic databases.

1.3.3 Fluorescence microscopy and image analysis

Studying the DNA sequence of plant genes may reveal important information about their cellular function and regulation, however, in order to understand the role of a particular gene in plant development, information about the time and location of its expression is required. One way to gain this information is through fluorescent microscopy.

Fluorescent microscopy exploits the fluorescent properties of materials to create imaging contrast. During image acquisition, a sample is illuminated by a light source with a narrow bandwidth, e.g. a laser or broad-spectrum lamp in combination with a narrow-band-pass filter, with wavelength matching the excitation wavelength of the fluorophore under study. The fluorophore then reemits the light at a longer wavelength and is collected either by the eyepiece or a sensor through another narrow-band-pass filter with the matching wavelength. By doing so, it is possible to achieve a high signal-to-noise ratio during the imaging of the fluorophore.

The identification^[107] and cloning^[108] of the green fluorescent protein (GFP) from *Aequorea* jellyfish saw an opportunity to create fluorescent reporters for highlighting cellular processes and studying biological patterns using fluorescent microscopy. More than 50 different variants of fluorescent proteins are currently known, with emission

wavelengths covering most of the visible spectrum^[109].

Two major strategies are used for exploiting fluorescent proteins for *in vivo* plant imaging. The first involves the introduction of artificial DNA into plants containing the coding sequence of the GFP downstream of a promoter of interest. This allows the expression pattern of the promoter to be studied. For example, a synthetic DR5^[110] promoter containing several copies of an auxin response factor binding motif was developed to study patterns of auxin signaling and was successfully used in combination with GFP to study auxin signalling in Arabidopsis roots^[111,112].

Alternatively, the coding sequences of the fluorescent proteins can be fused to the coding sequences of endogenous proteins to study their intracellular localisation or highlight other cellular features. For example, fluorescent proteins have been fused to microtubule-associated proteins to highlight the cortical microtubules^[113] or the coding sequence of the *PIN1* gene in the study of auxin transport^[114].

Fluorescence microscopy in combination with fluorescent reporters allows the capture of spatio-temporal patterns of plant development. The subsequent analysis of the captured data relies on the image processing algorithms that allow, for example, the segmentation of the features of interest^[115] or tracking of anatomical structures in time^[116].

These methods allow patterns of growth and gene expression to be extracted and can be used as inputs for the models of plant development. An example of such an approach, termed *in planta* cytometry, has been introduced in Arabidopsis roots for the simultaneous recording of tissue expansion rates and auxin signalling^[112] using the combination of a fluorescent membrane marker and fluorescent promoter fusion.

1.3.4 Computational models of plant development

A useful tool for gaining intuition about the relationship between different levels of hierarchical organisation is computational modelling^[117].

One type of numerical simulation combines the mathematical description of plant tissue growth, as outlined in subsection 1.2.2, with the hypothetical regulation of the tissue growth (γ) by genetic and hormonal processes. One common approach is to assume that γ is dependent on the concentration of a morphogen, which is synthesised, degraded and transported across the tissue according to the proposed hypothesis^[118,119]. For example, this approach was used to model the development of flowers^[120], petals^[121] and pollen tubes^[122]. While these types of simulations do not account for tissue cellularity, it is possible to introduce physiological aspects of growth to this continuous representation of the plant tissue. Such an approach, integrating the physiology of cellular water movement and turgor-driven expansion into a continuous tissue model, was applied to model the development of tomato fruit^[123].

To capture the cellular aspects of plant growth, vertex models^[124–127] can be used. The vertex models represent the geometry of individual cells as polygons with shared edges, or cell complexes^[128], which are generic representations of geometric structures related to each other through common boundaries. Such cellular models capture the mechanical properties of the cell walls and the cell expansion driven by hydrostatic pressure. An alternative approach for modelling the cellularity of the plant tissue is through particle-based simulations^[129], where the precise geometry of the cells is ignored. Instead, the cells are modelled as spherical or circular particles with a set of equations modelling forces for physical cell–cell interactions.

1.3.5 Marchantia as a model organism

Engineering morphogenesis is challenging because of the number of individual components and types of interactions involved in the development of plant shape. One way to reduce this complexity is to use a model system that is inherently simple. The thalloid liverwort, *Marchantia polymorpha* (Marchantia), offers a simple shape advantageous for morphogenetic studies^[130]. Among other benefits, Marchantia can propagate asexually, producing small ($\sim 1\text{ mm}$ diameter), flat clonal propagules, the development of which can be studied at a sub-cellular resolution using a microscope equipped with a low-power objective (Figure 1.3).

These propagules, or gemmae, have been previously used for the quantitative analysis of cell expansion^[130] and gene expression patterns^[131] through fluorescent microscopy.

Analysis of the Marchantia genome revealed low genetic redundancy in the most regulatory pathways^[132]. For example, the auxin transcriptional response system in Marchantia contains only one *TIR1/AFB* gene, one *Aux/IAA* gene and three *ARF* genes, compared to 29 *Aux/IAAs*, 6 *TIR1/AFBs* and 23 *ARFs* in Arabidopsis^[133,134].

A variety of powerful genetic tools and techniques has been assembled in Marchantia^[135]. Among them are several nuclear^[136–142] and plastid^[143,144] transformation methods, CRISPR/Cas9-mediated targeted mutagenesis^[131,145], transactivation using GAL4-VP16^[134], heat-shock and chemically-inducible systems^[131,146] and laser-induced gene expression^[131].

The morphological and genetic simplicity of the Marchantia gemma, along with the diversity of the genetic tools, were the main factors for choosing it as a model system for this work.

1.4 Aims

The biggest challenge on the way to engineering morphogenesis is to aggregate the data from various sources, e.g. growth dynamics, mechanical properties of a tissue, activity of genes, into a single model, and to identify the minimal set of key components and interactions that would be sufficient for understanding and controlling the development of plant shape.

Following Richard Feynman's quote, "What I cannot create, I do not understand", the ability to engineer plant morphogenesis not only offers an exciting opportunity to design living structures, but would also test our understanding of the complex mix of biological, chemical and mechanical processes behind the development of plant shape.

The main goal of this work is to establish methods and approaches that would allow one to link key developmental processes across the levels of biological organisation in Marchantia, from genes to cells and from cells to tissue. The key objectives are to:

- 1. Develop a novel data model for the Marchantia genome** Although the Marchantia genome is publicly available, it is based on a physical representation of the genome, rather than functions encoded into it. Because of this, it is not efficient for the extraction of genetic parts and high-throughput analysis of annotation data. A computational tool for the storage, mining and export of Marchantia genetic parts would simplify the design of synthetic constructs and enable access to the genomic data for multi-scale models of morphogenesis.

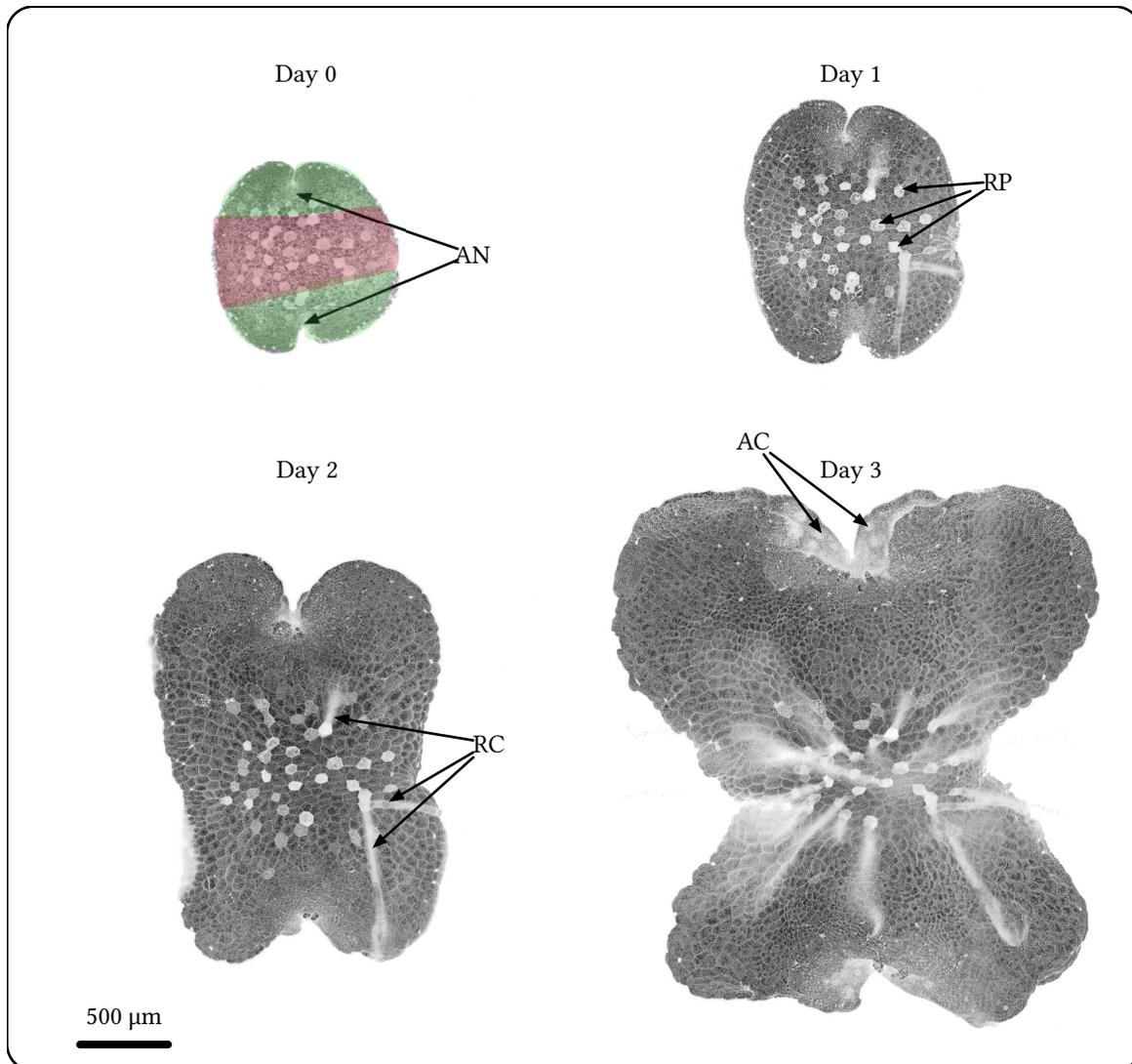


Figure 1.3: *M. polymorpha* gemma development. The figure shows four confocal microscopy images of a gemma during first three days of development. During the first two days, the cells of the apical region (green) expand and divide, with the resultant growth being polarised along the axis connecting the two apical notches (AN). The rhizoid precursor cells (RP) start developing into rhizoid cells after the day one on both dorsal and ventral surfaces of the gemma. By the day three, gemma splits in dorsal-ventral orientation and the repetitive cellular structures, air chambers (AC), start developing on the internal walls of the split.

In order to establish such tools, the *Marchantia* genome should be reduced to the collection of genes and corresponding functional annotations. Furthermore, splitting the gene records into the functional elements, i.e., promoters, untranslated regions, coding sequences and terminators, would create a direct link with the standard representation of plant DNA parts^[92] for the design of synthetic gene constructs.

2. Design fluorescent constructs for recording genetic, cellular and tissue-wide processes A database of *Marchantia* genetic parts would facilitate the extraction of parts and the assembly of fluorescent constructs for tracking the genetic, cellular and tissue-wide processes of *Marchantia* gemmae. A particular interest is the design of reporters for genes involved in auxin signalling and cell wall modification.

In order to assemble these constructs, the promoter regions of the genes of interest can be amplified from the *Marchantia* genomic DNA, and domesticated for Type IIS restriction enzyme assembly^[92,147]. These promoters can then be fused to the coding sequences of nuclear-localised fluorescent proteins and reintroduced into the *Marchantia* genome, creating a visual representation of the expression patterns of the target genes.

3. Develop imaging and image-processing tools Although the methods for sample preparation, imaging and quantitative analysis of gemmae development have been previously reported^[130,131] simultaneous recording of tissue dynamics and gene expression has not yet been achieved. A major practical challenge is to establish an assay for continuous time-lapse imaging of gemmae, which has previously been limited to less than 12 *h*; an analytical challenge is to develop methods for the comparison of gene expression and tissue growth patterns from different gemmae.

The practical challenge can be addressed by developing a growth chamber specialised for imaging purposes. It should be small enough to fit under a microscope, have a sufficient supply of water and nutrients, and prevent gemmae movement throughout the period of a time-lapse acquisition. In order to standardise the description of gene expression and tissue expansion patterns in gemmae, the patterns can be represented relative to the closest apical notch and normalised by the distance between the two notches^[130]. This will allow comparison and calculation of average patterns from several recorded samples. In order to do this automatically for every frame of a time-lapse image, a notch tracking algorithm should be developed.

4. Measure and correlate patterns of fluorescent reporters with patterns of resultant growth The development of continuous imaging assays will enable the simultaneous measurement of both the dynamic patterns of tissue expansion and the output of the fluorescent reporters. Such observations will allow verifying whether the underlying genetic and cellular processes, are sufficient to explain and predict the kinetics of tissue expansion.

Chapter 2

Materials and Methods

This chapter contains information about the main materials and methods used for this work. Online supplementary information and files can be found at the GitHub repository^[148]:

<https://github.com/HaseloffLab/MihailsDelmansThesis>

2.1 Marchantia methods

2.1.1 Chassis

Marchantia plants (male *Cam-1* and female *Cam-2*) were isolated in Cambridge by Prof. Jim Haseloff, and were used for the work presented in this thesis.

Marchantia is a relatively new plant model system with actively developing methods and community. Therefore, the complete protocols for the most critical methods were summarised in the Appendix A: Marchantia protocols, and are referenced in the main text of this chapter.

2.1.2 Tissue culture

Marchantia plants/spores were grown on solid 1 % (w/v) agar 1/2 Gamborg B5 medium (Duchefa Biochemie, 0.51 mM $CaCl_2 \cdot 6H_2O$, 12.36 mM KNO_3 , 0.5 mM $MgSO_4$, 0.55 mM NaH_2PO_4 , 0.5 mM $(NH_4)_2SO_4$) in 90 × 15 mm petri dishes sealed with 3M 1.25 cm Micropore tape (Protocol 1.1). Approximately once a month, the explants were transferred to fresh media plates.

Gemmae were propagated to a fresh plate using micropipette with 20 μ l tip. The tip was filled with sterile water, and one or several gemmae were picked up from the gemma cup and transferred to the surface of the solid media with a drop of water formed by pushing the pipet plunger. The plants were grown under fluorescent lamps with a 24 h light cycle at 22 – 24 °C.

2.1.3 Production of spores

In order to produce spores for *Agrobacterium*-mediated transformation, apical regions ($\sim 1\text{ cm}^2$) of *Marchantia Cam-1* and *Cam-2* thalli were excised and grown on Levington Advance M2 Potting & Bedding Compost (ICL) in trays with transparent lids, segregated by sex. The plants were grown inside growth cabinets under white LED lamps with 16/8 h light–dark cycle at 23°C and irrigated daily.

After two weeks of growth, the trays were placed under white LED lights supplemented with far-red LEDs (Duchefa GreenPower LED) to induce the development of antheridiophores and archegoniophores.

After a further two weeks, *Marchantia* sperm was collected from the young $\sim 3\text{ cm}$ tall antheridia. A drop of water was placed on top of the antheridia and collected after 1 – 2 min of soaking in a 15 mL Falcon tube. The collected sperm was diluted 1 : 3 with sterile water, and spread over the female *Cam-2* plants using a finger sprayer. The collection of the sperm and insemination of the female plants were repeated three times over nine days.

Three weeks later, the emerged sporangia of the female plants (spore-heads) were cut and collected in 50 mL Falcon tubes half-filled with self-indicating silica gel (Fisher Chemical), left to dry at room temperature for two days and transferred to a -80°C freezer for storage.

2.1.4 *Agrobacterium*-mediated transformation

The original protocol for transforming *Marchantia* spores^[137] modified by the members of the Haseloff Lab (Nuri Purswani, Bernardo Pollak, Susana Sauret-Gueto, Linda Silvestri, Mihails Delmans) was used to introduce synthetic DNA constructs into the *Marchantia* genome.

On day one, the spores of *Marchantia* were extracted from the sporangia, sterilised, plated on 1 % (w/v) agar 1/2 Gamborg B5 medium (each plate containing spores from individual sporangia) and left to germinate in the inverted orientation to promote rhizoid growth away from the solid medium to make it easier to scrape it from the plate during the following protocol steps (Protocol 2.1). On the same day, *Agrobacteria* GV2260 carrying *pSoup* plasmids^[149] were transformed with a binary vector, containing the level 2 construct of interest (Protocol 3.1, protocol 3.2). The transformed *Agrobacteria* were plated on solid LB medium (Protocol 1.4) supplemented with rifampicin, carbenicillin, tetracycline and spectinomycin, and left in a 30°C incubator.

On day four, a single colony was picked from the transformed *Agrobacterium* plates using a sterile loop, and used to inoculate 5 mL of the liquid LB medium (Protocol 1.4) supplemented with rifampicin, carbenicillin, tetracycline and spectinomycin (Protocol 1.2) in 14 mL Falcon round-bottom tubes. The tubes were placed into a shaking 28°C incubator at 150 rpm.

On day six, the liquid *Agrobacterium* cultures were spun down for 15 min at 3000 g. The supernatant was discarded, and the bacterial pellets were resuspended in 5 mL of 1/2 Gamborg B5 medium supplemented with sucrose, N-Z amine A, L-glutamine (Gamborg B5 Supplemented, protocol 1.3) and 100 μM final concentration of acetosyringone (5 μL of 100 mM stock in DMSO). The tubes were then returned to the shaking incubator for 6 h. At the end of the sixth hour, the spores were collected from the plates and resuspended in 7.6 mL of 1/2 Gamborg B5 Supplemented medium per plate (Protocol 2.2). The suspension of spores was aliquoted into a six-well plate, 3.8 mL per

well, and supplemented with $100\ \mu\text{M}$ of the final concentration of acetosyringone ($4\ \mu\text{L}$ of $100\ \text{mM}$ stock in DMSO). The *Agrobacterium* suspensions were removed from the shaking incubator and added to the six-well plate, $200\ \mu\text{L}$ per well. The six-well plate was sealed with the micropore tape and left under continuous white light on a shaking incubator at $200\ \text{rpm}$, $22 - 24\ ^\circ\text{C}$.

On day eight, the spores, co-cultured with *Agrobacterium*, were washed with 1/2 Gamborg B5 Supplemented medium + cefotaxime by adding $4\ \text{mL}$ of the media into each well and mixing by pipetting until a homogeneous consistency was achieved. The spores were then separated from the *Agrobacterium* by filtering through a $70\ \mu\text{m}$ strainer (EasyStrainer, Greinerbio-one) into a $50\ \text{ml}$ Falcon tube. Finally, the spores were transferred into 1% (w/v) agar 1/2 Gamborg B5 plates supplemented with hygromycin by inverting the strainer and flushing with $1\ \text{mL}$ of the 1/2 Gamborg B5 Supplemented medium. After a week of selection on hygromycin, the transformants were examined under a Leica M205 FA fluorescent microscope (Subsection 2.3.2). Twelve living sporelings with a strong fluorescence signal of expected wavelengths and localisation were transferred to fresh 1% (w/v) agar 1/2 Gamborg B5 medium plates and grown under tissue culture conditions (Subsection 2.1.2).

2.1.5 Surgical assay

For surgical assay, young gemmae were removed from the gemma cups and transferred to a 1% (w/v) agar 1/2 Gamborg B5 plate. From there, individual gemmae were moved onto the plastic lid of a $65\ \text{mm}$ petri dish set up under a Leica stereo microscope. The apical regions of the gemmae were excised using a scalpel (Swann-Morton) using a *Nr.10* surgical blade (Swann-Morton). The separated apical and central regions were then transferred onto a 1% (w/v) agar 1/2 Gamborg B5 plate or into the well of a growth chamber (Subsection 2.3.1) using a $0.1\ \text{mm}$ tungsten needle.

2.2 Construct Design

The synthetic DNA constructs for transforming *Marchantia* were built using Loop Assembly^[147], a variant of a Type IIS assembly. The assembly mix is prepared by combining the DNA parts with restriction enzymes and DNA ligases in a single tube. The tube is then loaded into a thermocycler, which cycles between $37\ ^\circ\text{C}$, favourable for the restriction enzymes, and $16\ ^\circ\text{C}$, favourable for the DNA ligases. During the $37\ ^\circ\text{C}$ phases, the restriction enzymes ‘chew off’ the ends of the DNA parts exposing the 4 bp-long overhangs that are identical between the ends of consecutive parts, hence allowing the DNA parts to bind to each other. During the $16\ ^\circ\text{C}$ phase, the DNA ligase joins the bound ends of the DNA parts. The two DNA constructs that were eventually used to transform the *Marchantia* spores were assembled in two steps. First, the genetic parts (level 0 parts) – promoters combined with 5’ untranslated regions (PROM5); coding regions (CDS); and terminators combined with 3’ untranslated regions (TERM3) – were combined into gene assemblies (level 1 parts). The assembly schema, compliant with the common syntax for plant DNA assembly, was used to standardise the overhangs between level 0 parts (Figure 2.1). Then, the gene assemblies were combined into the multi-gene constructs (level 2).

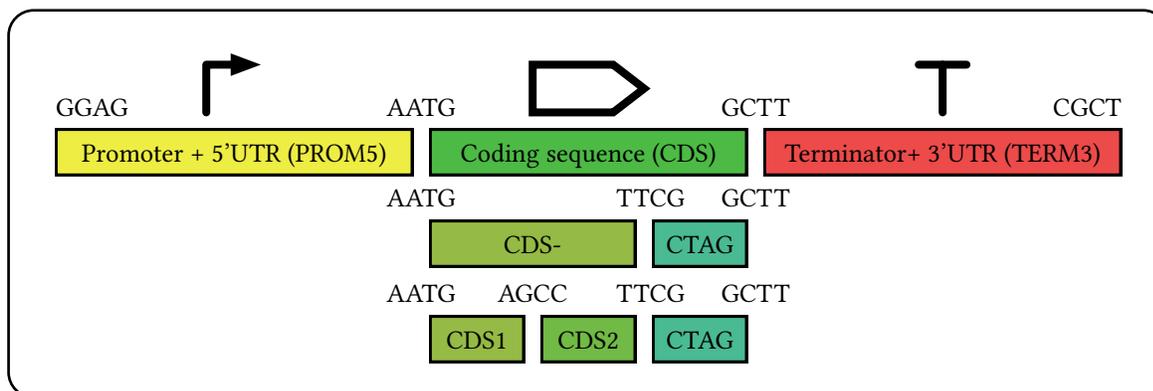


Figure 2.1: **Assembly schema.** Rectangles denote different genetic, level 0, parts, the nucleotide quadruplets denote standardised overhangs for the Type IIS assembly, exposed by the *BsaI* restriction enzyme digestion.

The summary of the parts is presented in table 2.1, while their sequences and plasmid maps can be found in the online supplementary resources^[148,150].

The level 0 parts (Table 2.1: Level 0) were either taken from the previous work of Bernardo Pollak^[131,147], amplified from the *Marchantia* genome (Protocol 4) or produced via DNA synthesis.

2.2.1 Domestication of parts amplified from the *Marchantia* genome

The sequences of the promoters of the *MpARF3* and *MpYUC2* genes were extracted from MarpoDB (Chapter 5) and used to design primers for the amplification of the 2 kb regions upstream from the translation start site and including 40 bp homologous overhangs with the cloning vector. The primers also included the sequence of the *BsaI* recognition sites and the corresponding overhangs (Table 2.2). The primers were used to domesticate parts for the Loop Assembly by PCR amplification from the genomic DNA (gDNA) (Protocol 4).

The amplicons were checked to have the expected size by gel electrophoresis, purified using a Gel Extraction Kit (Qiagen) and cloned into the modified *pUDP2(BBa_P10500)*^[147] plasmid by the Gibson assembly (Protocol 5.1).

2.2.2 Assembly of the Level 1 and Level 2 constructs

Level 0 parts were assembled into the gene constructs using the Loop Assembly and further into level 2 multi-gene constructs (Protocol 5.2, table 2.1: Level 1, Level 2).

The parts were cloned into *pOdd*^[147] or *pCk* plasmids for level 1 assemblies and *pEven*^[147] or *pCs* plasmids for level 2 assemblies. *pCk* plasmids are derivatives of *pDGB3_Omega1*^[158] with a modified selection marker, restriction enzyme sites and overhangs matching the specification of the Loop assembly level 1 plasmid. They were built and kindly provided by the members of the Patron Group (Earlham Institute). *pCsA* plasmid is a derivative of the *pCk* plasmid with modified restriction enzyme sites and overhangs matching the specification of the Loop assembly level 2

Level 0

Part name	Type	Description	Plasmid
<i>proMpEF1α</i>	PROM5	Promoter of the Marchantia translation elongation factor <i>MpEf1α</i> that encodes a protein essential for translation ^[151] (mpdb.gene.56948/Mapoly0024s0116)	<i>pL0_MpEF1α</i> ^[LA]
<i>mTurquoise2</i>	CDS-	A variant of a cyan fluorescent protein ^[152]	<i>pL0_mTurquoise2</i> ^[LA]
<i>Venus</i>	CDS-	A variant of a yellow fluorescent protein ^[153]	<i>pL0_Venus</i> ^[LA]
<i>N7</i>	CTAG	A nuclear-localisation tag ^[154]	<i>pL0_N7</i> ^[LA]
<i>nosT-35ST</i>	TERM3	A fusion of the <i>nosT</i> ^[155] and <i>35sT</i> ^[149] terminators	<i>pL0_nosT-35ST</i> ^[LA]
<i>eGFP-Lti6b</i>	CDS	Membrane-localised green fluorescent protein ^[156]	<i>pL0_eGFP-Lti6b</i> ^[LA]
<i>hygR</i>	CDS	Hygromycin B resistance marker ^[149]	<i>pL0_HygR</i> ^[LA]
<i>pro35S</i>	PROM5	35S promoter of a Cauliflower mosaic virus ^[157]	<i>L0-35S</i> ^[BP]
<i>proMpARF3</i>	PROM5	Promoter of the Marchantia auxin response gene <i>MpARF3</i> (mpdb.gene.51957/Mapoly0043s0098)	<i>pL0_prom5ARF3</i> ^[gDNA]
<i>proMpYUC2</i>	PROM5	Promoter of the Marchantia gene encoding an enzyme involved in auxin biosynthesis (<i>MpYUC2</i> , mpdb.gene.8238, Mapoly0063s0040)	<i>pL0_prom5YUC2</i> ^[gDNA]
<i>DII</i>	CDS1	A synthetic part containing DII region of <i>MpAUX/IAA</i> gene (mpdb.gene.52065, Mapoly0034s0017)	<i>pL0_cds1DII</i> ^[syn]
<i>VenusCDS2</i>	CDS2	A variant of a yellow fluorescent protein ^[153] with a poly-A N-terminus linker	<i>pL0_cds2Venus</i> ^[syn]

Level 1

Construct	Backbone	PROM5	CDS / CDS1	CDS2	CTAG	TERM3
<i>35s:HygR</i>	<i>pOdd-1</i> ^[LA]	<i>pro35s</i>	<i>hygR</i>	-	-	<i>nosT-35ST</i>
<i>ARF3:Venus</i>	<i>pOdd-4</i> ^[LA]	<i>proMpARF3</i>	<i>Venus</i>	-	<i>N7</i>	<i>nosT-35ST</i>
<i>YUC2:mTurquoise2</i>	<i>pOdd-2</i> ^[LA]	<i>proMpYUC2</i>	<i>mTurquoise2</i>	-	<i>N7</i>	<i>nosT-35ST</i>
<i>EF1:eGFP-Lti6b</i>	<i>pCk-4</i> ^[*]	<i>proMpEF1</i>	<i>eGFP-Lti6b</i>	-	-	<i>nosT-35ST</i>
<i>ARF3:DII-Venus</i>	<i>pCk-3</i> ^[*]	<i>proMpARF3</i>	<i>DII</i>	<i>Venus</i>	<i>N7</i>	<i>nosT-35ST</i>
<i>ARF3:mTurquoise</i>	<i>pCk-2</i> ^[*]	<i>proMpARF3</i>	<i>mTurquoise</i>	-	<i>N7</i>	<i>nosT-35ST</i>

Level 2

Construct	Backbone	Part 1	Part 2	Part 3	Part 4
<i>mDYUC2</i>	<i>pEven-1</i> ^[LA]	<i>35s:HygR</i>	<i>YUC2:mTurquoise2</i>	<i>L1.3-EF:GLT</i> ^[BP]	<i>ARF3:Venus</i>
<i>mDDII</i>	<i>pCsA</i> ^[*]	<i>L1.1-RKD:Hyg</i> ^[BP]	<i>ARF3:mTurquoise2</i>	<i>ARF3:DII-Venus</i>	<i>EF1:eGFP-Lti6b</i>

Table 2.1: **List of parts and constructs.** The source of the plasmid is shown in square brackets. LA - plasmids from the Loop Assembly collection^[147]; BP: plasmids from the work of Bernardo Pollak^[131]; gDNA: parts were amplified from the Marchantia genomic DNA; syn: parts obtained through DNA synthesis service.; *: *pDGB3_Omega1*^[158] - derived plasmids build by the members of the Patron Group, Earlham Institute (*pCk*) and Haseloff Lab, University of Cambridge (*pCsA*).

Part name	Sequence	Orientation	$T_m, ^\circ C$
<i>proMpARF3</i>	AATTCGCGGCCGCTTCTAGAGCGATGAGT <u>GGTCTC</u> AggagGTAGTTGGGAAGCCATGC	Forward	53.5
<i>proMpARF3</i>	TGCAGCGGCCGCTACTAGTCGCGATGAGT <u>GGTCTC</u> GcattCCCAAACAGAAAGGAGCC	Reverse	53.3
<i>proMpYUC2</i>	AATTCGCGGCCGCTTCTAGAGCGATGAGT <u>GGTCTC</u> AggagTACCTCGGACGGTTCT	Forward	54.2
<i>proMpYUC2</i>	TGCAGCGGCCGCTACTAGTCGCGATGAGT <u>GGTCTC</u> GcattTGTCGACGAAATCAGCA	Reverse	51.7

Table 2.2: **Primers for genomic PCR.** Underlined sequence denotes the *BsaI* recognition site; lowercase sequence denotes the assembly overhang sequence; highlighted sequence is homologous to the left and right borders of the insert site of the modified *pUDP2(BBa_P10500)* plasmid^[147]; unformatted sequence corresponds to the homologous region of the genomic DNA. T_m denotes the melting temperature of the primer.

plasmid. They were built and kindly provided by the members of the Haseloff Lab, Susana Sauret-Gueto and Eftychis Frangedakis (Department of Plant Sciences, University of Cambridge).

The assembly reactions were used to transform *E. coli TOP10* cells (Protocol 3.3). After transformation, the cells were plated on LB agar plates supplemented with kanamycin for level 1 and spectinomycin for level 2 (Protocol 1.2, protocol 1.4), and incubated at $37^\circ C$ overnight.

Two colonies were picked per plate and grown in 5 mL of liquid LB media with the corresponding antibiotic in 14 mL Falcon round-bottom tubes overnight in a $37^\circ C$ incubator shaking at 120 rpm . The plasmid DNA was extracted from the cell cultures using QIAcube (Qiagen) following the manufacturer’s instructions. The concentration of the extracted plasmids was measured and $1\ \mu\text{g}$ of the DNA was digested using a FastDigest enzyme (Thermo Fisher Scientific) in a FastDigest Green Buffer (Thermo Fisher Scientific) following the manufacturer’s protocol. A restriction enzyme that was expected to cut the plasmid in at least three parts was chosen. After digestion, the DNA was subject to gel electrophoresis using 1% (w/v) agarose gel together with the 1 kb DNA Ladder (HyperLadder, Bionline) for reference. The digestion pattern was compared to the one expected to verify the success of the assembly.

2.3 Microscopy

2.3.1 Sample preparation

For the time-lapse imaging a custom chamber (Figure 3.1A) was designed using OpenSCAD, and 3D-printed using Ultimaker 2 (Ultimaker). The chamber was cuboid in shape ($36 \times 25 \times 10\text{ mm}$) with $24\ 3.5 \times 3 \times 9\text{ mm}$ wells in 6×4 array. The design files can be found in the online supplementary resources^[148].

Two adhesive $125\ \mu\text{L}$ Gene Frames (Thermo Fisher Scientific) were attached to the top and bottom of the chamber. The bottom of the chamber was attached to a $75 \times 26\text{ mm}$ microscope slide. The wells of the chamber were filled with $\sim 90\ \mu\text{L}$ of molten 3% (w/v) agar 1/2 Gamborg B5 medium (Protocol 1.1) and allowed to cool for 30 min . *Marchantia gemmae* were picked from the gemma cups of the mature plants and transferred to the wells of the chamber, one per well. Finally, the chamber was sealed with a $0.16 - 0.19\text{ mm}$ -thick $22 \times 50\text{ mm}$ coverslip pretreated with B-Clean anti-fog spray (Bollé Safety).

For time-course imaging, gemmae were grown on the 1% (w/v) agar 1/2 Gamborg medium in 65 mm gridded

plates with closed lids sealed with Micropore Medical Tape (3M). The lid was removed during imaging.

2.3.2 Fluorescent stereo microscopy

Stereo-microscopic observations were made using a Leica M205 FA fluorescent microscope. The set of filters is summarised in table 2.3. The images were captured using a Leica DFC465 FX camera using LAS X software.

Signal	Filter excitation wavelength (band centre/band width)	Filter emission band (band centre/band width)
GFP	470/40 nm	525/50 nm
CFP	436/20 nm	480/40 nm
YFP	500/20 nm	535/30 nm
Chlorophyl autofluorescence	480/40 nm	610 nm long pass

Table 2.3: Overview of Leica M205 FA filters

2.3.3 Laser-scanning confocal microscopy

A Leica SP8 confocal system based on a DM6 microscope and configured with a white light laser, 442 nm laser diode (Leica CTR6 DIODE), four 8-bit hybrid detectors and HC PL APO 10x 0.4 CS2 DRY objective was used for the confocal microscopy.

For multispectral 3D imaging of the gemmae, z-stacks with 5 μm step were collected using sequential imaging (Table 2.4). Three-dimensional stacks were reduced to two-dimensional images using maximal intensity projection along the z-axis. The resultant multispectral, two-dimensional time-lapses were used to calculate the average tissue expansion rates and the average signals of the *Marchantia* lines transformed with *mDYUC2* and *mDDII* (Table 2.1: Level 2).

Sequence	Excitation wavelength, nm	Emission collection band, nm	Fluorophore	Channel
1	470	501 - 519	eGFP	Membrane
	470	650 - 680	Chlorophyll	Chlorophyll
2	422	465 - 483	mTurquoise2	CFP
3	518	570 - 588	mVenus	YFP

Table 2.4: Summary of image acquisition settings.

2.4 Image processing

2.4.1 Image-processing pipeline

The acquired multispectral time-lapses, containing recordings of individual gemmae during the first 50 *h* of development, were passed through an image-processing pipeline (Figure 2.2). The aim of the pipeline was to find the correlations between the patterns of tissue expansion and the patterns of gene expression.

The time-lapse processing pipeline was implemented in Python^[159] using the generic image-processing functions of the SimpleITK library^[160] and the image registration methods of the SimpleElastix library^[161]. The source code of the Python scripts that implement the pipeline is available in the online supplementary resources^[148] in the *timelapseProcessing* directory.

The pipeline consisted of three steps. During the first step (Figure 2.2 A), the signal from the *membrane* channel was used for estimating tissue transformations between the consecutive time frames of the time-lapse using the image registration methods of SimpleElastix (Algorithm IP.2). Noise in the image background can negatively impact the image registration algorithm. In order to exclude the background noise, the signal from the *chlorophyll* channel was blurred and converted into a *binary mask* (Algorithm IP.1), giving an estimate of the image foreground. The YFP and the CFP channels were used for detection of the nuclei and sampling the expression levels of the nuclei-localised fluorescent markers (Algorithm IP.3).

These image-processing steps were applied iteratively for each time frame of a time-lapse. The calculated masks, transformations and signal samples were grouped together with the original channels into a single processed package for each original time-lapse. Since the first step of the pipeline was the most time-consuming, preparation of these processed packages allowed the saving of time while optimising the remaining steps of the pipeline.

During the second step (Figure 2.2 B), the calculated transformations were used to estimate the velocity fields of the gemma tissue. This in turn allowed for the calculation of the tissue expansion rate as a divergence of the velocity field (Algorithm IP.4).

In order to aggregate the patterns of tissue expansion and gene expression from several time-lapses the coordinate frames in the time-lapses should be aligned with a biologically-relevant origin. In this pipeline, the coordinate frames were then transformed relative to the positions of the apical notches and normalised by the size of a gemma at a corresponding time frame (Algorithm IP.5). This allowed for the representation of the patterns of gemma growth and gene expression in relation to the distance from the apical notches. The coordinates of the apical notches were manually assigned at the first time frame and calculated using the estimated velocity field for every consecutive time frame.

Finally, during the last step of the pipeline, the normalised patterns of the reporter signals and relative tissue expansion rates, collected from six gemmae of *mDYUC2* and six of *mDDII*, were averaged to produce the mean patterns (Figure 2.2 C).

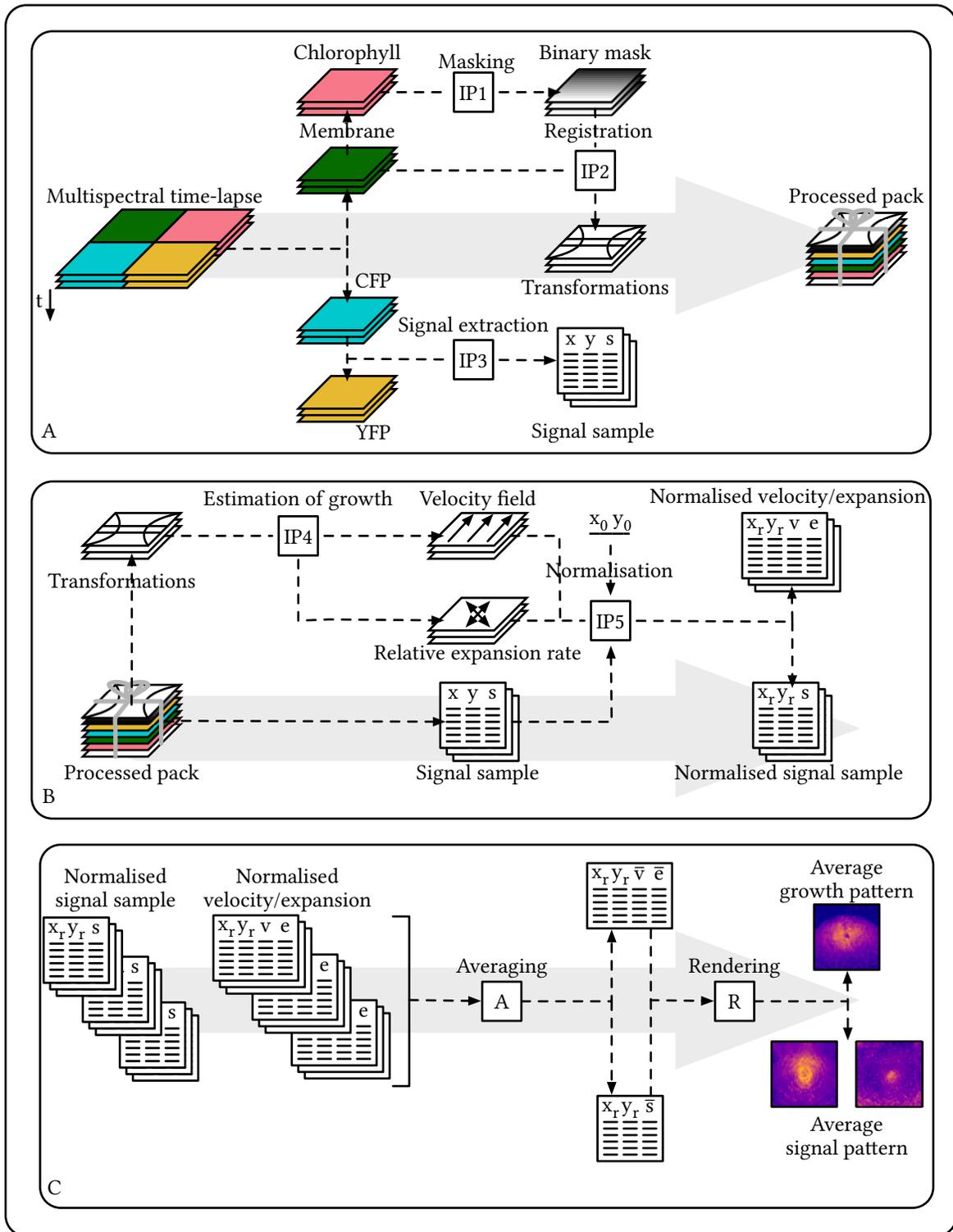
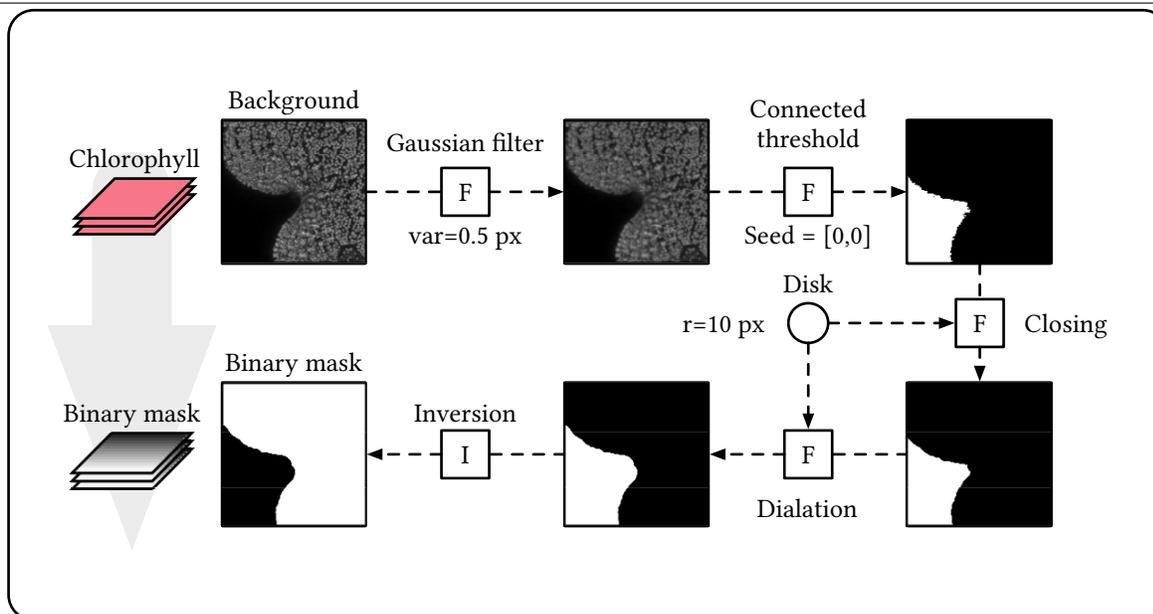


Figure 2.2: **Summary of the image-processing pipeline.** **A.** Processing step. **B.** Normalisation step. $x_0 y_0$ correspond to the positions of the apical notches. **C.** Averaging step. The normalised samples originating from the gemmae of both *mDDII* and *mDYUC2* were averaged (A) and rendered (R). The pipeline follows the direction of arrows. Square stacks represent stacks of time frames. White squares with IPx correspond to image-processing algorithms, summarised in subsection 2.4.2.

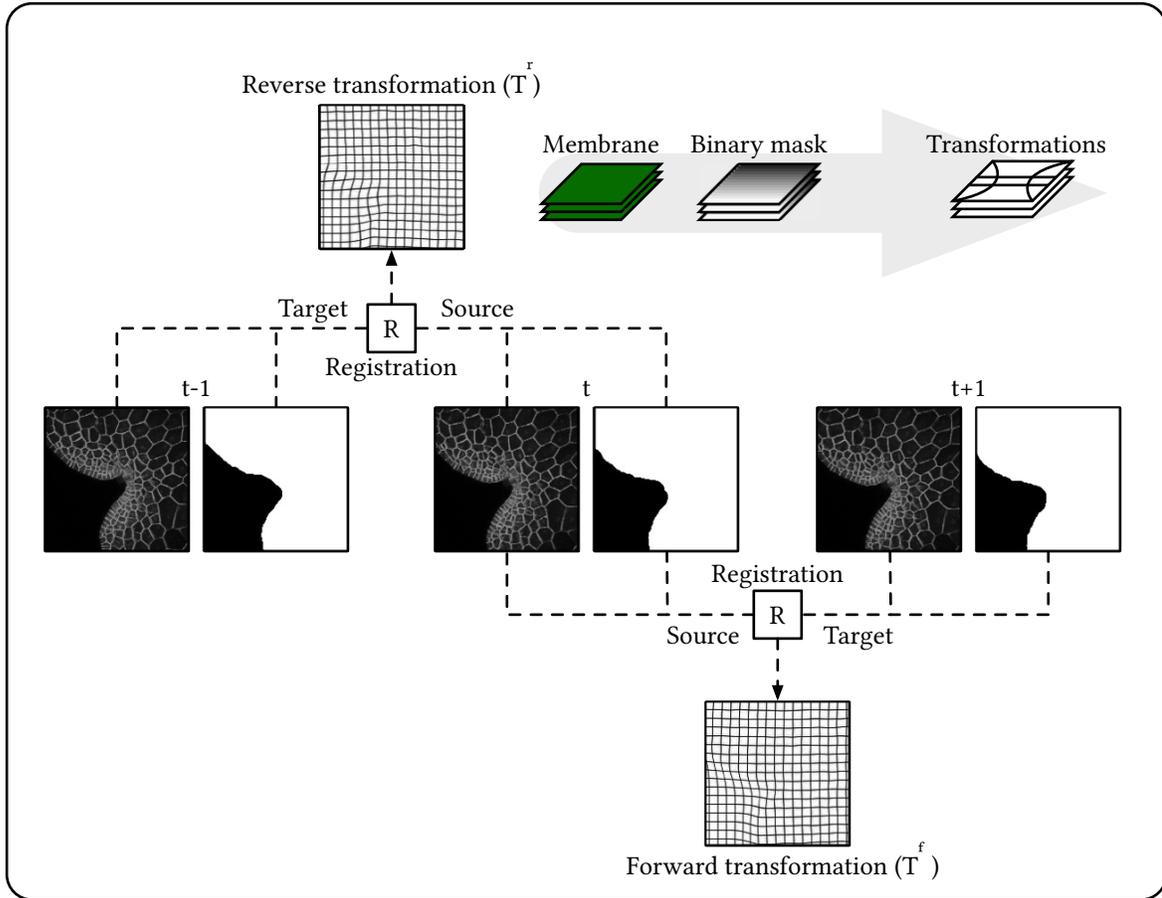
2.4.2 Image-processing algorithms

Algorithm IP.1 Masking



Chlorophyll channel is used to produce a binary mask separating the gemma tissue from the background. An image is processed with a Gaussian filter to eliminate noise in the image background. The background is then detected using a connected threshold filter that assigns the value 1 to all pixels with intensity values below 2, connected to an upper-left corner of the image. The application of a closing filter, followed by the application of a dilation filter, removes small artifacts and smoothes the edge of the background region. Finally, the image is inverted, thus pixels with value 0 denote background, while those with value 1, regions of interest. F denotes the SimpleITK filters, I denotes the signal inversion operator. Circles denote kernel elements used for the *Closing* and the *Dilation* operators.

Operator	Simple ITK filter (itk::simple)	Parameters
Gaussian filter	DiscreteGaussianImageFilter	MaximumKernelWidth: 32; Variance: 0.5
Connected threshold	ConnectedThresholdImageFilter	Lower: 0; Upper: 1; ReplaceValue: 1; Connectivity: 0; Seed list: [0,0]
Closing	BinaryMorphologicalClosingImageFilter	ForegroundValue: 1; SafeBorder: 1; KernelRadius: [10, 10, 10]; KernelType: Ball
Dilation	BinaryDilateImageFilter	BackgroundValue: 0; ForegroundValue: 1; BoundaryToForeground: 0; KernelRadius: [10, 10, 10]; KernelType: Ball
Inversion	$(i) \rightarrow 1 - i$	-

Algorithm IP.2 Registration

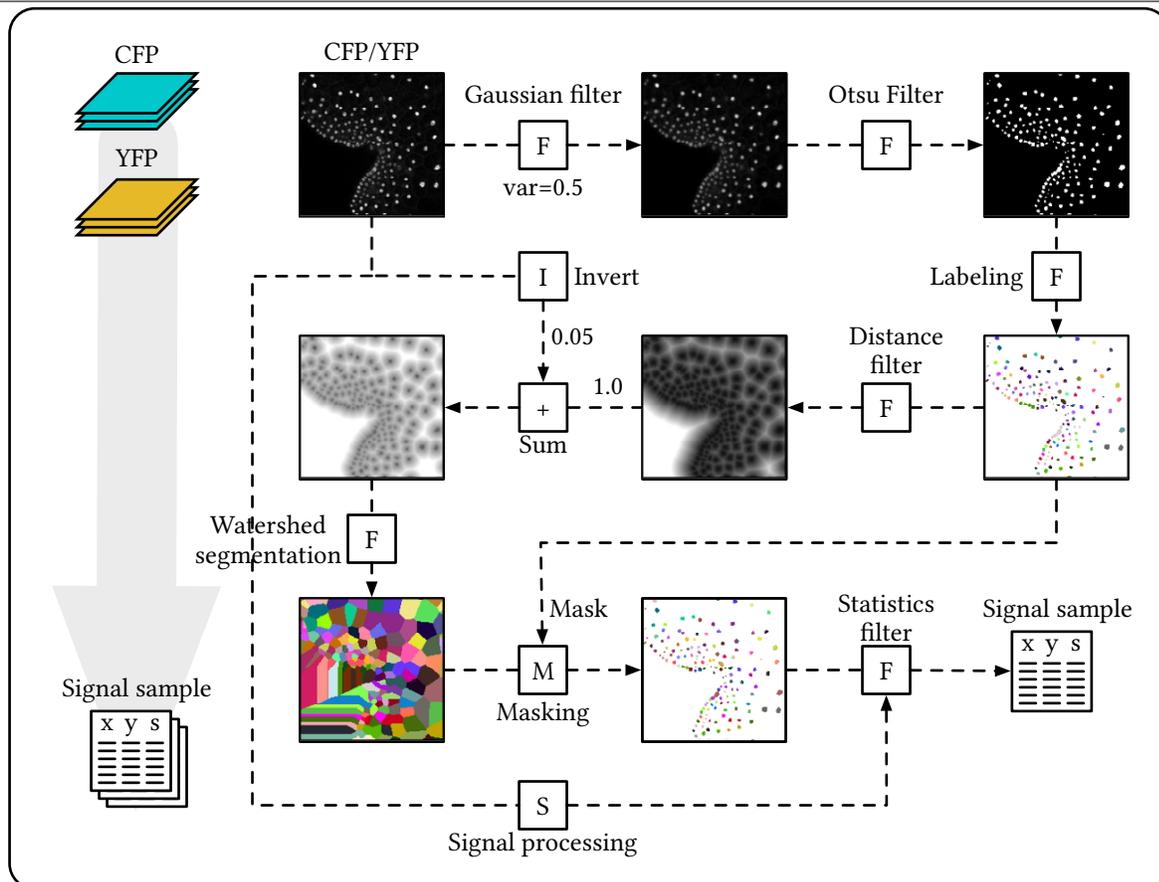
An image registration algorithm of the SimpleElastix package is used to estimate forward and reverse transformations (T^f and T^r). R denotes the registration operator. The binary images are used to exclude the image background from the calculation of the transformations. Image registration works by finding a transformation function T that should be applied to a source image $I_s(\mathbf{x})$, such that it minimises the difference with the target image $I_t(\mathbf{x})$. Mattes mutual information between the images was used as an objective function to minimise during solving an optimisation problem. The transformation functions were constrained to B-spline transformations with an additional bending energy penalty term. The stochastic gradient descent algorithm was used to optimise the transformations. The forward transformation contains information about displacement that should be applied to material points, sampled at each pixel location, for them to match the locations of the corresponding material points in the next time frame. The reverse transformation contains analogous information, but relative to the previous time frame.

Operator	Function	Parameters
----------	----------	------------

Registration	ElastixImageFilter	
--------------	--------------------	--

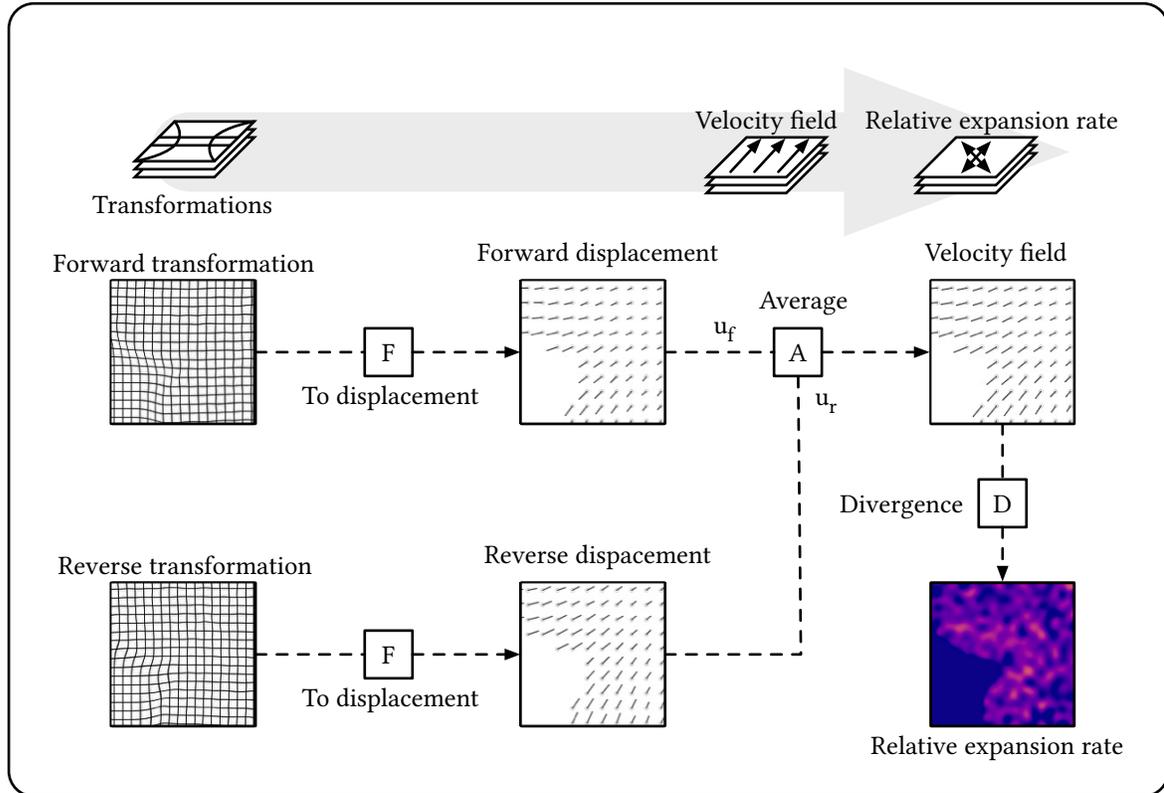
		DefaultPixelValue: 0; FinalBSplineInterpolationOrder: 3; FinalGridSpacingInPhysicalUnits: 8; FixedImagePyramid: "FixedSmoothingImagePyramid"; GridSpacingSchedule: 2.80322: 1.9881: 1.41: 1; ImageSampler: "RandomCoordinate"; Interpolator: "LinearInterpolator"; MaximumNumberOfIterations: 512; MaximumNumberOfSamplingAttempts: 8; Metric: "AdvancedMattesMutualInformation"; "TransformBendingEnergyPenalty"; Metric0Weight: 1; Metric1Weight: 1; MovingImagePyramid: "MovingSmoothingImagePyramid"; NewSamplesEveryIteration: "true"; NumberOfResolutions: 4; NumberOfSamplesForExactGradient: 4096; NumberOfSpatialSamples: 2048; Optimizer: "AdaptiveStochasticGradientDescent"; Registration: "MultiMetricMultiResolutionRegistration"; ResampleInterpolator: "FinalBSplineInterpolator"; Resampler: "DefaultResampler"; ResultImageFormat: "nii"; Transform: "BSplineTransform";
--	--	--

Algorithm IP.3 Signal extraction



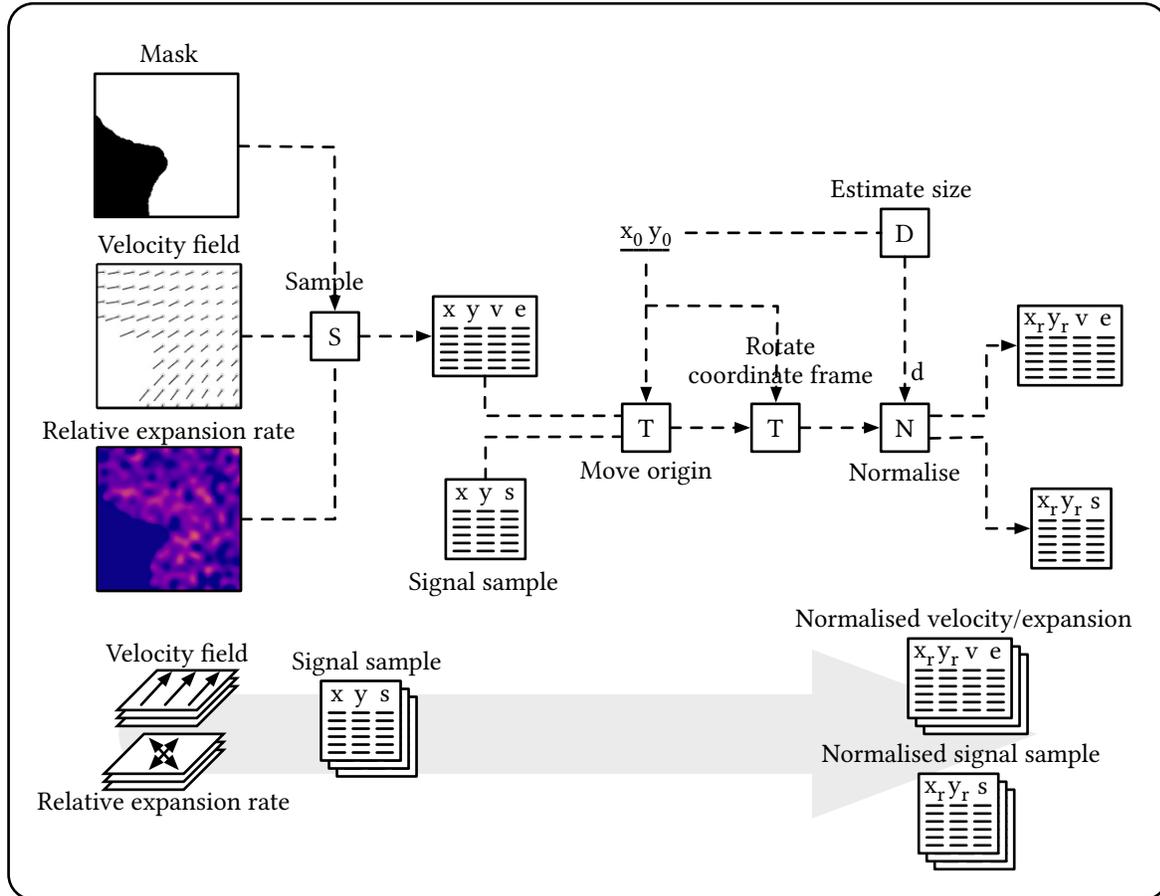
The nuclear-localised signal is used to sample the reading of the reporters (YFP channel for the *mDYUC2* construct and CFP channel for the *mDDII* construct) using the threshold-based segmentation and separation of joined nuclei through the distance filter. An image is smoothed by passing it through a Gaussian filter. It is then subject to a segmentation using an Otsu^[162] threshold. The result of the segmentation is a binary image, where pixels with value 1 correspond to nuclear regions. The binary image is then labelled by assigning a unique pixel value to each closed region, i.e. to every segmented nucleus. If nuclei in the original image are close together, they might be joined into a single region during the Otsu segmentation step. In order to segment apart these joined nuclei, an additional step, based on the watershed segmentation, is used. The image is subject to a distance filter, which assigns pixel value based on the distance to the closest pixel with value 1. The distance image is then combined with the original image multiplied by 0.05 to create sharper boundaries around the nuclear regions. Finally, the resultant image is subject to a watershed segmentation. The labels of the watershed segmentation are then transferred to the nuclear regions identified by the Otsu segmentation, resulting in the separation of joined nuclei. Finally, the average signal is recorded in the region of each segmented nucleus. *F* denotes the SimpleITK filters, *I* denotes the signal inversion operator, *M* denotes the masking operator, *S* denotes the signal processing operator.

Operator	Simple ITK function/filter	Parameters
Gaussian filter	DiscreteGaussianImageFilter	Variance: 0.5; MaximumKernelWidth: 32
Otsu filter	OtsuThresholdImageFilter	NumberOfHistogramBins: 128; MaskOutput: 1;
Labelling	ConnectedComponentImageFilter	FullyConnected: 0; ObjectCount: 0
Distance filter	SignedMaurerDistanceMapImageFilter	InsideIsPositive: 0; SquaredDistance: 0;
Watershed	MorphologicalWatershedImageFilter	Level: 1; MarkWatershedLine: 0; FullyConnected: 0
Masking	Mask	-
Statistics filter	LabelStatisticsImageFilter	-

Algorithm IP.4 Estimation of growth

The velocity field is calculated as a difference between the forward and reverse displacements, normalised by the time step (A). The forward displacement is the difference in material point locations between the following and the current time frame, i.e. $u_f(t) = x(t+1) - x(t)$. The reverse displacement is the difference in material point locations between the previous and the current time frame, i.e. $u_r(t) = x(t-1) - x(t)$. The divergence of the velocity field (D) is used to calculate the relative tissue expansion rate at each time frame. F denotes the SimpleITK filters.

Operator	Function/filter
To displacement	TransformToDisplacementFieldFilter
Average	$(u_f, u_r) \rightarrow \frac{u_f - u_r}{2dt}$
Divergence	$(v_x, v_y) \rightarrow \text{VectorIndexSelectionCast}(\text{Gradient}(v_x), 0) + \text{VectorIndexSelectionCast}(\text{Gradient}(v_y), 1)$

Algorithm IP.5 Normalisation

Each pixel in the velocity field and relative expansion rate channel, covered by the mask, is used as a sample (S). The coordinates of the sampled growth parameters are transformed relative to the position of the closest apical notch (x_0y_0) and rotated so that the x-axis matches the direction of the velocity field at the closest apical notch (T). The size of the gamma (d) is calculated as a distance between the two apical notches at the corresponding time frame (D). The coordinates of the sampled signal and velocity/expansion rates are normalised by the gamma size (N).

Operator	Function
Move origin	$(x, y) \rightarrow (x - x_0, y - y_0)$
Rotate coordinate frame	$(r, \theta) \rightarrow (r, \theta - \arg(\mathbf{v}(x_0, y_0)))$
Normalise	$(x, y) \rightarrow (x/d, y/d)$

2.5 MarpoDB

The core and the *server* modules of MarpoDB were written in Python^[159], while the *frontend* was implemented as a JavaScript^[163] web application. The source code of MarpoDB is hosted at GitHub^[164], while the public version of the

database is available at www.marpodb.io.

The database schema was implemented in a PostgreSQL^[165] database that allowed the records to be stored and queried using a Structured Query Language (SQL). Graphical representation of the tables (Figure 5.2) was generated from MarpoDB database schema using pgModeler software^[166].

For communicating with the PostgreSQL database, an SQLAlchemy^[167] library was used. For the DNA sequence operations, the Biopython^[168] library was used.

Visualisation of the gene models was implemented using Scribl^[169]; visualisation of the sequence using Sequence-viewer^[170]; copy to clipboard using clipboard.js^[171].

The *server* module was built with the help of the Flask^[172] library.

Isolated DNA and RNA libraries of Cam-1 were sequenced using the Illumina HiSeq 2000 platform and assembled *de novo* using Meraculous 2.0^[173] for genome, and Bridger^[174] for transcriptome. CDS prediction was accomplished using Transdecoder^[96]. The predicted CDSs were translated into the protein sequences and compared to a collection of *Viridiplantae* protein sequences, extracted from Uniprot^[175], using a protein BLAST^[176] algorithm. The BLAST hits with at least 35 % identity and 20 % coverage were considered positive. In addition, the translated CDSs were analysed using the InterProScan^[101] pipeline to find conserved protein motifs that pass gathering thresholds^[97]. The mapping of the transcript to the genome was done using Splign^[103].

Alignment of the genomic Cam reads against the Tak 3.1 genome assembly was done using the Burrows – Wheeler Aligner^[177]. Duplicated reads were removed from the alignment using MarkDuplicates tool from the Picard package^[178]. SNP and indels were called using The Genome Analysis Toolkit^[179] featuring a variant quality score recalibration algorithm^[180].

2.6 Simulating auxin transport

The simulation of the auxin transport model was implemented in Python. Partial differential equations were solved using the DOLFIN library^[181], released as a part of the FeniCS project. The source code is available in the online supplementary resources^[150] in the /morphosolver directory.

2.6.1 Generation of the gemma meshTo simulate the geometry of a day 0 gemma, a polygon with an outline governed by the following equation was extruded by 0.1 arbitrary units:

$$r = a \cdot \cos(\pm\theta + 0.5\pi)^{0.25} + b$$

with $a = 0.5, b = 0.7$.

2.6.2 Numerical solution of the reaction–diffusion equation

Reaction–diffusion of a chemical on the generated mesh 3 was simulated using the following numerical approach.

Let $c(\mathbf{x}, t)$ be the concentration of a chemical at time t and position \mathbf{x} inside a domain Ω , defined by the 3D mesh. Then the reaction–diffusion equation for the chemical can be described using the following partial differential equation:

$$\frac{\partial c(\mathbf{x}, t)}{\partial t} = D\nabla^2 c(\mathbf{x}, t) + R(c, t)$$

where D is a diffusion coefficient and R is a generic reaction term.

The reaction–diffusion equation can be solved by applying an implicit Euler discretisation in time, leading to the following form:

$$\frac{c(t) - c(t - \delta t)}{\delta t} = D\nabla^2 c(t) + R(c(t), t)$$

where δt is an arbitrary time step.

In order to solve the discrete equation using the finite element method, it should be presented in a variational form. Setting $c(\mathbf{x}, t)$ as an unknown trial function, the variational form can be obtained by multiplying both sides of the discretised equation by a test function $e(x)$ and integrating both sides over the domain Ω :

$$\int_{\Omega} ec(t) - ec(t - \delta t)dx = \int_{\Omega} eD\nabla^2 c(t)dx + \int_{\Omega} eR(c(t), t)dx$$

where dx is a differential element for the domain Ω . In order to get rid of the second-derivative in the diffusion term, integration by parts can be applied, leading to the following form:

$$\int_{\Omega} ec(t) - ec(t - \delta t)dx = \int_{\Omega} D\nabla e \cdot \nabla c(t)dx - \int_{\partial\Omega} \nabla \cdot \mathbf{n}c(t)ds + \int_{\Omega} eR(c(t), t)dx$$

where ds is a differential element for the domain boundary $\partial\Omega$, and \mathbf{n} is a vector perpendicular to the domain boundary. Assuming there is no diffusion allowed across the domain boundary, $\int_{\partial\Omega} \nabla \cdot \mathbf{n}c(t)ds = 0$ and the problem can be set as finding $c(t)$ such that

$$\int_{\Omega} ec(t) - ec(t - \delta t)dx - \int_{\Omega} D\nabla e \cdot \nabla c(t)dx - \int_{\Omega} eR(c(t), t)dx = 0$$

For the simulation in chapter 6, the above equation was solved iteratively for $c(t)$ using the DOLFIN library^[181] starting with an initial condition $c(\mathbf{x}, 0)$. The pseudocode for the one iteration step is presented in algorithm Algorithm 1, while the pseudocode for the simulation of auxin transport is presented in algorithm Algorithm 2. The detailed description of these methods can be found in the FeniCS book^[182].

Algorithm Algorithm 1 Update of concentration

```

function UPDATE( $c_0, D, R, \delta t$ )
   $c \leftarrow$  trialFunction
   $e \leftarrow$  testFunction
   $derivativeTerm \leftarrow e * (c - c_0)$ 
   $diffusionTerm \leftarrow dt * D * inner(grad(c), grad(e))$ 
   $reactionTerm \leftarrow -dt * R * e$ 
   $F \leftarrow derivativeTerm + diffusionTerm + reactionTerm$ 
   $c \leftarrow FeniCS.solve(F = 0)$ 
  return  $c$ 
end function

```

Algorithm Algorithm 2 Simulation of auxin transport

```

 $c_i \leftarrow 0$ 
 $c_o \leftarrow 0$ 
 $sf \leftarrow$  state function

 $p \leftarrow 0.05$ 
 $k_e \leftarrow 0.5$ 
 $k_i \leftarrow 0.5$ 
 $D \leftarrow 1.0$ 

 $t \leftarrow 0$ 
 $\delta t \leftarrow 0.1$ 
 $t_{final} \leftarrow 300$ 

while  $t < t_{final}$  do
   $R_i \leftarrow sf * p - sf * k_e * c_i + (1 - sf)$ 
   $R_o \leftarrow sf * k_e * c_i - (1 - sf) * k_i * c_o$ 
   $c_i \leftarrow$  UPDATE( $c_0 = c_i, R = R_i, D = D, \delta t = \delta t$ )
   $c_o \leftarrow$  UPDATE( $c_0 = c_o, R = R_o, D = D, \delta t = \delta t$ )
   $t \leftarrow t + \delta t$ 
end while

```

Chapter 3

Growth dynamics of *Marchantia gemmae*

3.1 Introduction

A natural first step towards understanding the regulation of growth is to collect observations of the spatiotemporal patterns of tissue expansion. This chapter presents the work on recording and analysing the growth dynamics of the gemmae during the first two days of their development.

According to the theory of morphoelasticity^[6], tissue growth results in the divergent movement of material points. Therefore, expansion of the tissue can be quantified by tracking notable features on its surface. To the best of my knowledge, two attempts to quantify the spatial patterns of tissue expansion have been previously made in *Marchantia*. The first^[183] relied on imaging of the mature thallus and using regularly-spaced air pores as fixed-place tissue landmarks. The drawback of this method is that it can only be applied to gemmae older than five days, as the air pores do not develop until then. Furthermore, the spatial resolution of the measurement is limited by the number of and distance between air pores. This is especially problematic in the areas surrounding the apical notches, where air pores do not develop fully.

The second study^[130] used an alternative approach, which exploited an image registration method^[184] and as a result provided a much finer spatial resolution. Image registration is a numerical algorithm that fits a transformation that should be applied to a source image for it to match the target image. In this way, every pixel of a source image is a landmark for tracking. In the study, an image registration-based method was applied to microscopy images in order to estimate the expansion rates of individual cells during the first 72 *h* of gemma development. However, the measurements were done with a low temporal resolution (12 – 24 *h* time step) and were not able to capture the fine details of the tissue movement.

Both approaches led to the conclusion that tissue expansion rates are the highest around the apical notches and decay exponentially as a function of the distance from the closest apical notch.

The challenge of collecting the patterns of tissue expansion with higher spatial and temporal resolution lies primarily with the ability to develop an imaging and image-processing assay that would allow continuous data col-

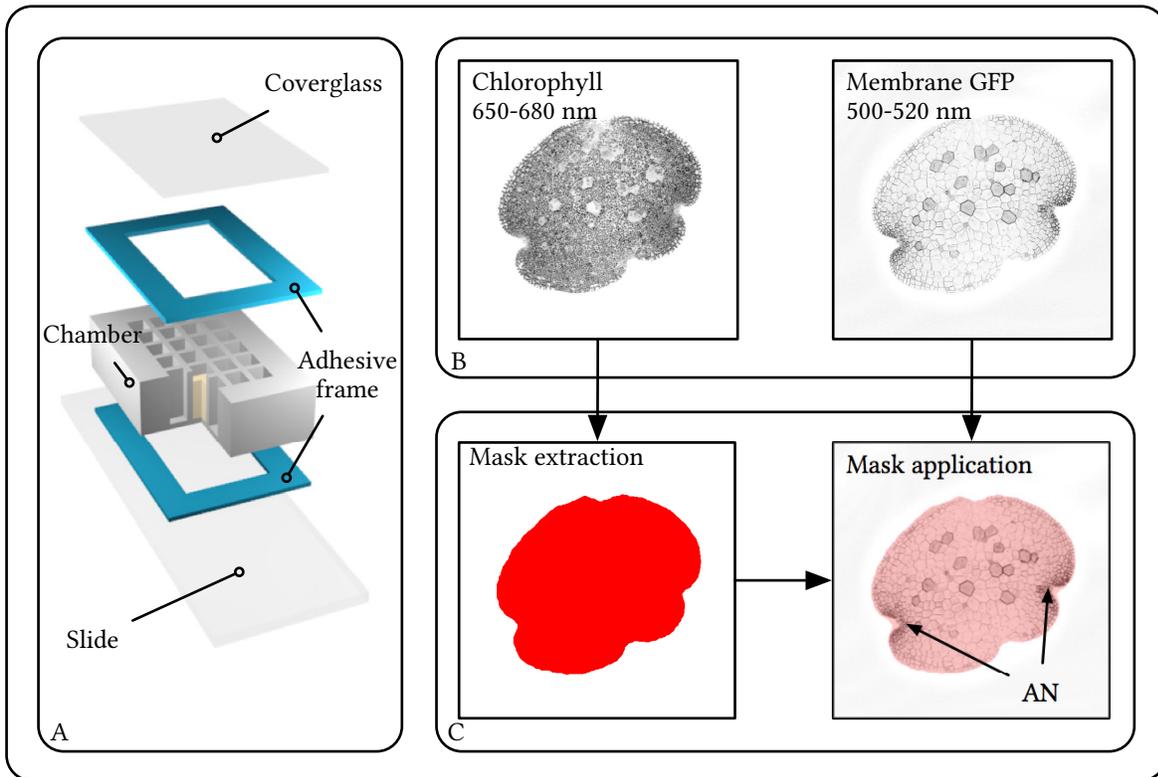


Figure 3.1: **Image acquisition and preprocessing.** **A.** Assembly of the imaging chamber. **B.** Examples of the signal collected using the two channels on the laser-scanning confocal microscope. **C.** The chlorophyll channel was used for separating foreground (*red*) from the background (*white*). AN denotes the apical notches.

lection and reliable analysis of tissue movement. In particular, imaging the live gemmae for several days requires growing them on a solid medium in a contained environment that prevents evaporation of the media and the resultant focus drift. The use of a growth chamber, sealed with a transparent coverslip, accomplishes this but is prone to collecting moisture condensation on the internal side of the coverslip, which obstructs the view of the gemmae beneath it. Furthermore, the coverslip should not touch the gemmae, as doing so would introduce a mechanical constraint to growth.

An additional challenge for analysis is that in order to obtain reliable and statistically-significant results, the data from multiple gemmae should be collected and averaged. However, due to the natural biological variation in the shape and size of the gemmae, direct comparison of the signal collected from different gemmae is not possible.

The work presented in this chapter addresses these challenges and results in a dataset that summarises the data collected from six individual gemmae over 35 time frames spanning the first two days of gemma development.

3.2 Image acquisition

The accuracy of the image registration, and hence the accuracy of the measurement of tissue expansion, relies on the type of landmarks used for tissue tracking. For example, tracking organelles would result in noisy measurements, because the organelles actively relocate inside the cells and their displacements do not represent the actual movement of the tissue. Cell walls, on the contrary, are rigid, do not move within the tissue and have been previously used for measuring expansion rates in *Marchantia*^[130].

To acquire a signal from the cell walls, a transgene encoding a membrane-localised green fluorescent protein (GFP) was introduced into the genome of *Marchantia* via *Agrobacterium*-mediated transformation (Subsection 2.1.4). The localisation to the membrane was achieved by fusing a peptide sequence encoding a *Lti6b* small membrane protein to the C-terminus of the GFP^[156]. The *Lti6b* tag does not facilitate localisation of the fused fluorescent protein to the cell wall itself, but instead directs the protein to the cell membrane. The thickness of the plant cell wall is $\sim 1 \mu\text{m}$, similar to the resolution of the collected images, and under normal osmotic conditions the membrane is tightly pushed against it. Therefore, the localisation of the *GFP-Lti6b* signal is indistinguishable from the position of the cell walls. The gene was driven by the *proMpEF1a* promoter^[151], which has a pattern of constitutively high expression in the young gemma tissue (Section 2.2).

Twelve transformants with bright green fluorescence in the membrane were selected and grown until they developed gemma cups (Subsection 2.3.2, subsection 2.1.2). Young gemmae were picked from the gemma cups of mature plants and examined under a stereo microscope to verify the presence of the fluorescent signal and absence of tissue damage. Gemmae that passed the examination were placed into a custom-made imaging chamber (Figure 3.1 A, subsection 2.3.1). The imaging chamber had a cuboid shape with 24 wells in a 6×4 configuration. The wells were filled with $\sim 90 \mu\text{l}$ hot media agar and left to cool for $\sim 30 \text{min}$. Once the media solidified, the gemmae were placed into the chamber, one per well. If left uncovered, the water from the agar would evaporate, causing the drying of the media and resulting in focus drift during microscopy. To prevent evaporation, the chamber was sealed using a double-sided adhesive frame and a coverslip treated with an anti-fogging solution. A 0.5mm gap was left between the gemmae and the coverslip to avoid mechanical constraints on gemmae growth.

The chamber was fixed onto a $75 \times 26 \text{mm}$ microscope slide using a double-sided adhesive frame and mounted onto the stage of a laser-scanning confocal microscope. The gemmae were imaged every 72min for 35 cycles using 470nm excitation laser. Signals from two bands of the emission spectrum were collected. The $500 - 520 \text{nm}$ band, contained the signal from the membrane-localised GFP, while the $650 - 680 \text{nm}$ band contained the chlorophyll autofluorescence signal (Figure 3.1 B, subsection 2.3.3). For each gemma at each time point, a $1024 \times 1024 \times 60$ 3D image with a voxel size of $1.5 \times 1.5 \times 5 \mu\text{m}$ was collected.

3.3 Image processing

3.3.1 Preprocessing

Expansion of gemmae tissue during the first two days of development mainly occurs in the horizontal plane^[130]. Therefore, the three-dimensional images were projected along the vertical axis using a maximum intensity function, and the subsequent analysis of growth was done in two dimensions.

The image background did not contain valuable information and could create artefacts during the image processing. Therefore, it was necessary to create a binary mask, that would separate the foreground from the background. For this purpose, the channel containing chlorophyll autofluorescence was used. To convert the signal to a binary mask, a Gaussian filter, followed by the background detection algorithm was applied (Figure 3.1 C, subsection 2.4.2: IP.1).

3.3.2 Quantifying tissue expansion

The collected time-lapse images contained information about the evolution of gemmae shapes during the first two days of development. In order to calculate the relative tissue expansion rates, the time-lapse images were processed in a two-step pipeline.

During the first step, the velocity field was estimated using the image registration algorithm (Subsection 2.4.1: IP.2). Two relative deformation functions, forward T_t^f and reverse T_t^r , were fitted such that they warped the signal of the membrane marker at time frame t for it to match the corresponding signal at the previous and the following time-frames at time $t - 1$ and $t + 1$ (Figure 3.2 A). These relative deformation functions contained information about the displacements, u_f and u_r , that should be applied to each pixel in order to minimise the difference between the signal of the current and either a following or a previous time frame (Figure 3.2 B). The velocity field was then approximated as an average of these displacements, normalised by the time step between the two time frames (Model 3.1 A).

In the second step, the divergence of the velocity field was calculated as an approximation for the relative tissue expansion rate (Model 3.1 B, subsection 2.4.1: IP.4).

3.4 Normalisation and averaging

As a result of the image processing, spatiotemporal patterns of relative tissue expansion rates were calculated for the six gemmae over $\sim 50 h$.

In agreement with the previous research^[130,183], the rates of tissue expansion were higher around the apical notches compared to that in the central part of a gemma(Figure 3.3 A).

Since the division and expansion rates of a gemma are clearly dependent on the distance from the notch^[130,183], it has been a common approach to describe spatial aspects of gemma development by placing the origin of the coordinate system at the apical notches. From a practical perspective, this requires labelling notch positions at every time frame (Subsection IP.5). In order to do so, the positions of the notches were manually assigned at the last

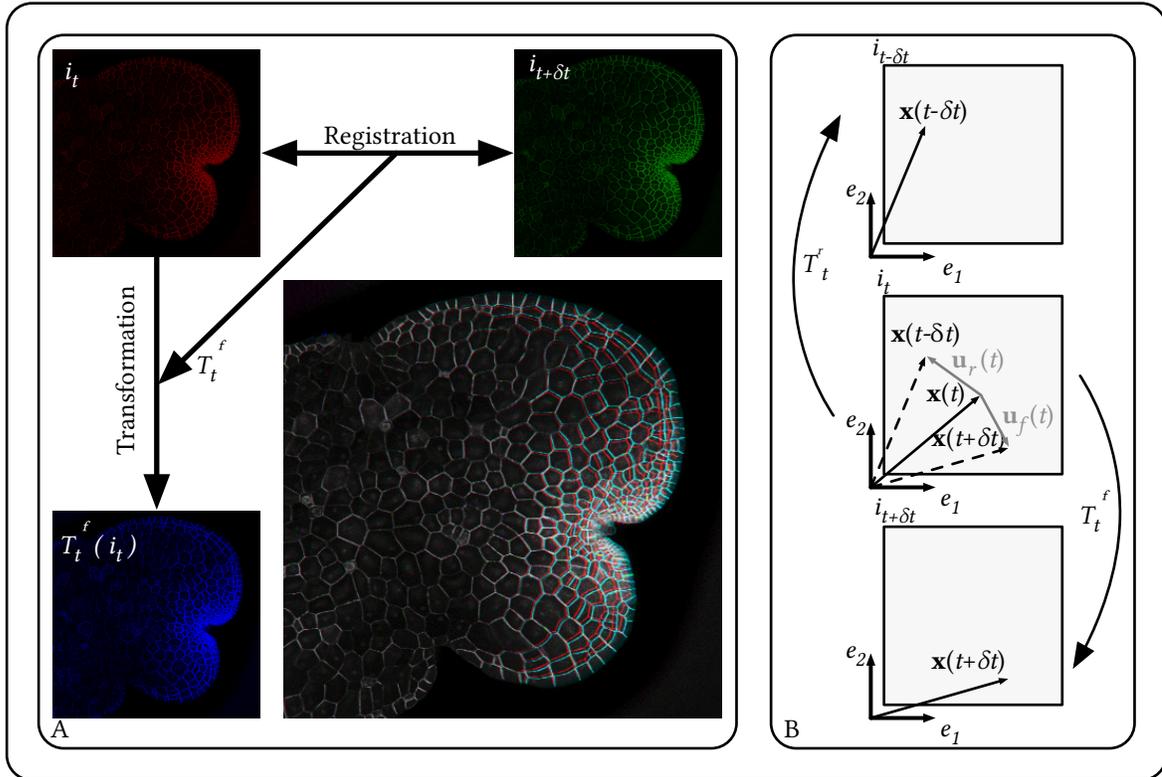


Figure 3.2: **Quantifying tissue expansion.** **A.** An example of image registration between two consecutive time frames. A time frame at time t is shown in *red*, the following time frame at time $t + \delta t$ is shown in *green*. A forward relative transformation function, T_t^f , is found by using the image registration algorithm between the two time frames. The transformation is then applied to the time frame at time t to obtain a transformed image, shown in *blue*. Colocalisation of the green and the blue signal in the overlaid image demonstrates the accuracy of the estimated transformation function. **B.** The details of the velocity estimation algorithm on an example of a single material point with coordinate $\mathbf{x}(t)$. Image registration algorithm is applied between the current (i_t) and both the previous ($i_{t-\delta t}$) and the following ($i_{t+\delta t}$) time frames. As a result, two relative deformation functions are found, forward T_t^f and reverse T_t^r . The deformation functions contain information about the displacements u_f and u_r that the material point has moved relative to the previous and the following time frames.

Model 3.1 Quantifying relative tissue expansion rates

A) Estimating velocity field

An approximation of the velocity field can be found by applying an image registration algorithm^[184] between consecutive pairs of the time frames. The image registration algorithm fits the relative deformation functions $T_t^f(\mathbf{x})$ and $T_t^r(\mathbf{x})$, which contain information about the displacements $\mathbf{u}_f(\mathbf{x}, t)$ and $\mathbf{u}_r(\mathbf{x}, t)$ that should be applied to each image pixel \mathbf{x} in order to minimise the difference in signal between the current and either following, or previous, time frames:

$$\begin{aligned} T_t^f(\mathbf{x}) &= \mathbf{x} + \mathbf{u}_f(\mathbf{x}, t) \\ T_t^r(\mathbf{x}) &= \mathbf{x} + \mathbf{u}_r(\mathbf{x}, t) \end{aligned}$$

The velocity field can be then estimated as

$$\mathbf{v}(\mathbf{x}, t) \approx \frac{\mathbf{u}_f(\mathbf{x}, t) - \mathbf{u}_r(\mathbf{x}, t)}{2\delta t}$$

B) Calculating relative tissue expansion rate

According to Lagrangian description of motion^[9], deformation can be represented by a *function* $\chi : B_0 \rightarrow B_t$ that maps coordinates of the tissue points at an initial configuration B_0 to coordinates of the same points at some later configuration B_t at time t , such that

$$\mathbf{x} = \chi(\mathbf{X}, t), \forall \mathbf{X} \in B_0$$

where \mathbf{X} is a two-dimensional position vector of a material point in the initial configuration B_0 , and \mathbf{x} is the position vector of the point at some later configuration B_t . In the context of the acquired images, the initial configuration B_0 corresponds to the state of the gemma tissue recorded at the first time frame, while B_t corresponds to configurations of the gemma tissue states at all subsequent time frames.

The information about the tissue strain is captured in the gradient of the deformation function:

$$\mathbf{F} = \frac{\partial \chi}{\partial \mathbf{X}}$$

Arial expansion of the tissue, J , defined as a ratio between the area of a tissue element at time t (A_t) to its original area (A_0), can be further expressed as a determinant of the deformation function:

$$J = \frac{A_t}{A_0} = \det(\mathbf{F})$$

Furthermore, from the law of mass balance, it can be shown that relative volumetric expansion rate, $\gamma = J^{-1} \frac{\partial J}{\partial t}$, is equal to the divergence of the velocity field \mathbf{v} :

$$\gamma = J^{-1} \frac{\partial J}{\partial t} = \nabla \cdot \mathbf{v} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y}$$

where

$$\mathbf{v} = \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \frac{\partial \mathbf{x}}{\partial t}$$

can be estimated as shown in **A**.

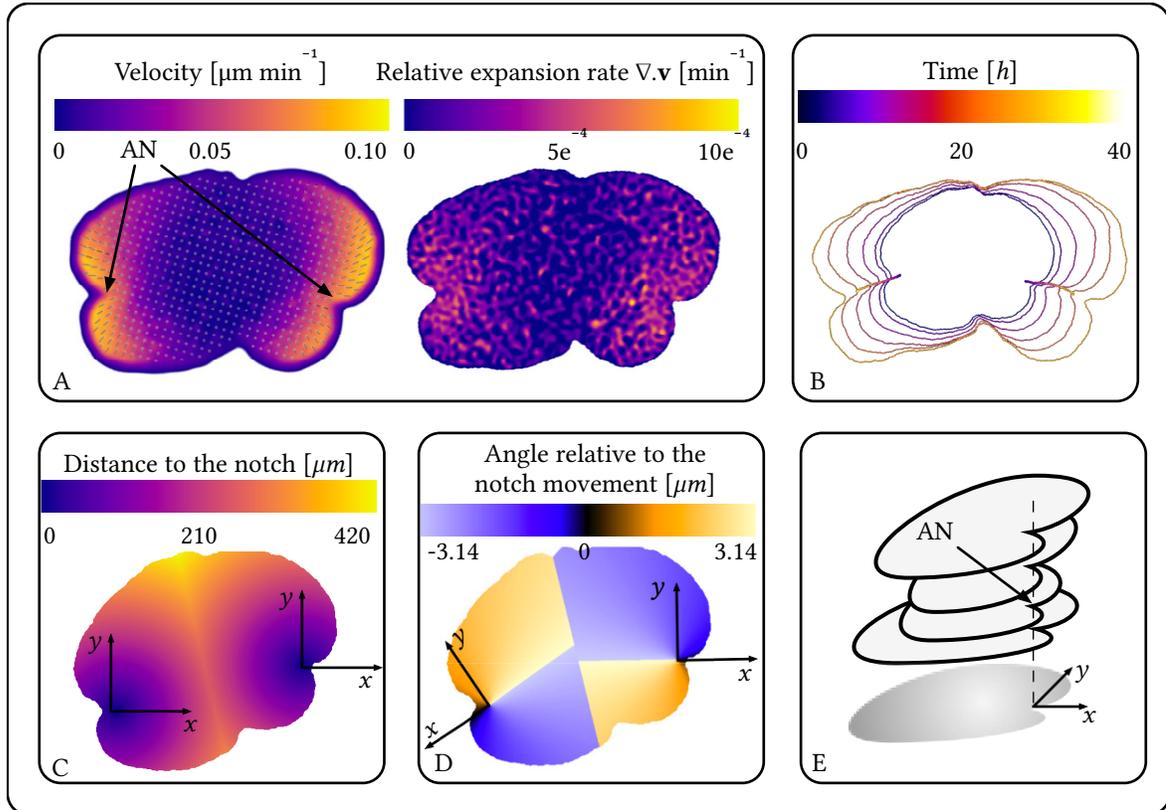


Figure 3.3: **Alignment, normalisation and averaging of relative expansion rate patterns.** **A.** Visualisation of the estimated velocity field and relative tissue expansion rates for one time frame of a single gemma. Orientation of the lines matches the direction of the velocity, while their lengths are proportional to the magnitude of the velocity. **B.** Result of the notch tracking algorithm in Model 3.2. The closed contours denote outline of the growing gemmae, while the circles denote location of the apical notches. **C.** The alignment of the origin of the coordinate system with the apical notches. The points on the gemma are split into two domains based on the distance to the closest notch. **D.** The alignment of the x-axis of the coordinate system with the direction of the notch movement. **E.** All material points from all time frames are aligned with the same coordinate system, where the origin denotes the position of the closest notch (AN), and the x-axis is aligned with the direction of the notch movement.

Model 3.2 Tracking the notches

Let $T_t^f(\mathbf{x})$ and $T_t^r(\mathbf{x})$ be forward and reverse relative deformation functions that map coordinates of the material points from frame at time t to the following and reverse time frames, respectively:

$$\begin{aligned} T_t^f(\mathbf{x}(\mathbf{X}, t)) &= \mathbf{x}(\mathbf{X}, t + \delta t) \\ T_t^r(\mathbf{x}(\mathbf{X}, t)) &= \mathbf{x}(\mathbf{X}, t - \delta t) \end{aligned}$$

Hence the coordinate of any material point $\mathbf{x}(\mathbf{X}, t)$, given \mathbf{X} , can be approximated as a composition of forward relative deformation functions

$$\mathbf{x} = \chi(\mathbf{X}, n\delta t) = T_{n\delta t}^f \circ T_{(n-1)\delta t}^f \circ \dots \circ T_{\delta t}^f(\mathbf{X}), \quad \text{for } n \in [1, N]$$

where δt is a time step, n is an index of a time frame and N is a total number of time frames. Therefore, in order to identify position of the notch at every time frame it is only necessary to manually specify its coordinates at the first time frame.

However, since the exact position of the notch is ambiguous at the first time frame, more accurate tracking is achieved, when the location of the notch is specified at the last time frame and then is tracked backwards using compositions of the reverse relative deformation functions:

$$\mathbf{x}(\mathbf{X}, (N - n)\delta t) = T_{N-n+1}^r \circ T_{N-n+2}^r \circ \dots \circ T_N^r(\mathbf{x}(\mathbf{X}, N\delta t)) \quad \text{for } n \in [1, N]$$

time frame of each time-lapse and were tracked backwards by using the pre-calculated velocity fields (Model 3.2, figure 3.3 B).

Next, the coordinate transformation was performed for every time frame by subtracting the coordinate of the closest notch from the coordinate of every collected data point of the corresponding gemma at the same time frame (Figure 3.3 C). In addition, the coordinate frame was rotated so that the x-axis was aligned with the direction of the notch movement, as prescribed by the calculated velocity field (Figure 3.3 D).

As a result, the frames of reference for every time frame were centred at the notches, and the x-axis pointed in the direction of the notch movement. Since each gemma has two apical notches, each time frame of each gemma image produced two sets of data split based on proximity to the closest of the notches. In other words, the spatial data from 12 apical regions from six different gemmae (two regions per gemma) were aligned to the single, notch-centred coordinate system (Figure 3.3 E, subsection 2.4.1: IP.5). By doing so, differential expansion of the tissue relative to the notch can be studied on a larger sample.

The natural variation of gemmae sizes, however, posed a question whether the tissue expansion rates recorded at the same coordinates, but originating from different gemmae, could be directly compared. Furthermore, due to growth, the number of data points recorded was inversely proportional to the distance of the data points from the closest notch.

In order to account for the biological variation in gemmae sizes and to distribute the recorded data more evenly across the range, the coordinates were normalised by the distance between the apical notches at the corresponding time frame.

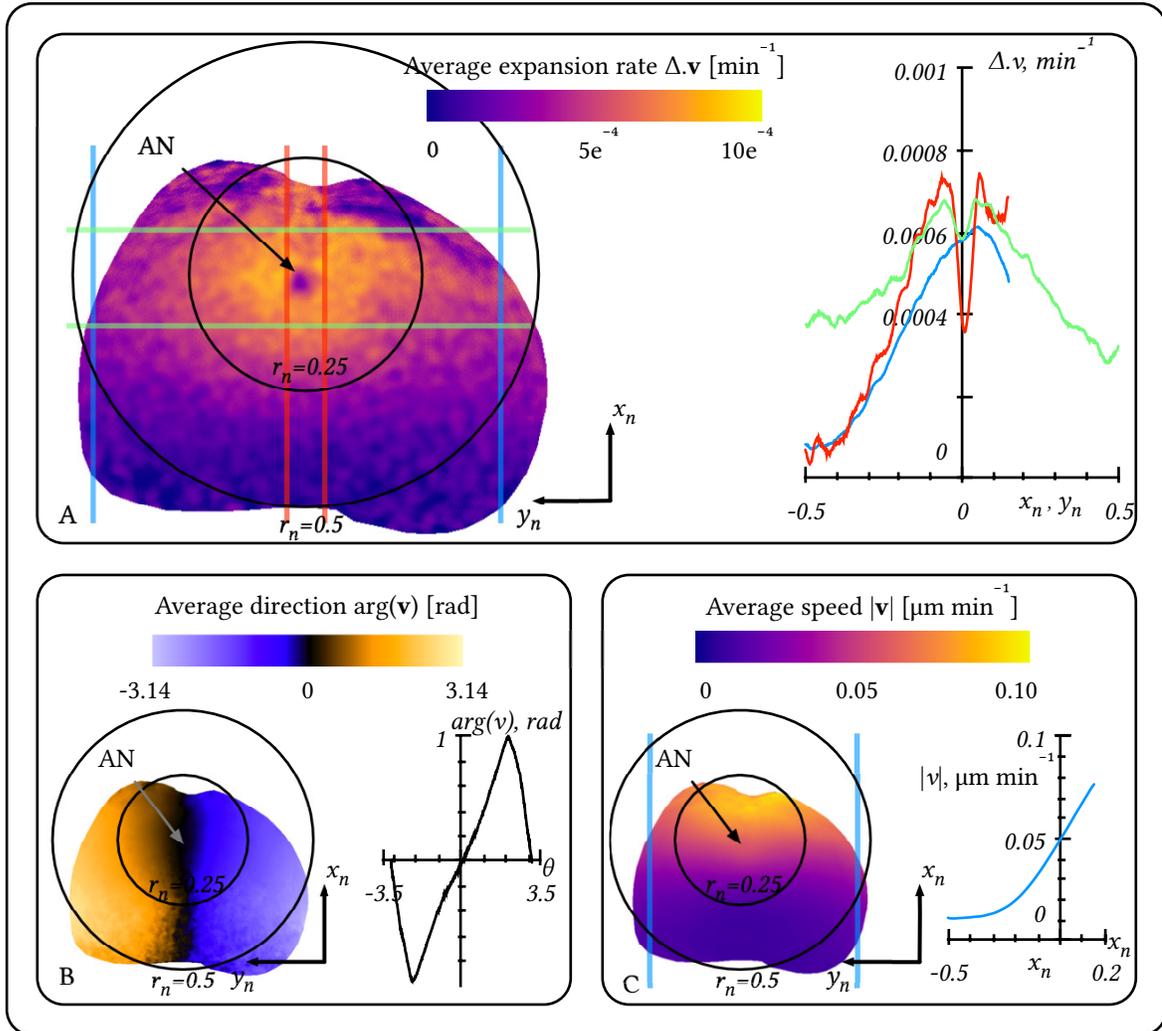


Figure 3.4: **Average velocity and tissue expansion rates.** A, B and C. The summary of the relative tissue expansion rates, direction of tissue movement and the speed of tissue movement averaged across a sample of 24 notches. The heatmaps show the spatial pattern of the measured parameters, and the graphs show the signal of the 1D slices through the heatmaps. The distance (r_n), and coordinates (x_n and y_n) relative to the apical notch were normalised by the distance between the two apical notches of a gemma at the corresponding time frame. The colours of the traces on the graphs correspond to the colours of the parallel lines on the heatmaps. The colour lines on the heatmaps are in pairs and are parallel to the direction of the slices. The distance between the lines in a pair denotes the thickness of the slice. The traces on the graph show the values of the parameters averaged perpendicular to the slice orientations, e.g., the blue trace in A corresponds to the variation of the tissue expansion rate along the x-axis. AN denotes apical notch.

Model 3.3 Normalisation and averaging of the tissue expansion rate patterns

A) Relative tissue expansion rate

Prior to normalisation and averaging, the relative tissue expansion rate (g) was recorded as a function (Γ) of coordinates in the image reference frame of gemma i at time t ($\mathbf{x}^{i,t}$):

$$g = \Gamma(\mathbf{x}^{i,t}), \mathbf{x}^{i,t} = (x^{i,t}, y^{i,t})$$

B) Alignment relative to the closest apical notch

In order to normalise the reference frames across all gemmae and all time frames, the positions of the closest apical notches ($\mathbf{x}_0^{i,t} = (x_0^{i,t}, y_0^{i,t})$) were subtracted from the original coordinates ($\mathbf{x}^{i,t}$). Furthermore, the reference frames were rotated such that the x-axis of the new coordinate frame matched the direction of the velocity of the apical notches. In result, the transformed coordinates ($\mathbf{x}_T^{i,t}$) can be represented as:

$$\mathbf{x}_T^{i,t} = (r_T^{i,t} \cos(\theta_T^{i,t}), r_T^{i,t} \sin(\theta_T^{i,t}))$$

where

$$r_T^{i,t} = \sqrt{(x^{i,t} - x_0^{i,t})^2 + (y^{i,t} - y_0^{i,t})^2}$$

$$\begin{aligned} \theta_T^{i,t} = \text{mod}(& \\ & \text{atan2}((y^{i,t} - y_0^{i,t}), (x^{i,t} - x_0^{i,t})) \\ & - \text{arg}(\mathbf{v}^{i,t}(x_0, y_0)) + \pi \\ & , 2\pi) - \pi \end{aligned}$$

C) Normalisation in relation to gemma size

To account for variation in gemmae size, the transformed coordinates were further normalised by the distance between the apical notches at the corresponding timeframe ($d_{i,t}$):

$$\begin{aligned} \mathbf{x}_n^{i,t} &= (x_n^{i,t}, y_n^{i,t}) \\ &= (r_n^{i,t} \cos(\theta_n), r_n^{i,t} \sin(\theta_n)) = \frac{\mathbf{x}_T^{i,t}}{d_{i,t}} \end{aligned}$$

The relative tissue expansion rate could be now expressed as a function of the normalised coordinates:

$$g = \Gamma_n(\mathbf{x}_n^{i,t})$$

D) Averaging

The average pattern of tissue expansion rate across all time frames for each gemma (\bar{g}^i) can be then found as

$$\bar{g}^i = \frac{\sum_t \Gamma_n(\mathbf{x}_n^{i,t})}{n_i}$$

where n_i is a total number of data-points for gemma i . Similarly, the average pattern of tissue expansion rate for each time frame (\bar{g}^t) can be found as

$$\bar{g}^t = \frac{\sum_i \Gamma_n(\mathbf{x}_n^{i,t})}{n_t}$$

where n_t is a total number of data-points for time frame t .

Then the average tissue expansion rate across all gemmae and time frames was calculated as:

$$\bar{g} = \Gamma_a(\mathbf{x}_n) = \frac{\sum_{i,t} \Gamma_n(\mathbf{x}_n^{i,t})}{n_{i,t}}$$

where $n_{i,t}$ is the number of data points collected for gemma i at time frame t .

The data from 12 individual notches were grouped by the normalised coordinates into 1000×1000 grid with limits $(-1, 1)$ for both axes. The mean value for each group was calculated to obtain an average pattern of relative tissue expansion rate (Figure 3.4). To assess the variability, the data was grouped by either gemma sample or time frame followed by spatial grouping into 1000×1000 as before. The average patterns for each gemma sample and each time frame are shown in figure 3.5 and figure 3.6.

3.5 Discussion

In this chapter, a method for quantifying tissue expansion rates was introduced. The method was applied to six *Marchantia* gemmae to record an average pattern of tissue expansion rates relative to the apical notch. For this purpose, a novel sample preparation and imaging assay was developed. The gemmae were grown under a laser-scanning confocal microscope inside a 3D-printed growth chamber, sealed with a cover glass treated with a solution to prevent the accumulation of condensate. This technique allowed for the collection of time-lapse images over the period of two days with a temporal resolution of 1 *h*.

The collected data were processed using an image-registration-based algorithm to extract the velocity fields of the imaged gemmae tissue. The divergence of the velocity fields was then calculated to obtain patterns of relative tissue expansion rates. Finally, the data from the six gemmae was normalised for the biological variation of the gemmae shapes and sizes, and averaged to produce a mean pattern of relative tissue expansion.

In agreement with the previous reports, the tissue expansion rates were highest around the apical notches and decayed gradually as a function of distance from the notch. The previous studies on *Marchantia* gemma tissue kinetics modelled the pattern of relative tissue expansion rates as being radially symmetric around the notch. However, the results of this work indicated that both the magnitude of the tissue movement and its relative expansion rates decayed quicker along the proximal-distal axis of a gemma (*x*-axis in Figure 3.4).

In summary, the tissue expands around the apical notches and remains relatively static in the central region of a gemma. For cells to exhibit such differential growth in different regions of the gemma, one could hypothesise that there exists a positional cue that informs the cells to either expand or be restrained from expanding. Hypothesis about the nature of this positional signal and the mechanism behind its control is introduced in chapter 4.

The collected data indicated a region of low growth right around the location of the apical notch ($r_n < 0.1$). To the best of my knowledge, this observation has not been previously reported, however, it is not clear whether this phenomenon has a biological origin. The numerical methods that were the base of calculation of relative tissue expansion rates relied heavily on the assumption that the dorsal surface of the gemmae can be considered flat, i.e. that it has a small variation in horizontal direction compared to the radius of a gemma. This assumption holds well in general^[130]: tissue in close proximity to the apical notch forms a sharp horizontal dip, which is a characteristic trait of the notch itself. Due to this fact, the low recorded values of the relative tissue expansion rates in the close proximity to the apical notch might be numerical artefacts.

The variability analysis of the data uncovered a pulsatile nature of gemma growth during the imaging experiment (Figure 3.6). The pulsatile growth was present in all six gemma samples with a period ~ 20 *h*. Whether this

pulsatile growth is a normal developmental trait of *Marchantia gemmae* or is an artefact resulting from experimental conditions, e.g. temperature variation in the microscopy room, was not determined. Influence of circadian-regulated mechanisms on this kinetic pulsation is unlikely as both the gemmae and the mother plants from which they were extracted had been grown under continuous light conditions both before and during microscopy.

A major limitation of the presented method is that it only works reliably as long as the object of observation can be considered flat, which does not hold for gemmae older than three days after their removal from the gemma cup. Even for the earlier stages, the reduction of the depth axis (*z*-axis) and considering the problem in only two dimensions results in up to 20 % errors in the estimations of relative expansion rates^[130]. This is because in reality, although being relatively flat, the gemmae are not constrained to, and in general do not, expand in two dimensions only. To address these issues, three-dimensional image registration can be employed. However, even in that case the signal obtained by fluorescence microscopy will only cover the epidermal surface of the gemma, and the tissue dynamics at the deeper layers will be inaccessible for observation. For precise three-dimensional registration, alternative imaging methods can be explored, e.g. micro-CT or optical coherence tomography.

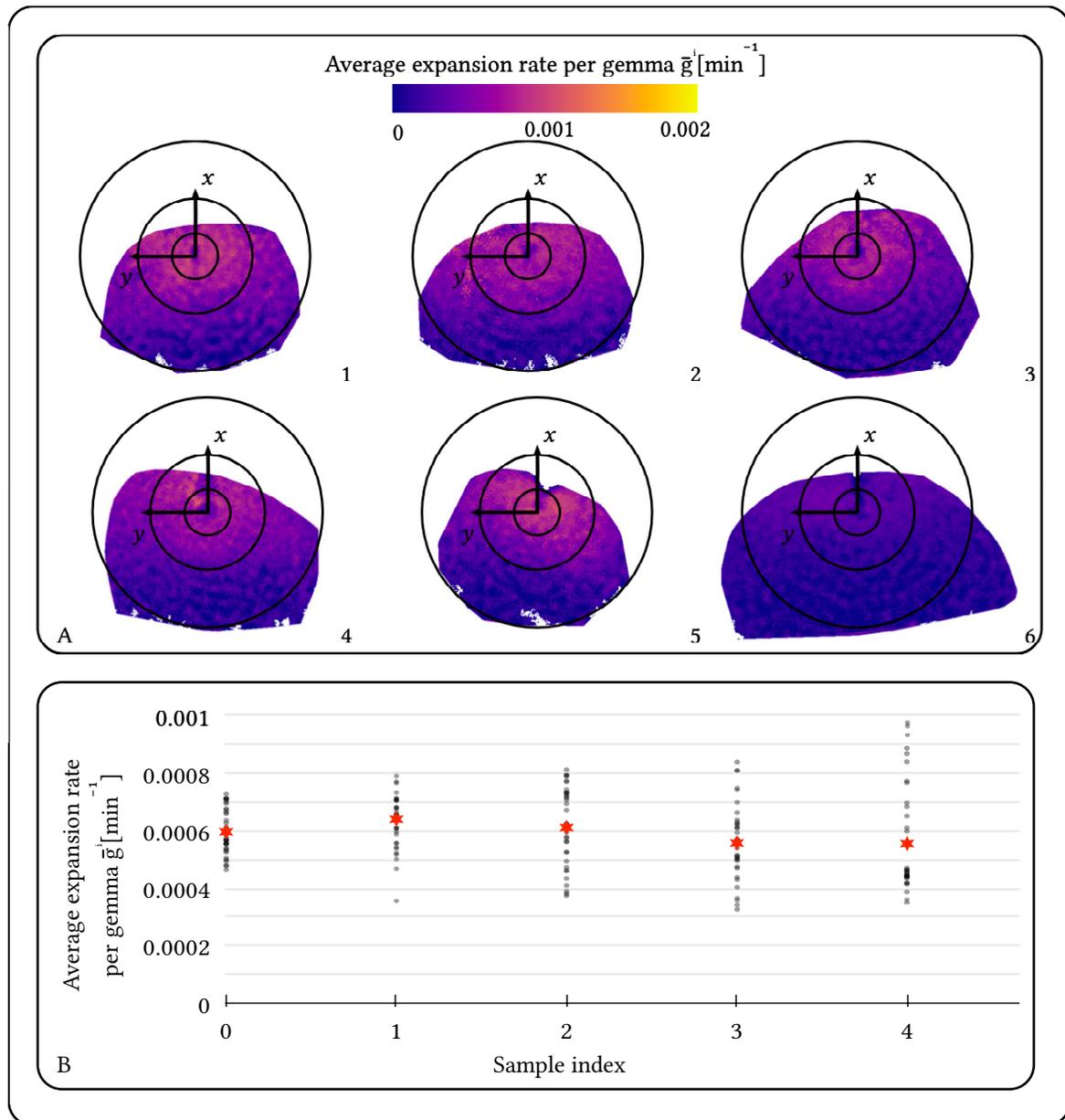


Figure 3.5: **Sample variability of tissue expansion rate.** **A.** Normalised patterns of relative tissue expansion averaged over all 35 time frames for each of the six gamma samples. The circles represent $r_n = 0.1, r_n = 0.25$ and $r_n = 0.5$. Each diagram represents cumulative data from both of the gamma notches. The numbers in the bottom-right corner represent gamma index. **B.** Average expansion rate for each gamma sample. Grey circles represent average per time frame, red stars denote the average across all time frames.

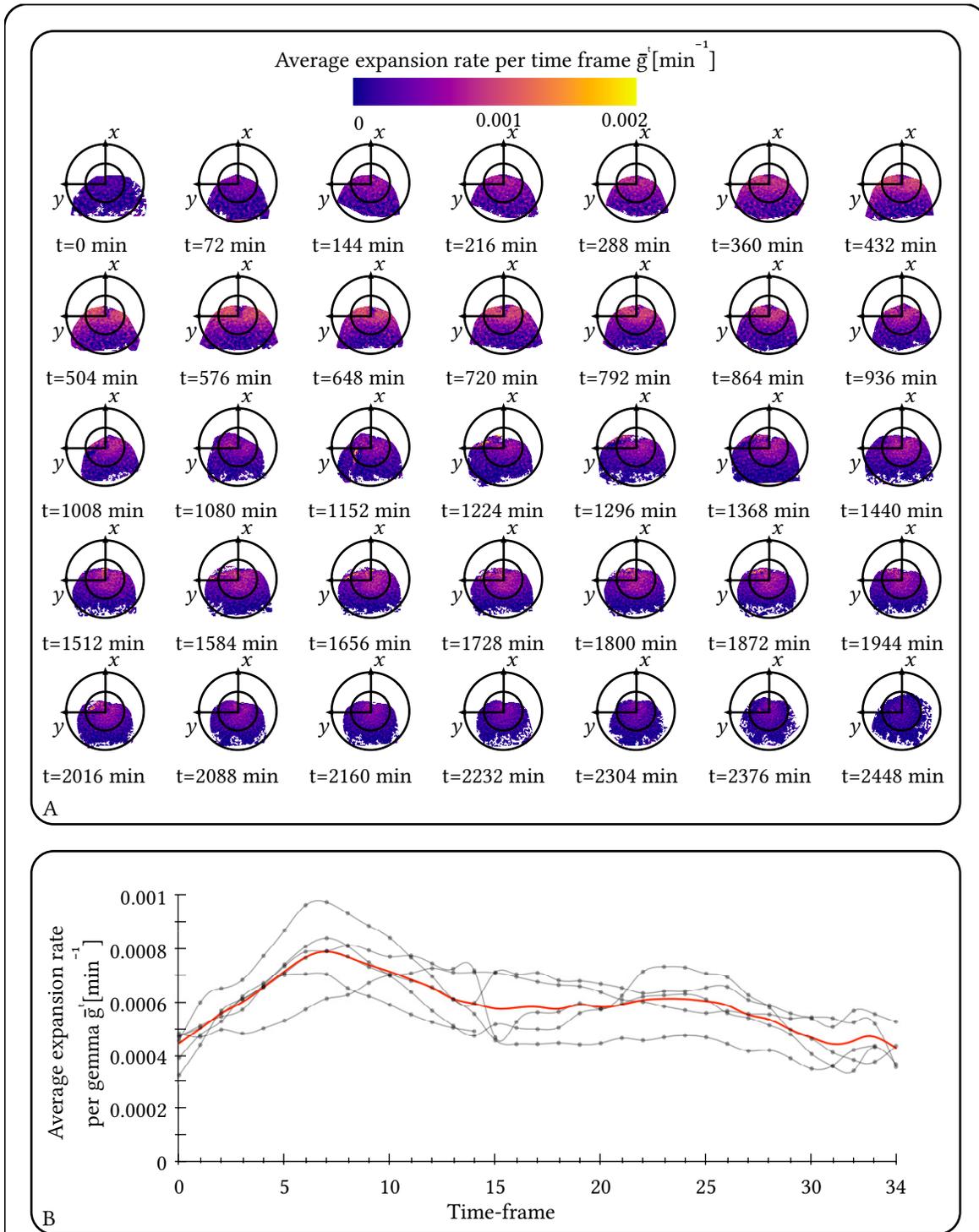


Figure 3.6: **Time variability of tissue expansion rate.** **A.** Normalised patterns of relative tissue expansion averaged over all gemma samples for each time frame. The circles represent $r_n = 0.25$ and $r_n = 0.5$. Each diagram represents cumulative data from both of the gemma notches. **B.** Average expansion for each time frame. Grey traces represent signals for each of the individual gemma samples, red trace is the average across all samples.

Chapter 4

Maintenance and reestablishment of the apical regions in *Marchantia*

4.1 Introduction

The apical notches of *Marchantia* are the centres of cell division, similar to the central zone of a shoot apical meristem in higher plants^[185]. Most commonly, a *Marchantia* gemma has two apical notches that branch out during the course of development.

As has been described in Chapter 3, the growth of a gemma during the first 48 *h* is polarised along the axis connecting the two apical notches; and the tissue can be conceptually separated into two types, with characteristic growth properties in relation to the distance to the closest apical notch. The cells surrounding the apical notches form *apical regions*, where they expand and divide. The remaining cells form a *central region*, where expansion and division are restrained. In other words, the growth of a gemma is driven by the apical regions. Cells divide in proximity to the apical notches, elongate in the apical regions, and as they leave the apical regions and enter the central region, cell division and expansion ceases. A key question for understanding and controlling the development of a gemma is how this *growth pattern* is established and maintained, i.e., what is the developmental mechanism behind the differential cell expansion and division between the apical and the central region.

In previous work^[183], Solly et al. took a top-down approach and hypothesised a mechanism for the maintenance of the apical regions. The proposed model explained the tissue expansion pattern during gemmae growth and its correlation with the distance to the apical notch by hypothesising the existence of a growth-inducing diffusible morphogen, termed *APEPROX*, produced in the apical regions. According to the model, the expansion of the tissue in the apical regions is driven by the high concentration of the *APEPROX*, while the lack of tissue expansion in the central region is explained by the degradation of the *APEPROX* as it diffuses away from its sources in the apical regions. Although the simulations of this transport mechanism were able to recapitulate the development of a gemma shape,

no direct evidence of the existence of the *APEPROX*-like morphogen in *Marchantia* was demonstrated. Furthermore, it was not clear whether any alternative transport mechanisms could explain the gemma growth pattern.

The authors proposed the phytohormone auxin as a candidate for the growth-inducing *APEPROX*. However, earlier experiments suggested that auxin inhibits the growth of *Marchantia* when applied externally^[186]. Furthermore, the depletion of auxin has been demonstrated to induce uncontrolled proliferation of tissue, similar to that in the early stage of sporelings development^[133]. Therefore, according to the experimental evidence, auxin, in contradiction to the model prediction, plays the role of the growth inhibitor.

In summary, there is a clear relationship between the cell proliferation and distance to the apical notch, suggesting a presence of a positional cue that controls cell differentiation. Therefore the mechanism, in which the positional cue is created by a diffusible morphogen deserves a further investigation. Considering that neither effect on the cell proliferation nor localisation of the morphogen production or accumulation are known, a systematic approach considering several possible scenarios was taken in the work presented in this chapter.

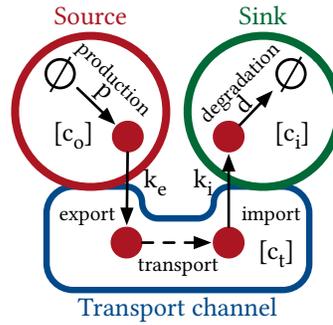
4.2 Analysis of the transport model

In order to explore alternative mechanisms that might explain a phenotypical difference between *Marchantia* cells inside the apical and the central regions, a mathematical source–sink model was developed (Model 4.1). The model described the transport of a morphogen between the two tissue types, *source* and *sink*, that were connected by a *transport channel*. The source and the sink represented the intracellular space of the apical or the central regions, while the transport channel represented the intercellular space connecting the two. The source produced and exported the morphogen into the transport channel, while the sink imported the morphogen from the transport channel for its further degradation.

According to the steady-state analysis, accumulation of the morphogen can be biased towards either the sink or the source depending on the relative values of the model parameters. The concentration of the morphogen is higher in the source compared to that in the sink when the degradation rate is higher than the export rate. This regime was captured in the model proposed by Solly et al.^[183], where the apically-produced morphogen accumulated in the apical regions. However, when the degradation rate is lower than the export rate, the concentration of the morphogen in the sink is higher than that in the source. An important conclusion from this analysis is that the production of a morphogen in a particular region does not imply its function. For example, production of the morphogen in the apical region does not imply its role as an inducer of cell proliferation, as the final pattern of the morphogen accumulation would be determined by the relative values of the transport parameters.

Taking into account that either of the regions can act as the source or the sink of a hypothetical morphogen, as well as the fact that the morphogen can, in theory, either induce or inhibit cell proliferation, there are four possible scenarios regarding how the growth pattern in the intact gemma can be maintained. In other words, there are four possible combinations of the transport parameters and type of morphogen action that could restricts cell proliferation to the apical regions (See Figure 4.1 A). The model of Solly et al. is represented by a mechanism *Nr 1*, i.e., an apically-produced inducer of the cell proliferation that accumulates at the source.

Model 4.1 Source–sink transport

**Description**

The model defines two separate compartments, a source and a sink, connected by a transport channel. The source produces a morphogen with the rate p and exports it to the transport channel following the first order kinetics with an export parameter k_e . The sink imports the morphogen from the transport channel following the first order kinetics with an import parameter k_i , where it is degraded or consumed with a degradation coefficient d . The following system of differential equations describes the dynamics of the morphogen concentrations in the source (c_o), the sink (c_i) and the transport channel (c_t):

$$\begin{cases} \frac{dc_o}{dt} = p - k_e c_o \\ \frac{dc_t}{dt} = k_e c_o - k_i c_t \\ \frac{dc_i}{dt} = k_i c_t - d c_i \end{cases}$$

Steady-state

The system above is in steady state when

$$\begin{cases} c_o = p/k_e \\ c_t = p/k_i \\ c_i = p/d \end{cases}$$

The ratio of the morphogen concentration in the source to that in the sink is then equal to

$$c_o/c_i = d/k_e$$

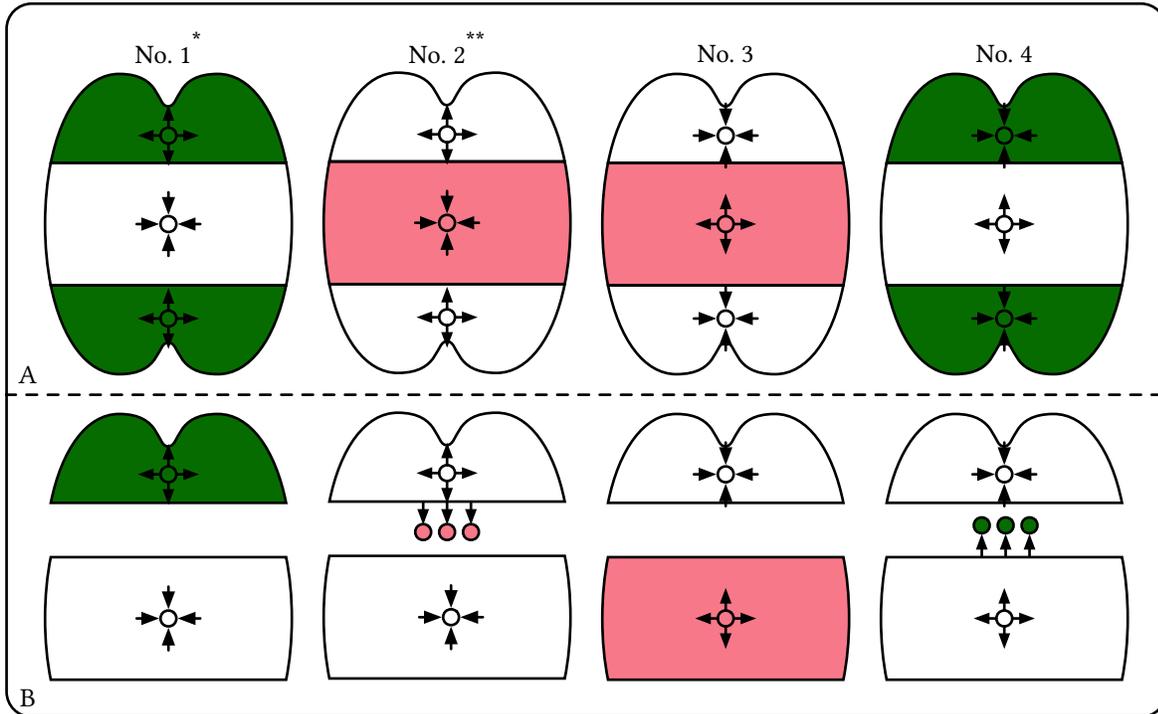


Figure 4.1: **Possible mechanisms explaining the maintenance of the apical regions.** **A.** Steady state of the proposed mechanisms in the intact gemmae. *No.1*: An inducer of the cell proliferation is produced and accumulated in the apical region. *No.2*: An inhibitor of cell proliferation is produced in the apical regions while being transported to and accumulated in the central region. *No.3*: An inhibitor of the cell proliferation is produced and accumulated in the central region. *No.4*: An inducer of cell proliferation is produced in the central region while being transported and accumulated in the apical regions. **B.** Steady state of the proposed mechanisms in the isolated apical and central regions. Red colour denotes regions of the inhibitor accumulation, the green colour denotes regions of the of the inducer accumulation. The divergent arrows denote source regions, convergent arrows denote sink regions.

* The mechanism proposed by Solly et al.^[183]

** The mechanism supported by the results of the surgical assay

4.3 Growth of the isolated apical and central regions

Since the proposed transport model is abstract and does not represent spatial representation of any particular plant morphogen, it is impossible to reject any of the possible mechanisms without acquiring further biological evidence. In an attempt to falsify some of the proposed mechanisms, the following assay was employed. Young gemmae were removed from the gemma cups of a mature thallus and were surgically separated into three parts with two cuts perpendicular to the axis connecting the two apical notches (Figure 4.2 A, subsection 2.1.2, subsection 2.1.5). By doing so, it was possible to examine the development of the apical regions and the central region in isolation from each other.

Under the assumptions of the model by Solly et al., the cells of both regions should have maintained their fates. Production of the morphogen in the apical regions should have continued to promote proliferation of cells, while the lack of morphogen influx into the isolated central region should have maintained the cells in a non-proliferative state. However, the results of the surgical assay contradicted the predictions of Solly et al.'s hypothesis.

In agreement with the previous experiments, in which the apical regions were excised from the gemmae^[187] or mature thallus^[188], the cells of the isolated central regions underwent several rounds of division (Figure 4.3), established new apical regions and regenerated the tissue architecture by the fifth day following the excision (Figure 4.2 B).

The regeneration of the central region can be divided into three stages. During the first stage, the majority of the cells on the dorsal surface undergo division, phenocopying cells of the apical regions. At the same time, the rhizoids on the dorsal surface elongate substantially. During the second stage, cell divisions continue to happen but are localised to one or several *apical precursor regions*; the cells around the apical precursor regions elongate, new borders between the apical and the central regions are established. During the third stage, the tissue returns to the normal developmental pathway: apical regions split, forming the new apical notches, and development of the air chambers takes place. This behaviour of the isolated central regions was noted in all observed samples, out of which 32 had microscopy evidence for the first stage, and eight had microscopy evidence for the second and the third stages.

The cells of the isolated apical regions continued to proliferate and followed a developmental pathway similar to that of the apical regions in the intact tissue (Figure 4.4 vs Figure 1.3). Eight out of eight isolated and imaged apical regions followed the same developmental course.

These results suggest that the cells of the central region are capable of reacquiring the proliferative state. However, this does not happen unless the apical regions are excised. Therefore, it was hypothesised that the apical regions act as lateral inhibitors of cell proliferation. This hypothesis is in agreement with the mechanism *No2*, i.e., apical regions acting as sources of a growth inhibitor, which is transported to and accumulated in the central region (Figure 4.1 A). Upon excision, the influx of the inhibitor to the central region ceases, decreasing its concentration and causing cell proliferation (Figure 4.1 B). Later, the regeneration of the apical precursor regions restarts inhibitor production and export, thus reestablishing the boundary between the apical and the central regions.

The other proposed mechanisms do not agree with the observed regain of proliferation in the isolated central region, as the central region would maintain high concentration of a growth inhibitor under assumptions of the mechanism *No3*; would maintain low concentration of a growth inducer under assumptions of the mechanism

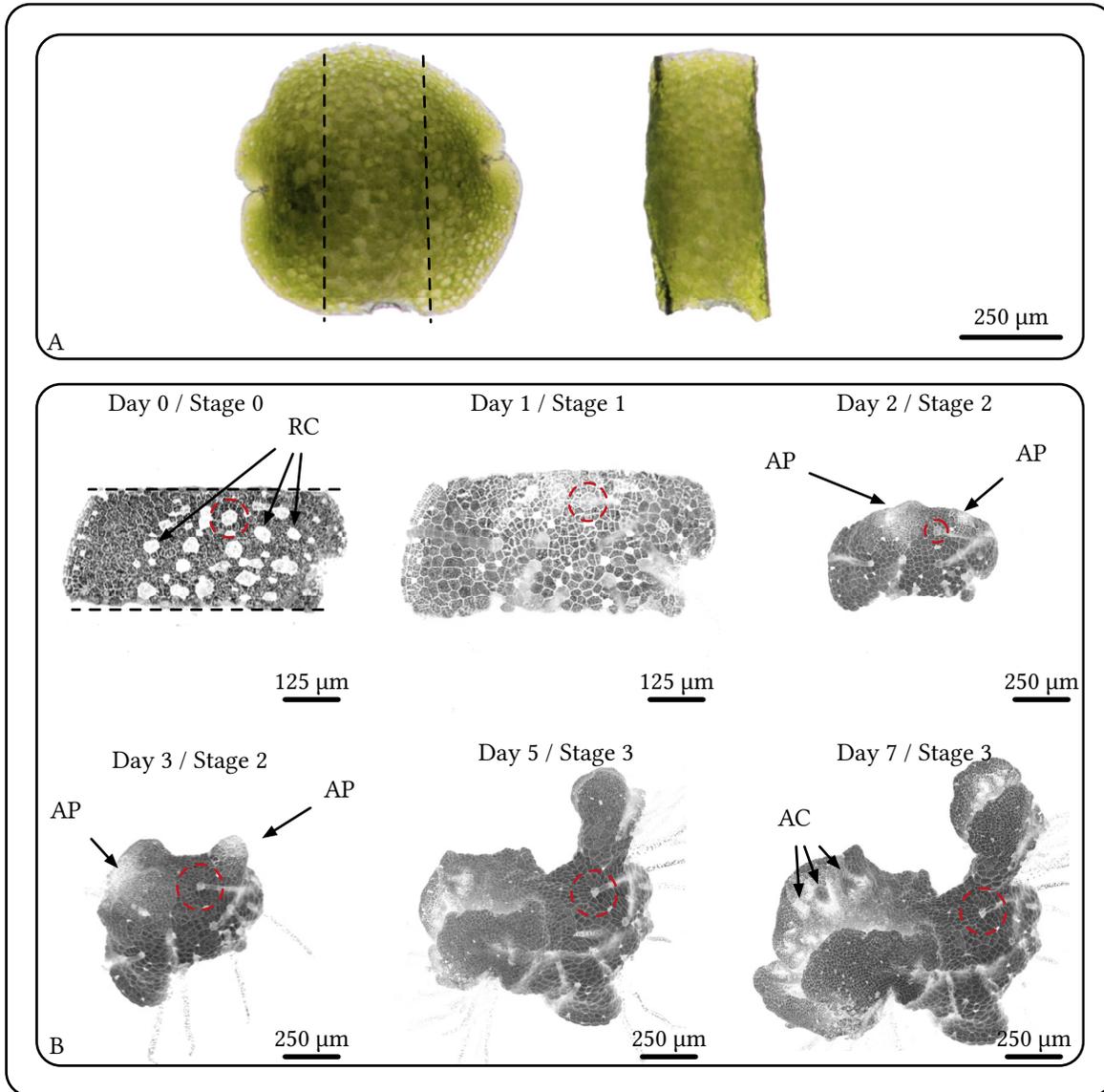


Figure 4.2: **Surgical assay.** **A.** colour photographs of a gemma before and after excision of the apical regions. The dashed lines denote the cut sites. **B.** An example of the regenerating central region after excision of the apical regions captured by the laser-scanning confocal microscopy. The contrast in the images was achieved by overlaying the signal from a membrane-localised fluorescent protein under control of the $proMpEF1\alpha$ and inverted signal of the chlorophyll autofluorescence. The *red circles* mark the same rhizoid cell at the different stages of the regeneration, AP denotes apical precursor regions, AC denotes air chambers, RC denotes examples of rhizoid cells, dashed lines denote the cut sites.

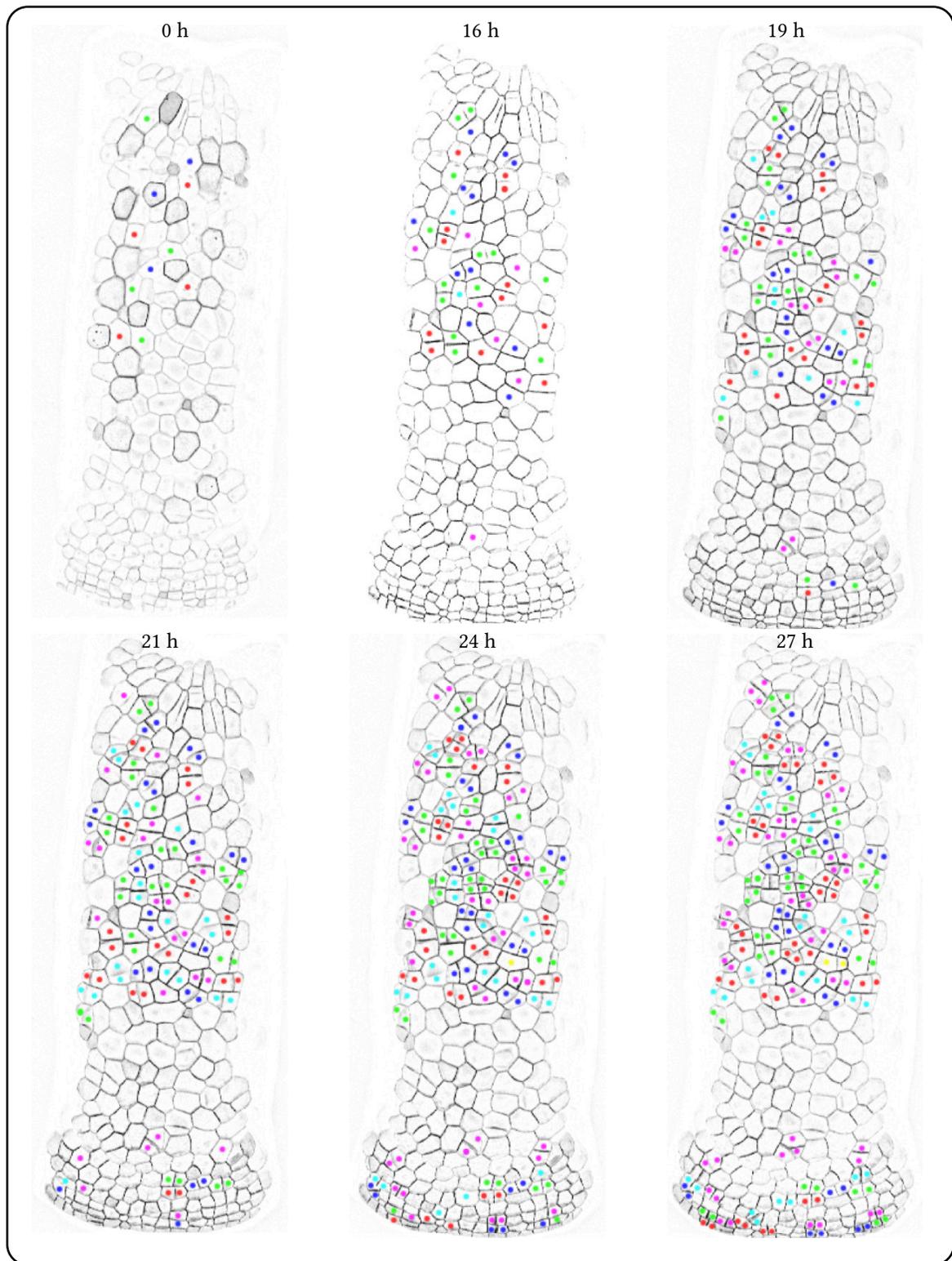


Figure 4.3: **Tracking cell divisions in the isolated central regions.** The figure shows inverted signal of the membrane-localised fluorescent protein under control of the *proMpEF1 α* in the isolated central region of a gemma. The dots mark cells that are either committed to division at the following time frame or have divided since the previous time frame. The daughter cells maintain the colour of the marking after the division. The cells that share a wall and are marked with the same colour have originates from the same parent cell. The titles above the images denote the time after excision of the apical regions.

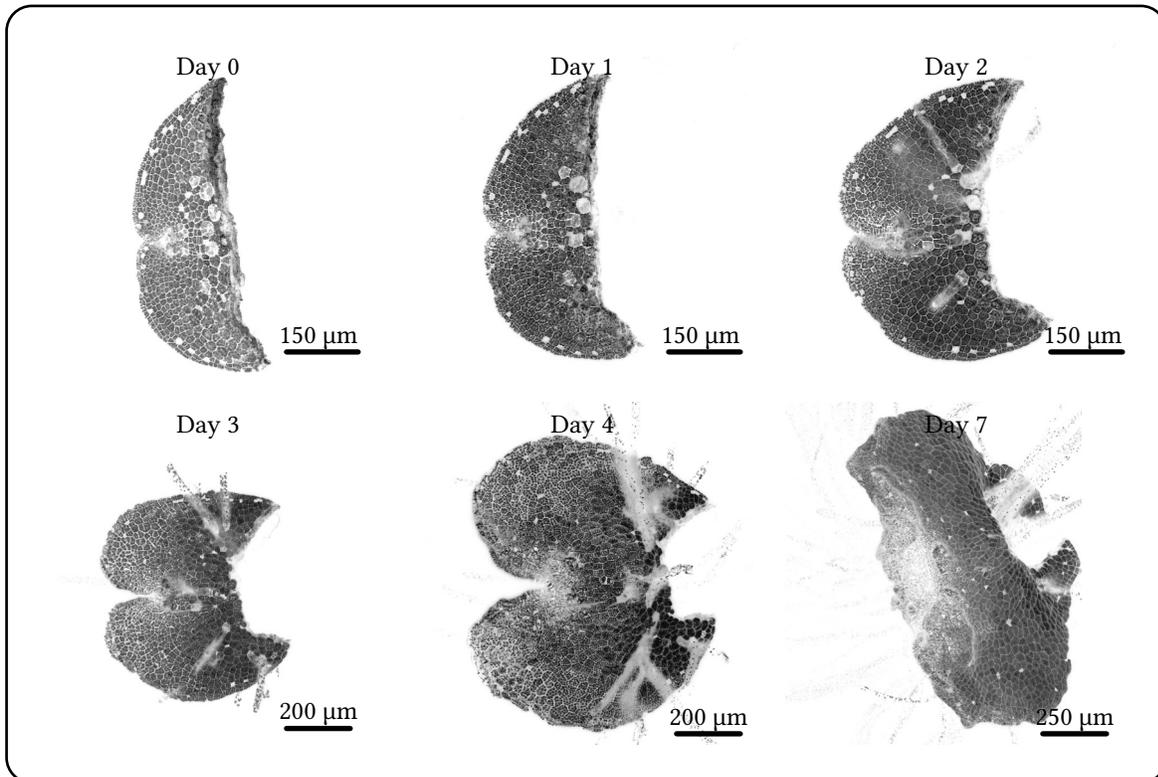


Figure 4.4: **Regeneration of the isolated apical regions.** Excised apical regions continue to follow the normal developmental pathway, i.e., expansion of the tissue during the first four days, followed by dorsal-ventral splitting of the notch. The contrast in the images was achieved by overlaying the signal from a membrane-localised fluorescent protein under control of the *proMpEF1α* and inverted signal of the chlorophyll autofluorescence.

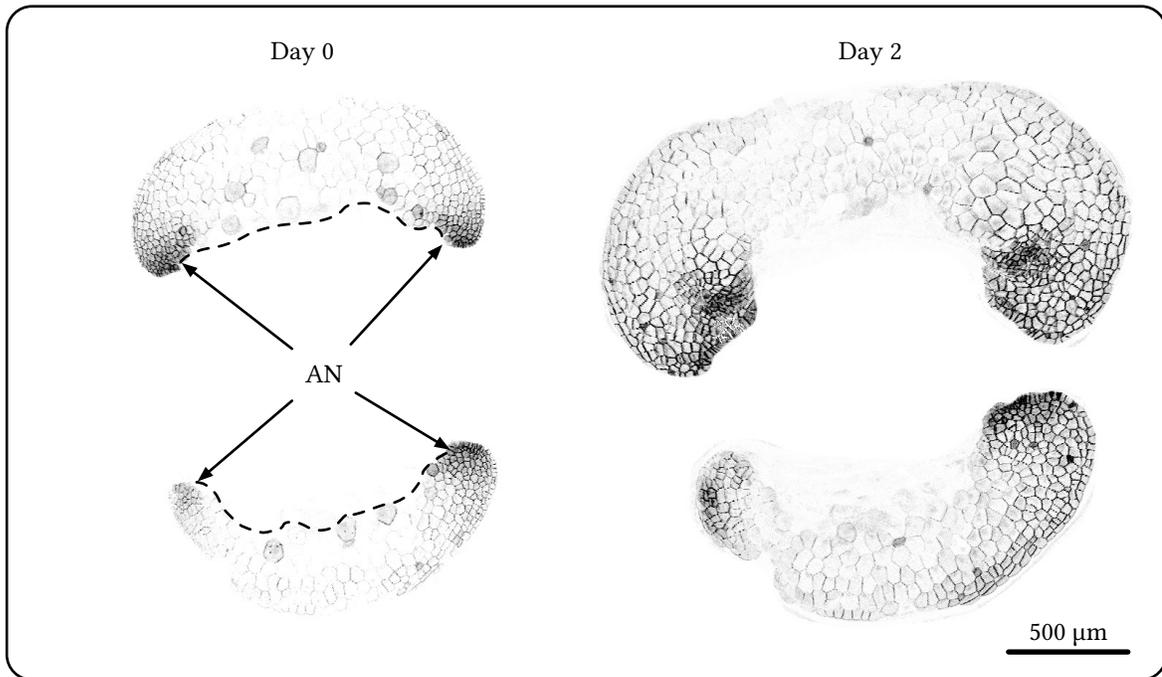


Figure 4.5: **Growth of a gemma split along the line connecting the two apical notches.** The figure shows inverted signal of a membrane-localised fluorescent protein under control of the *proMpEF1α*. AN denotes the location of the apical notches, the dashed line shows the cut site.

No 4; or would not receive any growth inducer under assumptions of the mechanism *No 1*.

In order to verify that the cell proliferation in the isolated central region resulted from the disturbance of the signalling between the apical and the central regions, rather than being related to a wounding response, young gemmae were cut along the imaginary line connecting the two apical notches. In this scenario, the split apical and central regions maintained their fates, and no division of cells in the central region was observed ($n = 2$, Figure 4.5).

4.4 Role of auxin in cell proliferation

In the previous sections, it was postulated that the cell division and expansion in *Marchantia* gemmae might be controlled by an inhibitor of the cell proliferation, produced in the apical regions. However, it was not clear whether a morphogen that fits the proposed hypothesis exists in *Marchantia*.

External application of auxin is known to restrain growth^[186] and promote dormancy^[189] of the gemmae. Therefore, auxin was selected as a primary candidate for the proposed inhibitor of cell proliferation.

In order to examine the cells' response to the exogenous auxin, gemmae ($n = 8$) were removed from the air chambers of a mature thallus and placed into an imaging chamber (Chapter 3, subsection 2.3.1) filled with solid medium

supplemented with $100 \mu\text{g}/\text{L}$ of synthetic auxin, 1-Naphthaleneacetic acid (NAA). The chamber was mounted on the stage of a laser-scanning confocal microscope, and the growth of the gemmae was examined for 48 h (Subsection 2.3.3). During the first 24 h, the cells of the apical region elongated without progressing through a cell cycle, and by the end of the 48 h period elongation of the cells ceased (Figure 4.6 A). When applied to the isolated central regions, auxin repressed their regeneration (Figure 4.6 B).

These experiments support the hypothesis that auxin acts as an inhibitor of cell proliferation in *Marchantia* and is transported from the apical regions to the central region, where it accumulates.

4.5 Discussion

In this chapter, a tissue-level regulation of *Marchantia* gemma growth was studied. In particular, the question of the maintenance of a border between actively proliferating apical regions and the stationary central region was examined.

A mathematical source–sink transport model was developed and analysed to propose four different mechanisms of the growth pattern maintenance. The possible mechanisms were tested by performing a surgical assay, where the apical and central regions of a gemma were physically separated, and their growth behaviour was studied in isolation from each other. While the cell of the isolated apical regions maintained in the proliferative state, the non-mitotic cells of the central region reentered the cell cycle and formed new apical regions, which then drove the regeneration of the tissue. This observation was in agreement with only one of the proposed mechanisms, namely production of a cell-proliferation-inhibiting morphogen in the apical regions that is transported to and accumulated in the central region. Proliferation of cells in the isolated central regions, as well as the apical regions of the intact gemmae, was suppressed by application of the exogenous auxin, suggesting its potential role as the proposed growth-inhibiting morphogen.

Although the proposed hypothesis was supported by the results of the surgical assays, there was no direct evidence presented about the spatial pattern of auxin production and accumulation predicted by the hypothetical mechanism. In order to present such evidence, the endogenous *Marchantia* genetic parts involved in auxin biosynthesis and response can be extracted from the genome and assembled in fluorescent reporter circuits. These circuits can be further introduced into the *Marchantia* genome to study the spatial patterns of their expression. A framework for the mining and extraction of functional genetic parts from the *Marchantia* genome is presented in chapter 5, while the design and analysis of auxin transport reporters is presented in chapter 6.

The model-driven top-down approach taken in this chapter was a key factor for designing an experiment that tested and rejected the state of the art model behind the regulation of tissue expansion in *Marchantia*. A criticism may be raised about the assumption that the tissue expansion rate is controlled by a single morphogen, auxin, whereas the experiments in other plant systems suggest several other morphogens being involved. A recent study done in *Marchantia*^[190] presented evidence that cytokinins might playing an antagonistic role to auxin, i.e. promoting tissue expansion.

For the purpose of this thesis, the involvement of cytokinins was not studied further based on the evidence that the addition of cytokinin to the growth media did not influence growth of the gammae during the first seven days,

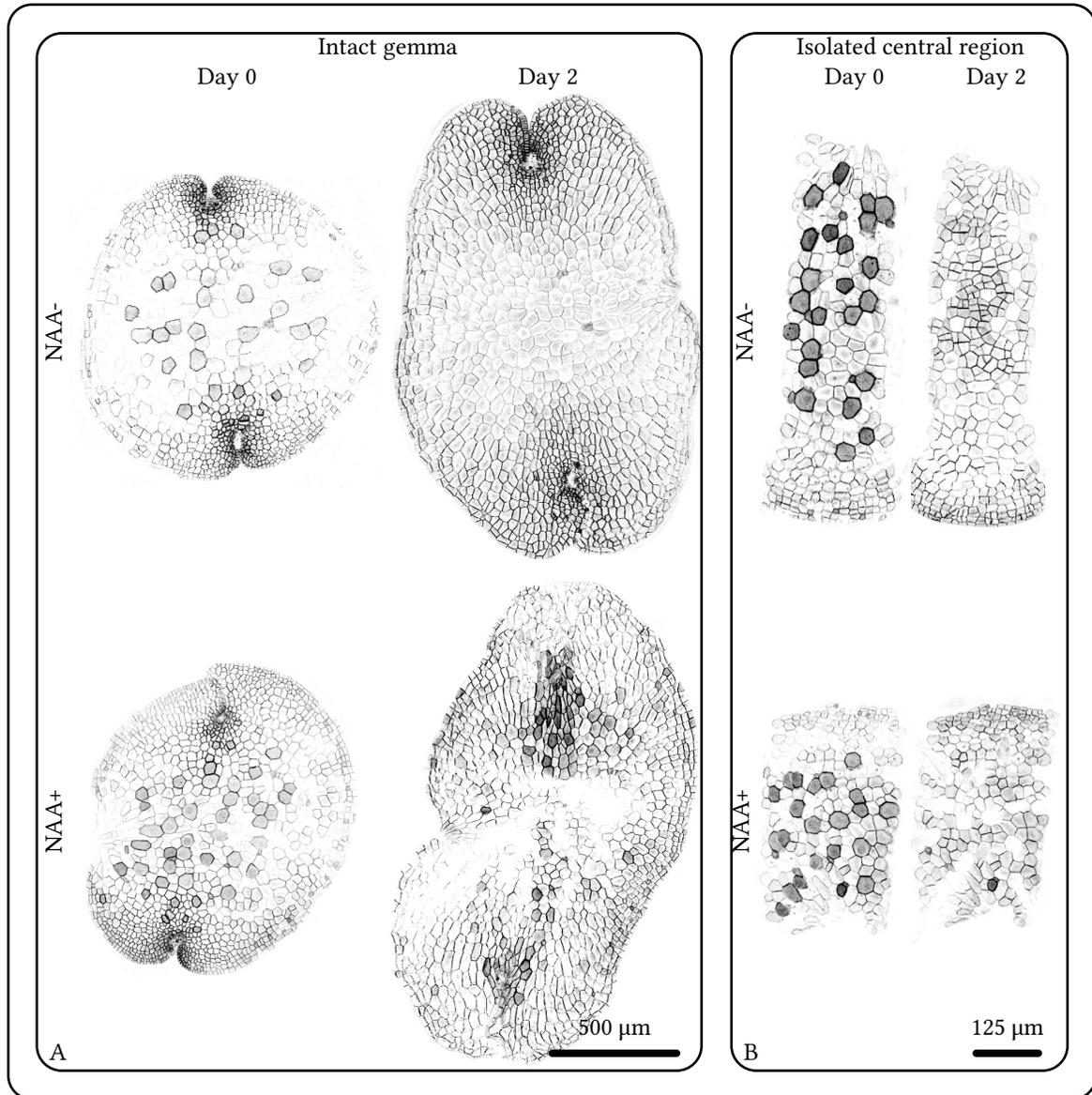


Figure 4.6: **Cellular response to exogenous auxin.** Signal of the membrane-localised fluorescent protein under control of the *proMpEF1α* in the **A.** intact gemma and **B.** excised central region. In both cases, external application of the NAA restrains cell proliferation.

nor did it qualitatively affect the regeneration process described in this chapter (Appendix B: Figure B.1, Figure B.2).]

Despite simplicity of shape, gemmae have several limitations as a platform for studying *Marchantia* morphogenesis. First, a gemma picked from a gemma cup already has an established cellular and tissue architecture. Therefore, the processes leading to the formation of this architecture are inaccessible to observation. Also, the apical regions of a gemma have highly curved boundaries that obstruct the view of tissue in proximity to the apical notches. As demonstrated in this chapter, cells of an isolated central region of a gemma undergo a series of divisions, followed by the establishment of new apical and central regions. Therefore, the isolated central regions can be used to study the differentiation of cells in *Marchantia*, and due to their flat geometry and availability of material, they are more accessible for this purpose compared to intact gemmae or sporelings at the early stages of development.

Chapter 5

MarpoDB

The work presented in this chapter was done in collaboration with Bernardo Pollak, who performed DNA and RNA extraction from the *Cam-1* accession of *Marchantia polymorpha*, prepared sequencing libraries, assembled the genomic and transcriptomic reads, and contributed to the frontend module. Mihails Delmans developed the software and implemented the data model.

Delmans, M., Pollak Williamson, B. & Haseloff, J., 2017, *MarpoDB: An open registry for Marchantia polymorpha genetic parts*. *Plant and Cell Physiology*. 58, e5.

Pollak Williamson, B., 2017, *Frameworks for Reprogramming Lower Plant Systems*. PhD Thesis, University of Cambridge.

5.1 Introduction

Advances in plant synthetic biology^[191,192] allow for the design and assembly of synthetic gene constructs, as well as their introduction into the plant genome. The modular representation of genes has been adopted by the plant synthetic biology community; the common syntax^[92], defining the standard for genetic parts for Type IIS restriction enzyme DNA assembly^[147,193–195], has been established.

The parts encoding promoter elements, five prime untranslated regions (5' UTRs) and coding sequences (CDSs) are of particular interest for the study and modification of developmental processes. Promoters and 5' UTRs can be fused to the coding sequences of fluorescent proteins to study the patterns of the endogenous gene expression. Similarly, the CDSs of endogenous genes can be fused to fluorescent proteins to study their sub-cellular localisation. In addition, by swapping endogenous or introducing exogenous regulatory elements, one can rewire the existing genetic circuits in an attempt to control the natural developmental processes.

In order to amass *Marchantia* genetic parts, genome-mining tools that enable the searching of genes, and their functional elements, by conserved protein motifs and homology with other plant systems, should be developed. Ad-

ditional tools should be developed for the extraction and preparation of part sequences for DNA assembly, generating registries of plant DNA parts.

A vast amount of data linking the function of proteins to their sequences has been collected over decades of biological research^[102,175]. This data comes from genetic screens, comparative sequence analysis, studies on protein localisation and structure, as well as research on protein–protein and DNA–protein interactions^[196]. This information is now available through public databases and can be projected onto any organism, whose genome has been sequenced, in order to infer the function of its genes and assist in choosing the right candidates for modification.

Such a projection has been performed on the *Marchantia* genome^[132]. The most notable discovery resulting from the study was a low genetic redundancy compared to other plant model systems. Whereas model plant systems contain gene families of up to several tens of members, *Marchantia* often has one or two representatives. This particular feature makes *Marchantia* an accessible platform for experimental genetics. However, the way the *Marchantia* genomic data had been stored is not directly applicable to its downstream use in genetic engineering. The official release of the *Marchantia* genome, hosted at Phytozome^[93], represents the data from a genome-wide perspective, with an emphasis on the relative location of the transcribed regions. This representation, although biologically accurate, does not bring out the modular structure of the genes and therefore is not tailored to the purposes of synthetic biology. Furthermore, the data also include the intergenic regions, which, while potentially functional^[197], are not commonly utilised in synthetic gene designs.

To better fit the engineering perspective, each gene can be represented as a modular, context-independent structure, composed of functional parts that can be disassembled and rearranged into new gene constructs. Additionally, the intergenic regions can be removed, transforming the genome into an unordered collection of genes.

In this chapter, the *Marchantia* genome was reduced to a collection of genetic parts. The collection features modular gene models, compliant with the common syntax; functional annotations of the putative proteins, provided by the InterPro database^[102], and comparative analysis with other plant species, powered by the UniProt database^[175]; and a simple and intuitive user interface, along with tools for the extraction and preparation of part sequences for DNA assembly.

5.2 Software architecture

MarpoDB is a software package that manages the database of *Marchantia* genetic parts. An overview of the software architecture is presented in figure 5.1 and is discussed in more detail in the following subsections.

5.2.1 Data model

The conventional data model of genome resources is built around the DNA sequencing data. The genome is represented as several strings that encode its sequence, one for each chromosome. In addition to the raw sequence data, a representation of a genome usually contains annotation data, otherwise known as *features*. Each feature record contains information about which DNA sequence (chromosome) it annotates, the index of the start and end of the

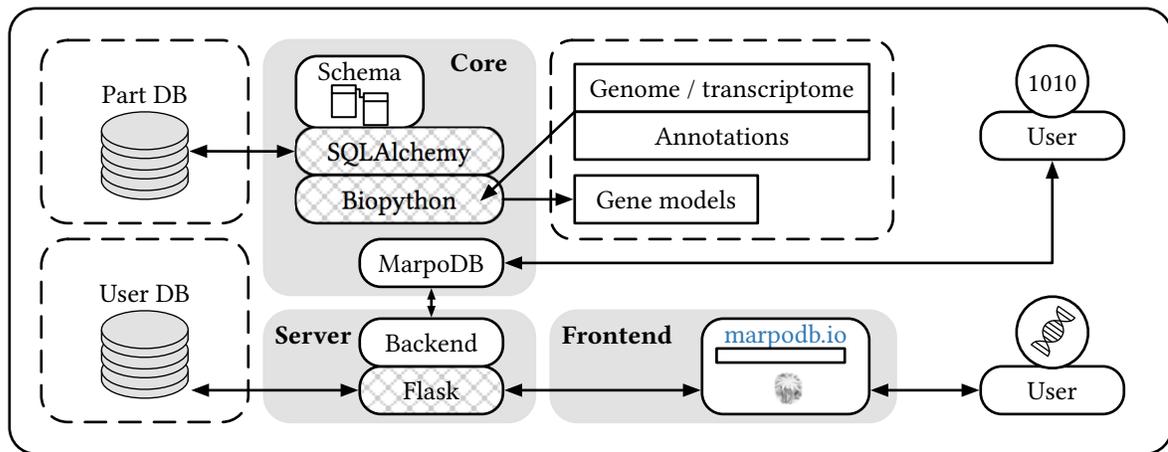


Figure 5.1: **MarpoDB software architecture.** MarpoDB is comprised of three modules. The core module handles import, export and storage of the genetic parts. A *MarpoDB* submodule of the core module abstracts the underlying data processing and provides a high-level access to the database for the *server* module, or directly for the *user*. The *frontend* module provides a graphical user interface for querying the database, exploring the annotation data and handling the DNA sequences. The *server* module acts as a bridge between the core and the *frontend*; it translates HTTP requests coming from the *frontend* into the calls to *MarpoDB* submodule using functions defined in the *backend* submodule. MarpoDB modules are shown in grey rounded rectangles; the submodules are shown in white rounded rectangles; hatched submodules represent third-party libraries; arrows denote the flow of information; data are represented by rectangles, databases by the stack of discs.

annotation relative to the start of the sequence, the strand of the DNA being annotated, the type of the feature and a list of tag-value pairs for additional feature attributes. In relation to the genome sequence, the features are used to annotate the transcribed regions, introns and exons, coding regions and untranslated regions. A collection of coding regions with introns removed and translated protein sequences are usually stored along with the raw genomic sequences and their annotations. The features of the protein sequences contain information about conserved protein domains.

The benefit of such a data model is that it captures the comprehensive set of data that accurately represents the genomic and functional aspects of a genome. This model, however, does not fit well the representation of a gene that is used in synthetic biology and where the functional elements of a gene, such as promoter, coding sequence and terminator, are considered to be modular and independent genetic parts. For a synthetic biologist, a genome is a resource from which the sequences of these modular genetic parts can be extracted. Not much consideration is given to the intragenic regions, which do not contain any coding regions, and hence cannot be used for mining the genetic parts. In order to represent *Marchantia* genome in a form that is more suited for synthetic biology application, a novel data model was developed.

In a classical relational database management system, data are divided into tables, each storing a particular type of records. Each row of a table corresponds to a single record, while each column to a data field that stores information of a particular type. Each table has a column that stores an identifier (ID), also known as a primary key, that is uniquely defined for each record. Two tables can be relationally linked together by storing the primary keys of each other's records. Combination of tables and their relations define a data model, also known as a *database schema*. The overview of the MarpoDB schema is shown in figure 5.2..

The central part of the MarpoDB schema is the *gene* table that stores a unique record for each *Marchantia* gene. The gene records do not store the DNA sequence themselves; it is instead distributed between the records of the corresponding parts, defined in compliance with the common syntax: promoter, five prime untranslated region (5' UTR), coding sequence (CDS), three prime untranslated region (3' UTR) and terminator. Each part type is stored in a separate table, and each gene record stores the relational links to the records of its five parts. In other words, the genes are represented as if they were assembled from a predefined repository of parts, where each part is only used once in a unique gene construct. Each gene is related to a record in the *dataset* and the *locus* tables that preserve information about where it was extracted from and position of the gene in the corresponding genome assembly. In addition to the genomic sequence, *CDS*, *5' UTR* and *3' UTR* records store the coordinates of the exons relative to the part sequence. This information allows reconstruction of the sequence of the final transcript of a gene.

Apart from the gene models, MarpoDB also stores annotation data that relates to the records of the parts. Each type of annotation data is stored in a separate table. However, all annotation tables share the same structure. Each annotation table has a *target ID* field, which stores primary keys of the corresponding part records, and a number of fields that store the information about the annotations themselves. The main fields of the annotation tables are: *origin*, *description* and *reference ID*. The first two store textual information about the nature of the annotation, while the third stores a cross-reference to an external annotation database. These fields create a list of keywords that can be queried against by the user in order to find the related gene of interest. In addition, the annotation tables have

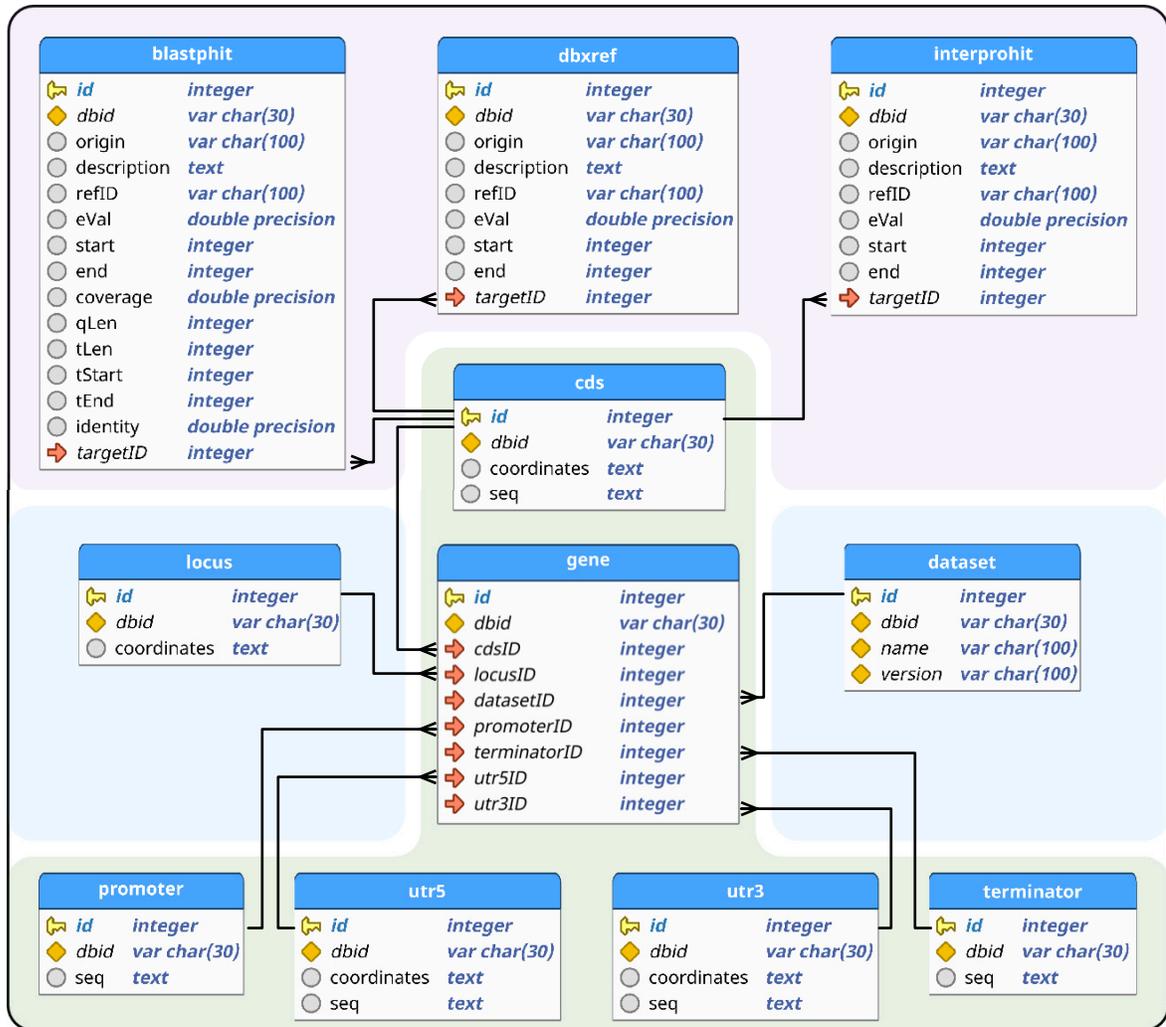


Figure 5.2: **MarpoDB schema.** Tables on a green background are related to a gene model. Blue background denotes tables that store gene annotation; purple coding sequence (protein) annotation. Fields marked with grey circles denote non-unique data; a key symbol denotes a primary key; red arrow a foreign unique ID from a relational table; yellow square a field that stores a secondary unique ID (dbid). Relationships are shown in black arrows.

e-value, *start* and *stop* fields. These fields store information about the significance of the annotation and its location relative to the sequence of the part. Apart from these six fields, annotation tables are allowed to have extra columns, if required.

All records in MarpoDB have a secondary unique identifier (*dbid*), in a format *mpdb.table.xxx*, where *table* is a name of the table and *xxx* is a primary unique ID, e.g., *mpdb.cds.1243* for a *CDS* record and *mpdb.blastphit.1134* for a BLAST annotation record.

5.2.2 Core module

After the data model was defined, the *core* module for handling the data was developed. The first function of the *core* module was to communicate with the SQL database, e.g., to process the queries, add or remove records from the tables. The second function of the *core* module was to handle the DNA sequences: conversion the raw genomic and transcriptomic data into a gene-centric representation, import of the annotations and export of the part sequences. The functionality of the core module was encapsulated through a *MarpoDB*¹ class, which provided a high-level interface for management of the database and the DNA sequences.

5.2.3 Frontend module

The core module was sufficient for establishing and operating a gene-centric database with given genomic and transcriptomic data. However, operating the core module required Python scripting skills, and was not convenient for an average user with no prior computational background. In order to create a publicly-available and accessible user interface for MarpoDB, a *frontend* module was developed. It was implemented as a web application^[198] that allowed the remote database to be queried by a keyword search, BLAST^[176] search or HMMER^[199] search. An overview of the *frontend* functionality is presented in Figure 5.3A.

The home page of MarpoDB presents an input field for querying the database by a keyword. Upon the query submission, the keyword is compared with the *description*, *origin* and *reference ID* fields of every annotation record present in the database. The user has an option to select a particular dataset or annotation tables to query against. All the annotation records with fields that start with the submitted keyword are grouped by the corresponding gene identifier and are presented to the user in a table. Alternatively, the user can submit a DNA or a protein sequence to search for homologous gene records using BLAST algorithms. The BLAST search can be customised by selecting a desired algorithm, similarity thresholds and a dataset to query against. For finding sequences, belonging to a protein family, a HMMER search can be used. It allows to submit a protein sequence alignment file, and finds proteins with the similar sequence patterns.

From the table of results, the user can navigate to a *details* page of a particular gene, which displays an interactive gene model, information about the associated annotations, gene name and identifiers in the other Marchantia databases. The interactive gene model can be used to select a particular part of the gene. Upon selection, the corresponding sequence highlights in the sequence viewer below. The selected sequence can be then copied to the

¹Note an ambiguity in naming. *Italic version* relates to a Python class, while *non-emphasised* to the whole framework.

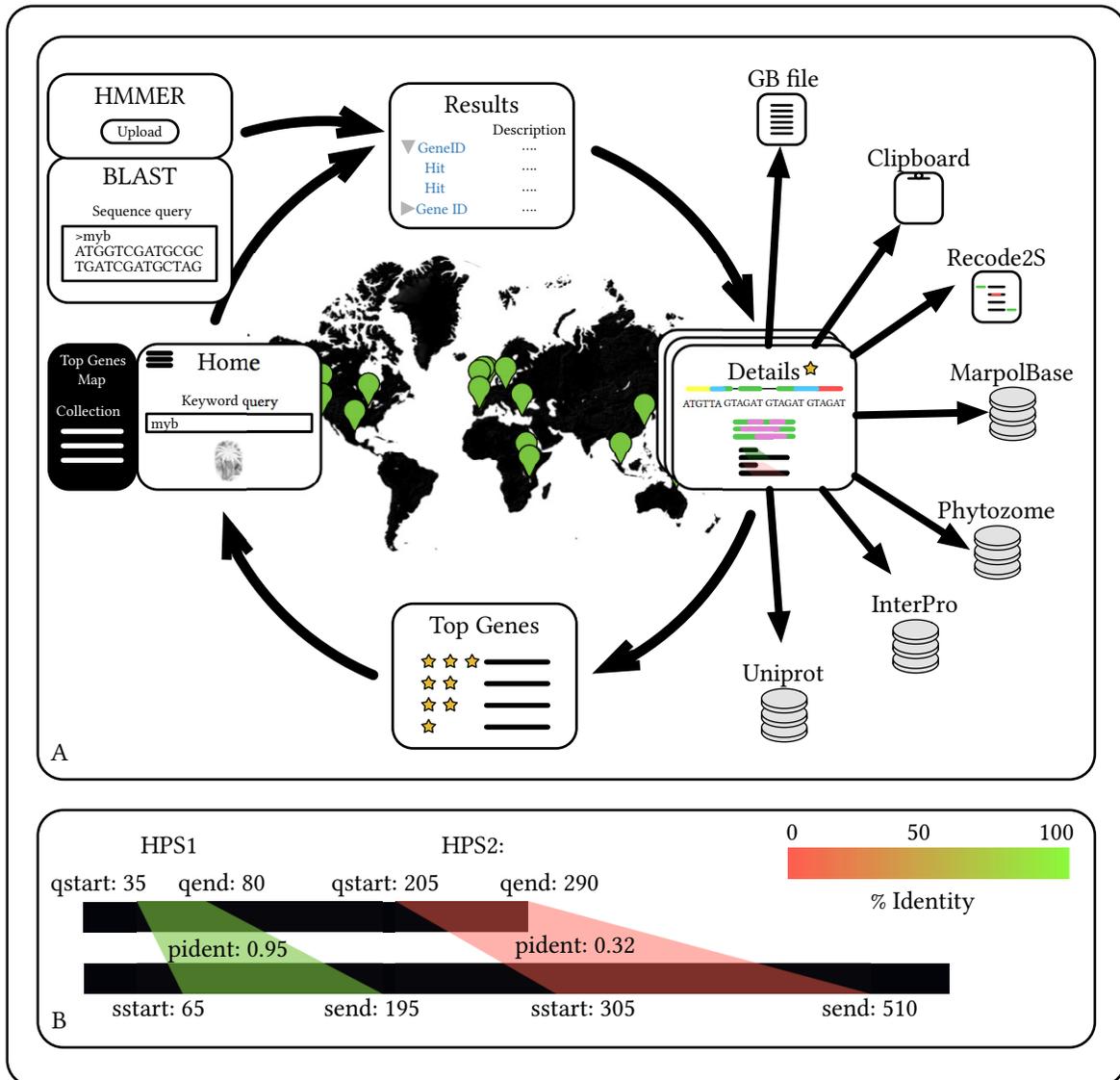


Figure 5.3: **MarpoDB frontend.** **A.** A schematic representation of the frontend workflow. The queries to the database are made either through a keyword, BLAST or HMMER search (left). The results of the query are displayed in a table (top). Each row of the table links to a *Details* page that presents an interactive gene model, sequence tools and annotation data from the external databases (right). Users can assign stars to the gene records, which increases the position of the gene in the popularity rank (bottom), and adds the record to a personal collection (right). The world map, marking the locations of MarpoDB users, is shown in the centre. **B.** The BLAST hit visualisation tool. The top and bottom black rectangles represent a query and a target sequences respectively. The lengths of the rectangles are proportional to the lengths of the sequences. High scoring pairs are represented by coloured trapezoids. The coordinates of trapezoid vertices denote the location of the overlap and colour denotes the identity score.

clipboard or domesticated with the help of the embedded *Recode2S* tool^[131,200], for its downstream use in a Type IIS DNA assembly. The domestication refers to addition of flanking sequences containing the recognition site of the enzyme used for the assembly together with the overhangs specific for the part type (Section 2.2), and removal of the recognition sites from the rest of the sequence by introduction of a silent mutation. In addition, a complete annotated sequence of the gene can be exported in a GenBank^[201] file format. The *details* page also contains annotation of conserved protein motifs and a custom BLAST hit visualisation tool (Figure 5.3 B) for exploring the comparative data between other plant species.

Apart from sequence search and export, the *frontend* provides an opportunity to save genes in a personal collection by pressing on a star image next to the title of the *details* page. Not only this allows to save the search results, but also increases the rating of the gene. The top-rated genes are listed on a special page for a quick access.

Finally, a *map* page shows geographical location of all MarpoDB users.

5.2.4 Server module

In order to serve the *frontend* to users across the globe, a *server* module was developed. It was used to translate the HTTP requests coming from the *frontend* into the calls to *MarpoDB* class in the core module. The translation of the requests was implemented in the *backend* submodule of MarpoDB. In addition, the *server* module handled the database of the users and their gene collections.

The MarpoDB source code was released under an MIT license^[164], along with a snapshot of the database. A local version of the database can be recreated on a local machine, such that users can have unrestricted access to the data.

5.3 Datasets and annotations

The previous section discussed the data model and the implementation details of MarpoDB. The following subsections will describe the pipeline that was implemented in order to collect Marchantia genomic and transcriptomic data and transform it into the gene-centric representation.

5.3.1 Genome and transcriptome

Initially, MarpoDB was developed to store genomic data of the *Cam* accession of Marchantia (Subsection 2.1.1). The collection of the genome and transcriptome data of the *Cam-1* accession was performed by DNA and RNA isolation from the young male Marchantia gametophyte, library preparation and sequencing, and subsequent *de novo* assembly of the short reads. As a result, 15,703 genome scaffolds with a total length of 195.3 *Mbp* and 93,315 mRNA isoforms were used as a starting point of the population and annotation pipeline.

In order to separate the coding isoforms from the non-coding ones, a CDS prediction algorithm, offered by a Transdecoder package^[96], was applied. The algorithm selected only those isoforms that contained a potential CDS, i.e., having length of a multiple of three, starting with a start codon (ATG) and ending in one of the three stop codons (TAA, TAG, TGA). In addition, the CDSs were further filtered by comparing the encoded protein sequences with the

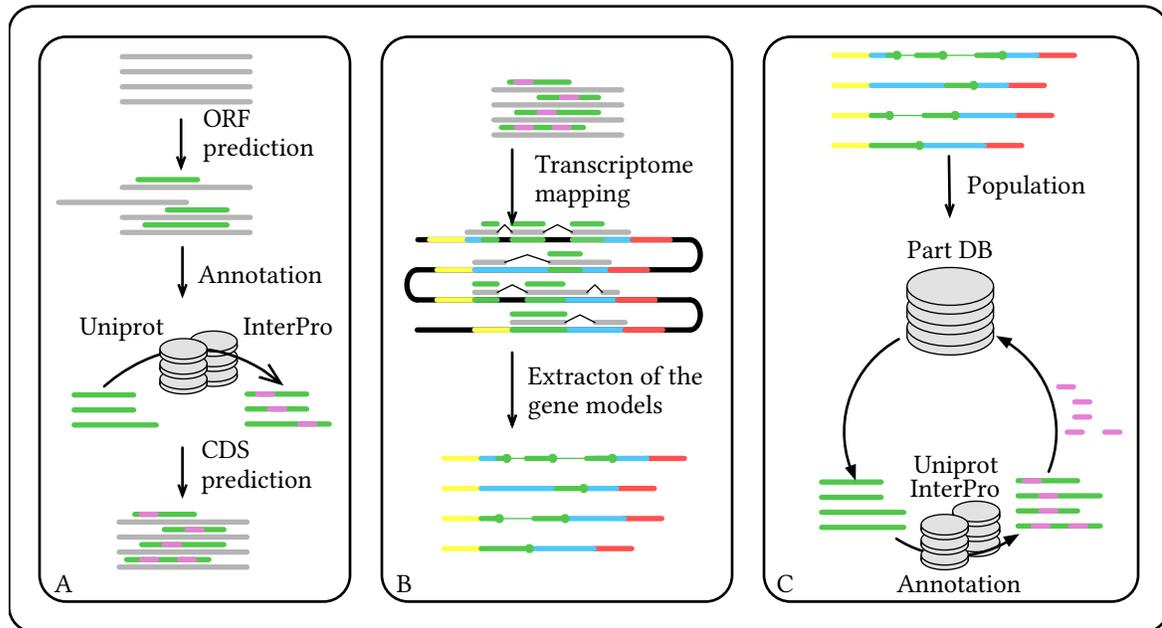


Figure 5.4: **MarpoDB population and sequence annotation pipeline.** **A.** The sequences of the assembled messenger RNA isoforms (grey lines) were passed through a coding sequence (CDS) prediction software. The sequences of the potential encoded proteins (green lines) were further subjected to a comparative analysis with known plant protein sequences and identification of the conserved protein motifs. The isoforms containing at least one predicted CDS and at least one match with either plant protein collection (Uniprot), or protein motif database (InterPro), were considered positive and were used in the next stage of the pipeline. **B.** The selected isoforms were mapped against the genome assembly (black line), which allowed to annotate the gene parts: putative promoters (yellow) and terminators (red), coding sequences (green) and untranslated regions (blue). **C.** The gene models, and the corresponding sequences were then used for the database population. Translated CDS sequences were then exported from the database and annotated using the similar tools as for the CDS prediction. The annotation data (purple lines) were then imported back to the database.

sequences of known plant proteins and conserved protein motifs. The isoforms with at least one potential CDS, and at least one match with the known protein sequence were selected for the following step of the pipeline (Figure 5.4 A). MarpoDB did not focus on preserving the accurate and detailed biological representation of the genome, but rather on creating a functional and concise gene-centric resource. It is a common practice to use the sequence of the longest CDS for downstream molecular and synthetic biology applications. To facilitate this, only the longest CDS for each selected isoform was kept.

Next, the selected isoforms were mapped onto the genome assembly, and the genomic sequences contained within the mapped regions were extracted together with 3 kb flanks upstream and downstream as proximal promoters and terminators (Figure 5.4 B). In total, sequences of 12,722 gene models were extracted from the *Cam* dataset.

In parallel to the *Cam* accession sequencing, the *Tak* accession was sequenced as a part of *Marchantia polymorpha*

genome project^[132]. The *Tak* genome version 3.1, together with annotation of exons, was acquired through Phytozome portal and was used to generate the gene models, following the same approach as for the *Cam* dataset, i.e., extracting the transcribed regions with 3 *kb* upstream and downstream flanks. In total, 19,287 gene models were extracted.

The gene models, acquired from both *Cam* and *Tak* datasets, were imported to MarpoDB using the functionality of the core module, described earlier. The translated CDS sequences were then exported back from MarpoDB, and were analysed using the annotation tools described below (Figure 5.4 C). The export of the sequences was required to retrieve the unique identifier, associated with the sequence, which was then used to create relational links to the corresponding annotation records.

5.3.2 Annotation pipeline

InterProScan5^[101] software was used to find the conserved protein motifs in the protein sequences extracted from MarpoDB. In total, 248,434 motifs matches were identified and exported in CSV and HTML format. The former was used to populate the *interprohit* annotation table of MarpoDB and the later one was used for visualisation of the motifs at the *details* pages of the *frontend*.

In order to enable a comparative analysis between *Marchantia polymorpha* and other plant species, the protein sequences were exported from MarpoDB and aligned with the *Viridiplantae* protein sequences, extracted from the Uniprot database. In order to minimise the number of hits, and hence the time it would take to process a query, only the sequences having experimental evidence for existence were used for the BLAST alignment. In total, 278,607 sequences were used for the analysis resulting in 714,814 hits that were imported into the *blastphit* annotation table of MarpoDB.

Since the *Tak 3.1* dataset had been already published at Phytozome, the cross-database gene identifiers were added for each gene record from the *Tak* dataset, and stored in the *dbxref* annotation table with the *origin* field set to “Phytozome Tak3.1”.

Alignment of the *Cam* dataset reads against the *Tak 3.1* genome assembly^[202] revealed an average of 1.48 single nucleotide polymorphisms per 1 *kb* of sequence in the predicted exon regions (Table 5.1). In order to find the corresponding gene pairs between the two datasets, the open reading frame sequences and the complete gene sequences of the *Cam* dataset were aligned against those in the *Tak* dataset using the nucleotide BLAST algorithm. All alignments with at least 99 % coverage and with at least 99 % identity were considered as true homologs. In cases, where a *Cam* sequence had several true homologs in the *Tak* dataset, the one with the best score, defined by a product of coverage and identity, was chosen. In total, 11,309 out of 12,722 *Cam* genes have been assigned a homologous gene in the *Tak* dataset, which in total contained 19,287 genes (Figure 5.5). The discrepancy in the number of genes between the two datasets might be explained by the differences in the methods for genome and transcriptome assembly, transcriptome mapping and CDS prediction. For *Cam* genes with an identified homolog the corresponding Phytozome ID was added as an annotation record.

The MarpolBase^[203] resource hosted by the *Marchantia* Working Group and Genome Informatics Laboratory

Region	Number of SNPs	% of all SNPs	Length (bp)	% of total length	SNPs per 1kb
Intergenic	1,010,971	82	153,390,713	68	6.59
Intragenic	224,762	18	72,370,426	32	3.11
Exons	48,423	4	32,752,581	15	1.48
Introns	176,339	14	39,617,845	18	4.45
Total	1,235,733	100	225,761,139	100	5.47

Table 5.1: **Single nucleotide polymorphisms between *Cam* and *Tak* datasets.** The *intragenic* regions correspond to those covered by an open reading frame; the *intergenic* to those between the two *intragenic* regions. *Intragenic* regions are further separated into *Exons*(coding) and *Introns*(non-coding)

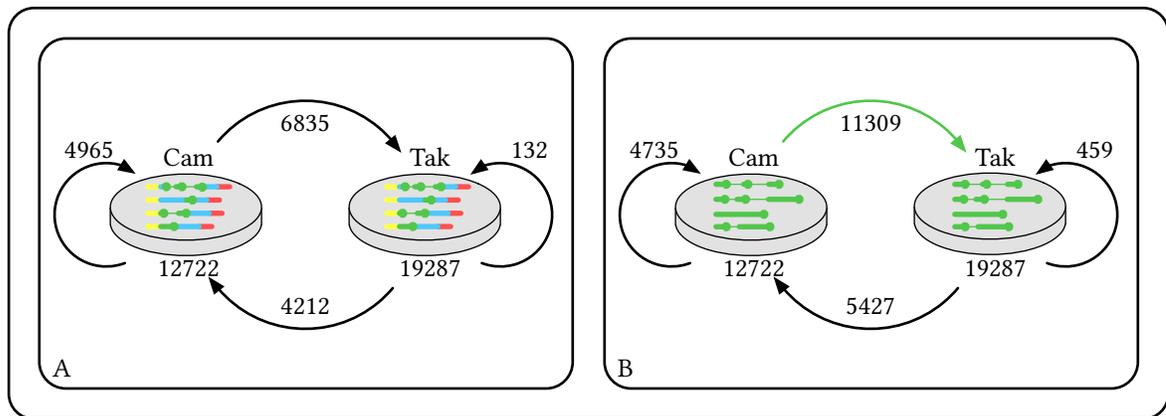


Figure 5.5: **Comparative analysis of *Tak* and *Cam* datasets.** **A.** A comparative gene sequence analysis. **B.** A comparative open reading frame sequence analysis. The datasets are represented by the grey cylinders, with the total number of sequences indicated below. The arrows denote comparison by BLAST, with the numbers showing how many sequences from the dataset the arrow is pointing from can be found within at least one sequence in the dataset the arrow is pointing to, with a minimal identity and coverage of 99%. Where the datasets are compared to themselves, the number excludes matches between sequences with identical ID, and represents the amount of duplication within the dataset. The green arrow denotes the comparison that was used for transferring annotation from the *Tak* dataset to the *Cam* dataset.

Name	MarpoDB ID	Phytozome ID	Description	Dataset
<i>MpYUCCA2</i>	mpdb.gene.8238	Mapoly0063s0040	Auxin biosynthesis	Cam
<i>MpAux/IAA</i>	mpdb.gene.52065	Mapoly0034s0017	Auxin response	Tak
<i>MpARF3</i>	mpdb.gene.51957	Mapoly0043s0098	Auxin response factor	Tak
<i>MpEF1α</i>	mpdb.gene.56948	Mapoly0024s0116	Translation elongation factor EF1A	Tak

Table 5.2: **The list of genes used in the work presented in this thesis.**

stores a database of the registered Marchantia gene names. The list of registered gene names was acquired, and the corresponding annotation records were added to the *dbxref* annotation table with the *origin* field set to ‘MarpoBase’.

5.4 Applications

The main advantage of MarpoDB compared to alternative resources is its application programming interface (API), which enables the batch retrieval of parts with a common functionality, and the modular nature of its stored gene models.

For example, I used MarpoDB to create a library of proximal promoters of Marchantia transcription factors (TFs). The protein sequences of all genes in the *Tak* dataset were exported and analysed using a TF prediction tool offered by the Plant Transcription Factor Database^[204]. In total, 398 genes were identified as belonging to one of the 51 TF families. The unique identifiers of the selected genes were used to query MarpoDB for the corresponding promoters. The promoter sequences were then extracted, domesticated using the *Recode2S* tool and sent for DNA synthesis.

MarpoDB was also used as a tool to identify genes and parts relevant to the work presented in chapter 6 (Table 5.2).

5.5 Discussion

In this chapter, a gene-centric database of the Marchantia genome was presented. The database allows one to search and extract sequences of gene elements for use in DNA assembly. In particular, the *Recode2S* tool allows *in silico* domestication of parts for Type IIS assembly. MarpoDB is especially efficient for the batch retrieval of parts and general programmatic access through the core module. Therefore, it can be easily integrated into applications that require access to genomic data. As well as low-level access, MarpoDB features a simplified web frontend for users lacking a computational background.

MarpoDB was used to create a library of synthetic transcription factor promoters, which can be fused to sequences of fluorescent proteins, creating reporters for the study of their expression patterns. Together with the recordings of the tissue expansion patterns presented in chapter 3, these expression patterns enable the observation of processes happening at the cellular and tissue scales. An example of this approach is presented in chapter 6, where the model of auxin-dependent growth pattern maintenance was validated using fluorescent markers of auxin biosynthesis and response.

MarpoDB provides fast access to the sequence of a particular gene part, and also allows the categorisation of annotation data for different functional aspects of a gene, e.g., annotation about promoter regulatory elements can be stored separately from the conserved protein motifs in the CDS. The benefit of this separation is that information about the parts, both sequence and annotation, can be distributed independently of other parts of the original gene.

The MarpoDB project was released open-source under the MIT license, and was designed with expansion in mind. New annotation tables can be added to the database schema without breaking the integrity of the data model. Therefore, new types of annotation data can be incorporated into the database as *Marchantia* research progresses. The core module has an inbuilt functionality to establish a local copy of the central MarpoDB resource, such that potential collaborators can have unrestricted access to both the source code and the data.

Finally, MarpoDB is agnostic to the source of the input sequences, and should be considered as a novel way to store genomic data, rather than a one-off resource designed specifically for *Marchantia*. In other words, MarpoDB's approach can be applied to genome and transcriptome data of any other organism in order to convert it into a gene-centric resource.

The comparative analysis between the Cam and Tak datasets presented in this chapter opens up a discussion about the sensitivity of the genome annotation result based on the chosen pipeline and starting material. Each step of the pipeline – DNA/RNA sample collection, library preparation, sequencing, genome assembly, transcript mapping, CDS prediction and annotation – has myriad approaches and implementations, which are in an active state of research. Therefore, each step introduces an additional source of noise to the final pipeline result, i.e. annotated genome.

Each genome assembly attempt will differ slightly from another, and without being able to compare to the ground truth, it is very hard to assess the quality of the final assembly. Long-read sequencing, like that offered by Oxford Nanopore Technologies, allows long reads of DNA to be produced and so might be a significant step towards increasing our confidence in the accuracy of genome assemblies^[205].

A second layer of variability comes from predicting genes inside the assembled genome. A classical approach, which featured in this chapter as well, is to map the transcript sequences to the genome assembly in an attempt to identify the transcribed regions that contain genes. The challenge lies in the fact that due to the transcript splicing, there may be several locations on the genome from where the transcript could have theoretically originated. Furthermore, the prediction of where the coding region is inside the transcript is also ambiguous and its quality is determined by the information gathered from the analysis of other genomes. The simplest way to predict a coding region without relying on any extra information is to consider all stretches of DNA sequence that start with an ATG codon and end in one of the stop codons. Usually there are several such stretches, while only one or several of them would actually encode a gene. However, without conducting a detailed experimental analysis, it is impossible to know for sure which of the potential CDSs is the true one. One may include additional information, like data on sequences of conserved protein domains, to make an educated guess regarding the accuracy of the CDS prediction. But even in this case, the decision relies heavily on the choice of the conserved protein domain database, its version and many choices of thresholds that would inform the binary decision. Since the collection of genomic resources is constantly growing, in order to keep up, the annotation pipeline should be re-run with every update of the databases used in the analysis. Although one may argue that the annotations of the genomes will eventually converge to a

single true state, it is not clear how much time is required for this to happen.

Taking into account this inbuilt noise, the way the genome resource is built and used should be in compliance with the biological question in mind. The approach presented in this chapter was aimed at providing a resource for engineering new life forms and hence was biased towards collecting less data but with the higher probability of it being true. While disregarding a lot of data that might hold valuable biological insight and form the basis for further basic questions, MarpoDB holds a collection of building DNA blocks that have a high probability of having the function predicted through the annotation pipeline; hence, it is more suitable for engineering purposes.

Chapter 6

Patterns of auxin production and accumulation

6.1 Introduction

In chapter 3, the growth of *Marchantia* gemmae was quantified, demonstrating that cell expansion and division rates decrease as cells progress from the apical to the central region of a gemma. In chapter 4, a mechanism that prevents cell proliferation outside of the apical regions was proposed. Based on the observation that the cells of surgically isolated central regions reenter a proliferative state (Section 4.3, subsection 2.1.5), it was hypothesised that cell proliferation in the central region of the intact gemmae is repressed by an apically-produced morphogen. The morphogen was proposed to be transported to and accumulated in the central region, while its concentration in the apical regions remained low, allowing cell division and expansion. According to this hypothesis, upon excision of the apical regions, the flux of the morphogen was disrupted, leading to the de-repression of the proliferative state in the isolated central region. The exogenous application of auxin was shown to prevent cell proliferation in the isolated central regions, as well as in the cells of the intact gemmae. Therefore, it was hypothesised that auxin is playing the role of the hypothetical proliferation-repressing morphogen.

In accordance with the postulated model, auxin should be produced in the apical regions, while accumulating in the central region of a gemma. In order to test the model predictions, evidence about spatial patterns of auxin biosynthesis and accumulation should be collected.

It has been demonstrated that the two-step tryptophan (TRP)-dependent auxin-biosynthesis pathway is essential for the normal development of *Marchantia*, and provides a limiting source of auxin (*IAA*)^[189]:



Two enzymes, tryptophan aminotransferase (TAA) and flavin monooxygenase (YUCCA) are involved in the pathway:

TAA catalyses the transformation of TRP into Indole-3-propionic acid (IPA), while YUCCA catalyses the transformation of IPA to IAA. In *Marchantia*, TAA is encoded by a single copy of the *MpTAA* gene, while YUCCA by two *MpYUCCA* genes (*MpYUCCA1* and *MpYUCCA2*). However, only *MpYUCCA2* is expressed in the gametophyte^[189] and hence is relevant to gemma development. Therefore, by finding the spatial concentration profile of the YUCCA enzyme it is possible to estimate the pattern of auxin synthesis. One way to do this is by fusing the promoter sequence of the *MpYUCCA2* gene (*proMpYUCCA2*) to the coding sequence of a fluorescent protein. In such a way, the strength of the fluorescent signal may be assumed to be proportional to the concentration of the intracellular YUCCA enzyme.

Marchantia has a simple auxin-signalling system containing a single *Aux/IAA* and a single *TIR1/AFB* gene, which are responsible for auxin sensing, and three *ARF* transcription factors, responsible for the regulation of the auxin response^[133,134]. It has been demonstrated that the DII motif, responsible for auxin-sensitive degradation of *Aux/IAA*, can be fused to the coding sequence of a fluorescent protein to create a reporter of relative auxin distribution in the roots of *Arabidopsis*^[206].

In this chapter, two fluorescent reporters, measuring auxin production and concentration by proxy, were designed based on *proMpYUCCA2* and the DII region of the *MpIAA*. Their patterns were studied in order to validate the predictions of the proposed mechanism of auxin-regulated cell proliferation in *Marchantia* gemmae.

6.2 Reporter design

6.2.1 Auxin-biosynthesis reporter

The gene model of *MpYUCCA2*, extracted from the MarpoDB (Chapter 5), was used to design primers for cloning *proMpYUCCA2* from the *Marchantia* genome: two kilobase-pair region upstream of the translation start site was PCR-amplified while adding the flanking regions, required for the Loop assembly^[147] (Figure 6.2 A). The size of the amplicon was verified by gel electrophoresis and cloned into the modified *pUDP2*^[147] level 0 vector (Subsection 2.2.1). The *proMpYUCCA2* part was used to assemble a synthetic fluorescent reporter by fusing it to the coding sequence of cyan fluorescent protein mTurquoise2^[152] (CFP) with an N7 nuclear localisation tag^[154] and nosT-35ST terminator^[147] (Figure 6.2 C, section 2.2, table 2.1). The resultant *proMpYUCCA2:mTurquoise2-N7-nosT-35ST* reporter was assembled into a level 2 *mDYUC2* construct, together with genes encoding for hygromycin resistance, membrane-localised green fluorescent protein eGFP-Lti6b^[156] (GFP), and nuclear-localised yellow fluorescent protein Venus^[153] (YFP), driven by the constitutively-expressed *proMpARF3* (Subsection 2.2.2, table 2.1).

Assuming that the difference in post-transcriptional regulation and degradation rates between *MpYUCCA2* and mTurquoise2 is small, and that the changes in transcriptional activity of *proMpYUCCA2* are slow, the output of the *proMpYUCCA2:mTurquoise2-N7* fluorescent reporter is proportional to the *MpYUCCA2* concentration (Model 6.1 A). Furthermore, since the *MpYUCCA2* enzyme catalyses the final step of the tryptophan-dependent auxin-biosynthesis pathway, the signal of the fluorescent reporter can be used as a proxy for the rate of auxin production (Model 6.1 B).

Model 6.1 Reporter of auxin biosynthesis

A) Proxy of MpYUCCA2 concentration

Dynamics of MpYUCCA2 concentration, y , can be modeled with the following ordinary differential equation:

$$\frac{dy}{dt} = p - d_y y$$

where p is a promoter-dependent production rate and d_y is the MpYUCCA2 degradation rate. The model assumes only the transcriptional control of the gene expression, and does not take into account any post-transcriptional factors. The solution to the above equation is

$$y = y^* + Ae^{-d_y t}$$

where $y^* = p/d_y$ is a steady state concentration and $Ae^{-d_y t}$ is a dynamic term, which decays with the timescale proportional to the degradation rate d_y .

Assuming that the temporal changes in promoter activity are slow compared to the degradation rate, and that there is no transport of MpYUCCA2 between cells, the dynamic term is negligible and the pattern of the MpYUCCA2 concentration is proportional to the pattern of promoter-dependent production rate, i.e.,

$$y \approx y^* = \frac{p}{d_y}$$

By analogy, the pattern of the fluorescent protein concentration (f) driven by the $proMpYUCCA2$ is

$$f = \frac{p}{d_f}$$

where d_f is a rate of fluorescent protein degradation. Assuming that the patterns of the degradation rates d_y and d_f are consistent, i.e., there exists a real constant Γ such that $d_f = \Gamma d_y$, the concentration of the MpYUCCA2 is proportional to the concentration of the fluorescent protein, and hence its fluorescent signal:

$$y = \Gamma f$$

B) Proxy for auxin biosynthesis

As shown in **A**, the signal recorded from the fluorescent protein driven by the $proMpYUCCA2$ can be assumed to be proportional to the concentration of the MpYUCCA2 enzyme, which catalyses the final step in the tryptophan-dependent auxin-biosynthesis pathway:



where TRP is tryptophan, IPA is indole-3-pyruvate, IAA is indole-3-acetic acid or auxin and MpTAA is an amino transferase. Following Michaelis–Menten kinetics, the rate of auxin biosynthesis is proportional to the concentration of the MpYUCCA2 enzyme, and hence the signal of the fluorescent protein:

$$\frac{di}{dt} \sim y \frac{s}{s + K_m}$$

where i is concentration of auxin (IAA), s is concentration of the reaction substrate (IPA) and K_m is a Michaelis constant of the substrate.

The model does not account for variation of the tryptophan and MpTAA in the tissue, nor for the other pathways leading to auxin biosynthesis. Therefore, the recorded signal can only be used as an approximate auxin production rate, which might be in disagreement with the actual rate of auxin production inside a cell.

6.2.2 Auxin-signalling reporter

An auxin-signalling reporter was built based on the previous work of ratiometric auxin sensors in Arabidopsis^[206]. The sensor design relies on a DII domain of an AUX/IAA protein gene. In the presence of auxin, DII binds to TIR1/ABF, promoting degradation of the AUX/IAA protein. By fusing the DII domain to an open reading frame of a fluorescent protein, it is possible to obtain a signal that is sensitive to intracellular auxin concentration.

In order to build an auxin-sensitive reporter, a *DII* level 0 CDS1 part (Section 2.2) was designed based on the sequence of the DII-containing N-terminus sequence upstream of the *ntdTomato* coding sequence from the *pGreenIIIMRPS5A-mDII-ntdTomato/RPS5A-DII-n3Venus* plasmid^[206]. The original Arabidopsis Aux/IAA DII domain was substituted by the sequence of the DII domain of Marchantia Aux/IAA, extracted from MarpoDB (Figure 6.1 A). To allow fusion of the *DII* part to a fluorescent protein, the coding sequence of the *Venus* level 0 part (Section 2.2, table 2.1) was modified by introducing a sequence encoding a poly-A linker at the N-terminus (Figure 6.1 A). In addition, the restriction digestion overhangs were changed to those characteristic for the CDS2 part, as defined by the assembly schema (Section 2.2, figure 2.1).

The resultant *DII* and *VenusCDS2* parts were synthesised using the Thermo Fisher Scientific GeneArt service, and delivered in *pMa* vectors with an ampicillin selection marker (Figure 6.2 B). The parts were then assembled into a fluorescent, nuclear-localised *proMpARF3:DII-VenusCDS2-N7-nosT-35ST* reporter (Figure 6.2 C, section 2.2, table 2.1). The reporter was further assembled into a level 2 *mDDII* construct, together with genes encoding for hygromycin resistance, membrane-localised GFP and nuclear-localised CFP driven by the *proMpARF3* (Subsection 2.2.2).

In the *mDDII* construct, two fluorescent proteins were driven by the same promoter, *proMpARF3*. However, only one was tagged by the DII domain and hence expected to degrade quicker in the presence of auxin. Assuming that the activity of the *proMpARF3* is context-independent, the ratio between the signals from the CFP and YFP can be used as a proxy for auxin concentration (Model 6.2).

6.2.3 Transgenic reporter lines

The *mDYUC2* and *mDDII* constructs were introduced into Marchantia spores (Subsection 2.1.3) using agrobacterium-mediated transformation (Subsection 2.1.4). The spores were grown on a selective medium that contained hygromycin. Therefore, only the sporeling that acquired the hygromycin resistance gene during the transformation carried on normal growth, in contrast to the untransformed sporeling that died due to a lethal concentration of hygromycin in the media. For each line, a dozen surviving transformants with bright green fluorescence in the membrane were selected using a fluorescent stereo microscope (Subsection 2.3.2). The selected sporelings were then rescued by being transferred to a growth medium without the antibiotic and were grown until they started to produce gemma cups (Subsection 2.1.2).

Individual gemmae were picked from the gemma cups of the transgenic Marchantia plants, transferred to a growth chamber (Subsection 2.3.1) and imaged using the laser-scanning confocal microscope for two days with a ~ 1 h time-step (Subsection 2.3.3). Four-channel 3D images were collected for each time frame and reduced to two dimensions using maximal projection along the *z*-axis. The *CFP* channel collected the signal from the nuclei-localised cyan

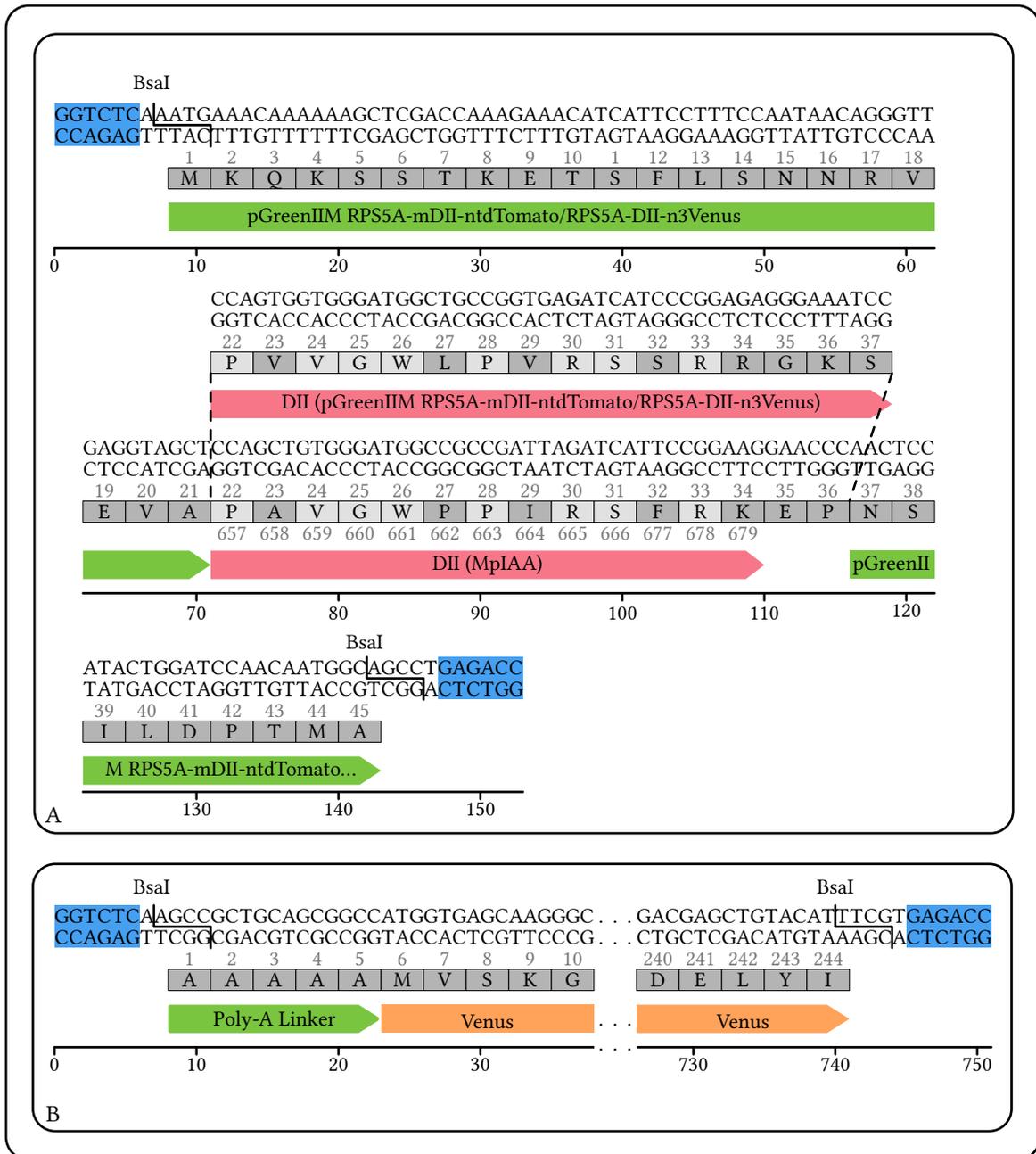


Figure 6.1: **Design of synthetic parts.** **A.** DII tag design. **B.** *mVenusCDS2* design. Amino acids coloured in light grey denote conservation between the DII of the *pGreenIIM RPS5A-mDII-ntdTomato/RPS5A-DII-n3Venus* and DII of *MpIAA*. The dashed lines denote the substitution region. The numbers on top of the amino acid sequence denote position in a synthetic construct; the numbers on the bottom denote the position in the *MpIAA* protein sequence. Blue denotes *BsaI* recognition site.

Model 6.2 Reporter of auxin signalling

A) Reporter design

DII is a protein domain of Aux/IAA responsible for auxin-dependent degradation and has been previously fused to a fluorescent protein to visualise the auxin-distribution pattern in the roots of Arabidopsis^[206]. A degradation rate of a fluorescent protein fused to the DII should be higher in the presence of auxin. Therefore, the steady state concentration of the fluorescent protein should be modulated by the intracellular auxin concentration.

Similar to Model 6.1, dynamics of the concentration of the cyan fluorescent protein (CFP) driven by the *proMpARF3* can be described with the following ordinary differential equation:

$$\frac{dc}{dt} = p - d_c c$$

where c is concentration of the CFP, p is *proMpARF3* activity and d_c is CFP degradation rate. The dynamics of the yellow fluorescent protein (YFP) fused to the same promoter but tagged with the *DII* can be described as

$$\frac{dy}{dt} = p - (d_y + d_a a)y$$

where y is concentration of the YFP, a is the concentration of auxin, d_y is the YFP degradation rate and d_a is an auxin-dependent degradation coefficient. The steady states of the fluorescent protein concentrations c^* and y^* are then

$$\begin{aligned} c^* &= p/d_c \\ y^* &= p/(d_y + d_a a) \end{aligned}$$

B) Estimating auxin concentration

Assuming that the system transitions through a steady state, and that the recorded signal is proportional to the concentration of the fluorescent proteins, the ratio between the signals is proportional to the local concentration of auxin:

$$s_c/s_y = g_c/g_y(d_y/d_c + ad_a/d_c)$$

where $s_c = g_c c_0$ and $s_y = g_y y_0$ are the signals of the CFP and the YFP channels amplified by a gain g_c and g_y respectively. Therefore, the ratio (R) of the two signals can be used as a proxy for auxin concentration:

$$a = kR - b$$

where

$$\begin{aligned} k &= g_y/g_c \cdot d_c/d_a \\ b &= d_y/d_a \\ R &= s_c/s_y \end{aligned}$$

The model assumes that the biological factors influencing the coefficients k and b are spatially homogeneous across the tissue.

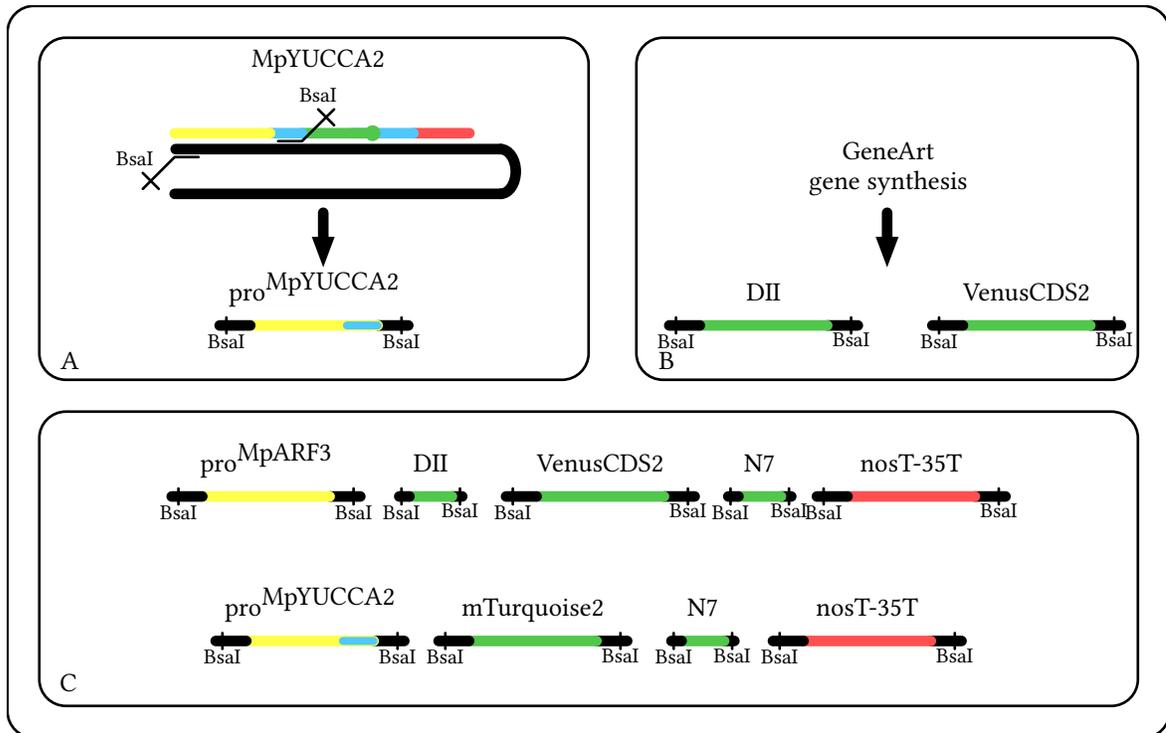


Figure 6.2: **Design of auxin signalling and biosynthesis reporters.** Panel A depicts PCR-amplification strategy of the *proMpYUCCA2* based on the *MpYUCCA2* gene model from MarpoDB. The thick black line denotes Marchantia genome, yellow bar annotates the promoter region, blue bar - untranslated regions, green bar - the coding sequence, red bar - terminator. The thin black lines denote PCR primers, containing *BsaI* restriction enzyme site and overhangs for the Loop assembly. Panel B depicts *DII* and *VenusCDS2* DNA parts synthesised via GenArt service. Panel C shows the level 0 DNA parts in the order they been assembled into level 1 constructs; auxin biosynthesis reporter on the top, auxin signalling reporter in the bottom.

Model 6.3 Effect of auxin on auxin sensor readings

In order to confirm that the auxin reporter is sensitive to auxin, three gemmae of a *mDDII* Marchantia line were grown on substrate supplemented with auxin while the signal of the reporter was captured by the means of fluorescent microscopy. At the same time, a control group of three gemmae of the same line was grown under the same conditions, apart from auxin supplementation.

The fluorescent signal from individual nuclei was collected as described in Algorithm IP.3 and the output of the reporter was calculated as a ratio between cyan and yellow channels (Model 6.2).

Since the global effect of auxin on the sensor output was of interest, the obtained data was averaged for each timepoint and sample. In addition, the signal was normalised by subtracting the starting ($t = 0$) signal for each sample independently. This normalisation allows to remove a bias of intra-sample variability. Finally, a following linear-regression model was fitted into the data:

$$R = b + k_1t + k_2a$$

where R is a ratio between readings of a cyan and yellow channel, supposedly capturing auxin response, t is time after treatment, a is a categorical variable, which takes value of 1 for the treated samples and 0 for the control samples.

Coefficients of the regression model and their p-values are summarised below:

Coefficient	Value	p-value
b	-31.8	-
k_1	1.33	$4.3 \cdot 10^{-29}$
k_2	27.8	$1.1 \cdot 10^{-15}$

The results suggest that application of auxin has a significant effect on the reporter output. The r^2 for the fitted model was 0.78.

fluorescent protein; the *YFP* channel collected the signal from the nuclei-localised yellow fluorescent protein; the *membrane* channel collected the signal from the membrane-localised GFP and the *chlorophyll* channel collected the chlorophyll autofluorescence signal (Table 2.4).

The pattern of the CFP expression driven by the *proMpYUCCA2* in *mDYUC2* had maxima in the apical regions and the rhizoid precursor cells of the central region (Figure 6.3). The pattern of the CFP/YFP ratio in the *mDDII* line had maxima in the proximal apical region and the central region (Figure 6.4).

In order to verify that the auxin-signalling reporter was responsive to auxin, the gemmae were placed in a growth chamber filled with $100 \mu\text{g/L}$ NAA and imaged for 46 hours with 1 h resolution. The signal of the reporter in the auxin-treated gemmae ($n = 3$) increased significantly compared to that of the untreated gemmae (Figure 6.5, Model 6.3).

6.3 Average patterns of auxin biosynthesis and signalling

In order to study the relationship between the tissue expansion patterns recorded in chapter 3, and the patterns of the auxin biosynthesis and signalling, the data recorded from both studies should have been represented in the same form. The difference between the recordings of the tissue expansion and those of the nuclear signal is that the former was sampled for each pixel of an image, while the latter contained information only within the nuclear regions. In

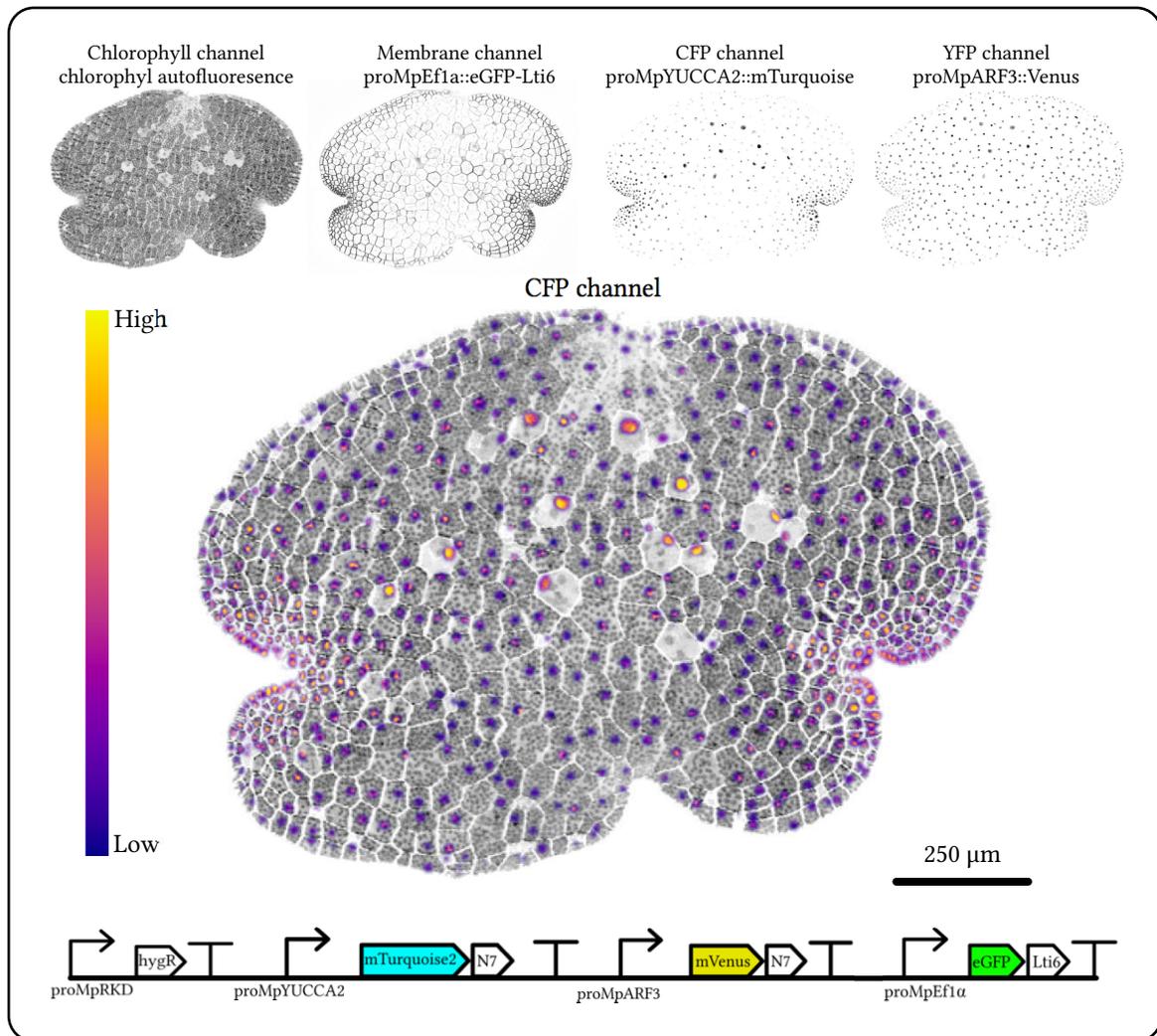


Figure 6.3: **Auxin-biosynthesis reporter.** The images at the top of the figure show inverted signal from the individual channels within a single time frame. The central image shows a false-colour signal of the CFP channel overlaid on top of the sum of the membrane channel and inverted chlorophyll channel. The CFP channel captured the signal of *mTurquoise2*, driven by the *proMpYUCCA2*, and was used as a proxy for the auxin biosynthesis rate.

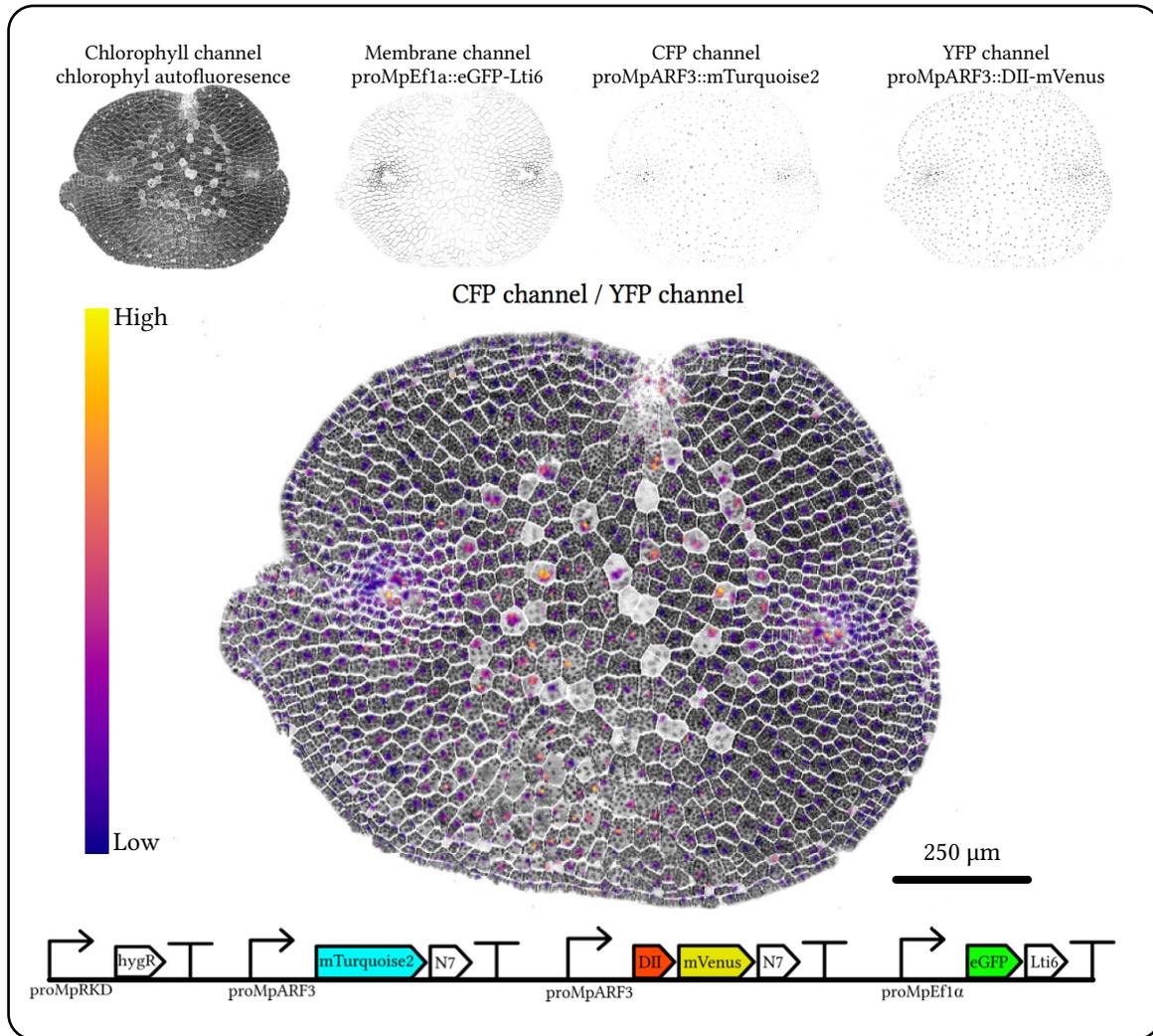


Figure 6.4: **Auxin-signalling reporter.** The images at the top of the figure show the inverted signal from individual channels within a single time frame. The central image shows a false-colour signal of a ratio between the CFP and YFP channels overlaid on top of the sum of the membrane channel and inverted chlorophyll channel. The YFP channel captures the signal from *DII-mVenus* driven by the *proMpARF3*, while the CFP channel captures the signal of *mTurquoise2* driven by the same promoter. The CFP/YFP ratio was used as a proxy for auxin signalling.

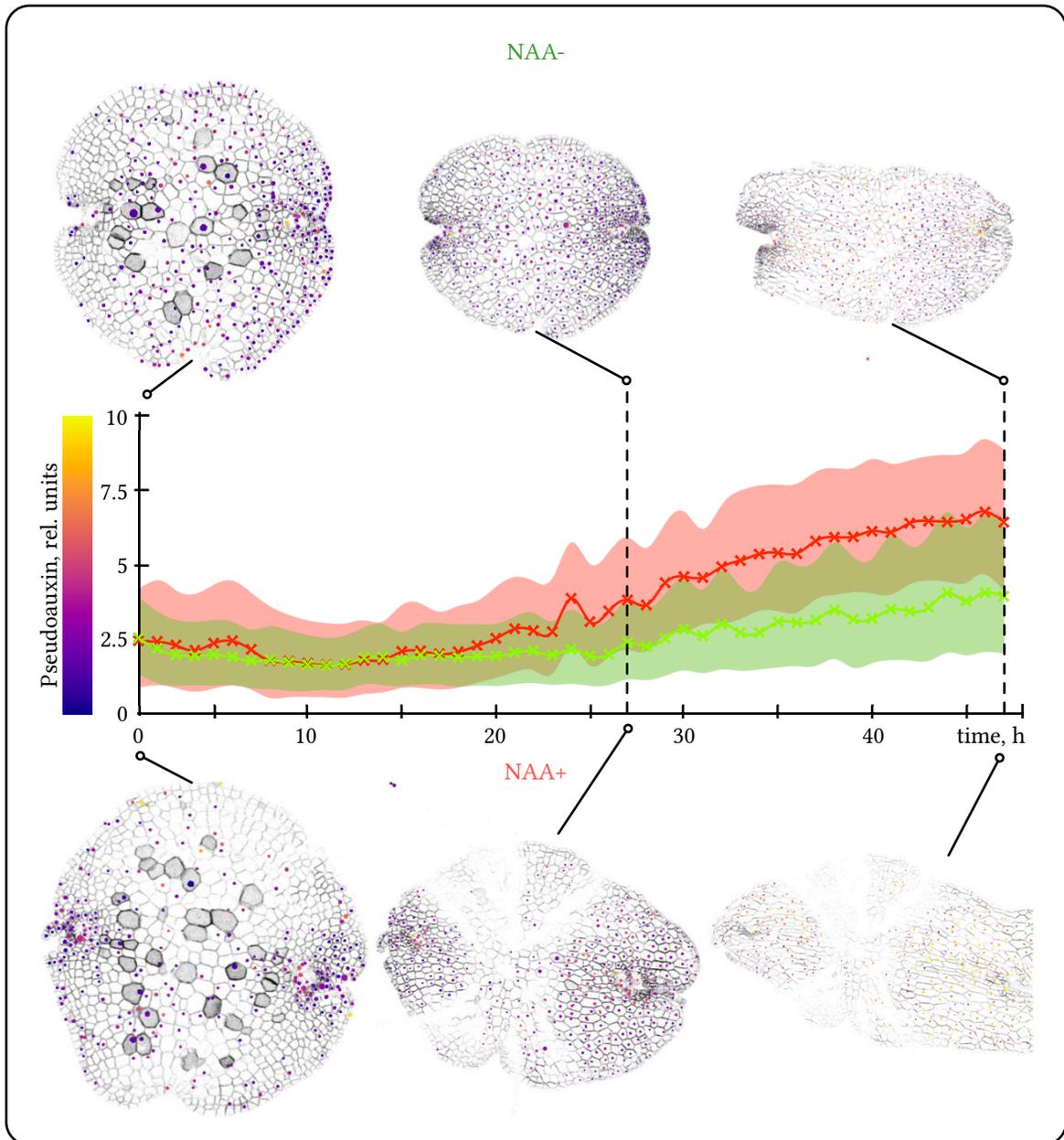


Figure 6.5: **Induction of the auxin-signalling reporter.** The traces on the graph denote the mean signal of the auxin-signalling reporter with added synthetic auxin (red) and control (green). The filled regions denote the signal range in 10% – 90% quantile. The images show the signal of the reporter at 0 h, 22 h and 47 h after removal from the gemma cups.

order to convert the microscopy images capturing the cell-nuclei signal into a discrete sample of observations, a nuclear segmentation algorithm was applied (Algorithm IP.3).

Due to technical constraints, the signal from the measurement channels (*CFP* in *mDYUC2* and *YFP* in *mDDII*) can fall below the sensitivity range of the microscope detector. In this case, it is impossible to distinguish a nucleus with a low signal from the absence of a nucleus in the corresponding location. In order to capture the signal from all nuclei independent of whether or not they have a detectable signal, a nuclear detection algorithm was applied to the reference channel produced by *proMpARF3:Venus* in *mDYUC2* or *proMpARF3:mTurquoise2*. The broad expression pattern of *proMpARF3* ensured that the locations of most of the nuclei in the picture were detected. By sampling the measurement channel in the detected nuclei locations, it was possible to record data points even in cases where the measurement signal was below the detection limit (Subsection 2.4.2:Algorithm IP.3, Figure 6.6).

As a result of nuclear segmentation, each nucleus was represented as a single datapoint containing information about the time since removal from the parent gemma cup, sample index, position relative to the image frame and average-per-nucleus intensity of the signals in the *YFP* and *CFP* channels. As discussed in the subsections 6.2.1 and 6.2.2, the intensity of the *CFP* channel in *mDYUC2* was used as a proxy for auxin biosynthesis, while the ratio of the *CFP* to *YFP* channels in *mDDII* was used as a proxy for auxin signalling.

The normalisation routine, previously applied to spatiotemporal data of relative tissue expansion rates in section 3.4, was applied to the collected nuclear expression data from $n = 6$ gemmae samples for both *mDYUC2* and *mDDII* lines. In short, the coordinates of the extracted nuclear regions were transformed relative to the position of the closest apical notch and normalised by the gemma size (Subsection 2.4.2: IP.5). For this purpose, the position of the apical notches was tracked using an image-registration algorithm (Subsection 2.4.2: IP.2, section 3.4, model 3.2).

The data were then averaged ($n = 6$) across all samples ($n = 6$) and time frames ($n = 35$) of the same line, *mDYUC2* or *mDDII* (Section 3.4, Subsection 2.4.1) and minmax normalised, i.e.:

$$y_i = \frac{x_i - \min X}{\max X - \min X}, x_i \in X$$

where X is a set of all observations of a particular property, and y_i is a normalised value. For example, a normalised auxin signalling value of 1 corresponds to the maximal recorded value, while value 0 corresponds to the minimal recorded value.

By applying this normalisation, the tissue expansion rate pattern and patterns of auxin signalling and biosynthesis may be compared despite being recorded from different gemma samples (Figure 6.7).

According to the collected pattern, auxin signalling was predominant in the central region with $r_n > 0.25$, where the r_n value corresponds to the distance to the closest apical notch, normalised by the distance between the two apical notches. It was in negative correlation with the tissue expansion rate (Figure 6.8 Pearson coefficient -0.85 , p-value 1.21×10^{-22}). In addition, the reporter suggested elevated auxin signalling in the proximal apical region with $r_n < 0.1$.

The auxin biosynthesis pattern, on the contrary, was prevalent in the apical regions with $r_n < 0.25$, and was also in negative correlation with auxin signalling rates (Pearson coefficient -0.69 , p-value 3.24×10^{-12}). Sample

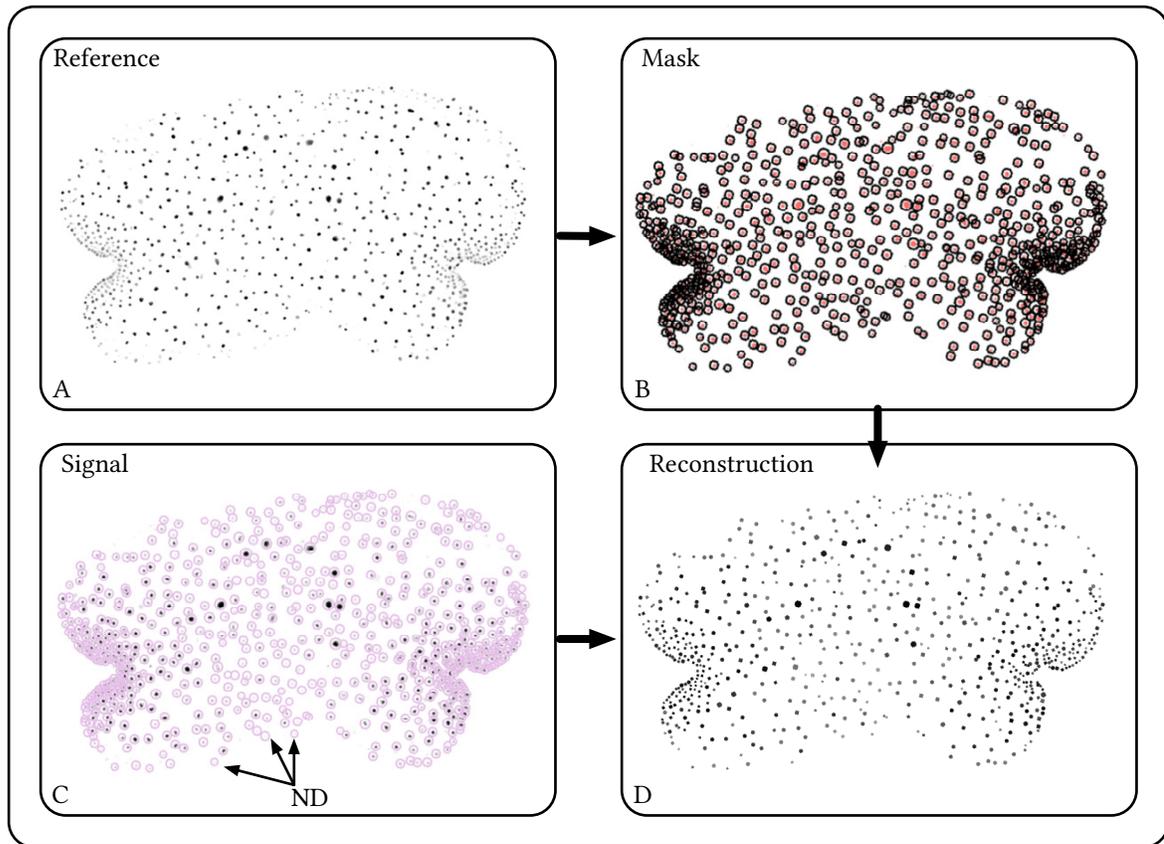


Figure 6.6: **Extraction of nuclear signal.** **A.** An inverted signal from the reference channel (*proMpARF3 :: mTurquoise2 – N7*) in *mDYUC2 gemma*. **B.** Result of the nuclei detection algorithm. The regions containing the nuclei are labeled in red. The centres of the black circles denote locations of the nuclei centroids, while their radii are equal to twice the estimated nuclei radii. **C.** An inverted signal of the measurement channel (black) and locations of the detected nuclei (pink). ND denotes examples of the nuclei that would not have been detected if measurement signal was used for the nuclei detection. **D.** Example of the reconstructed signal, where each detected nucleus is represented by a circle with the estimated radius and colour representing the corresponding signal from the measurement channel.

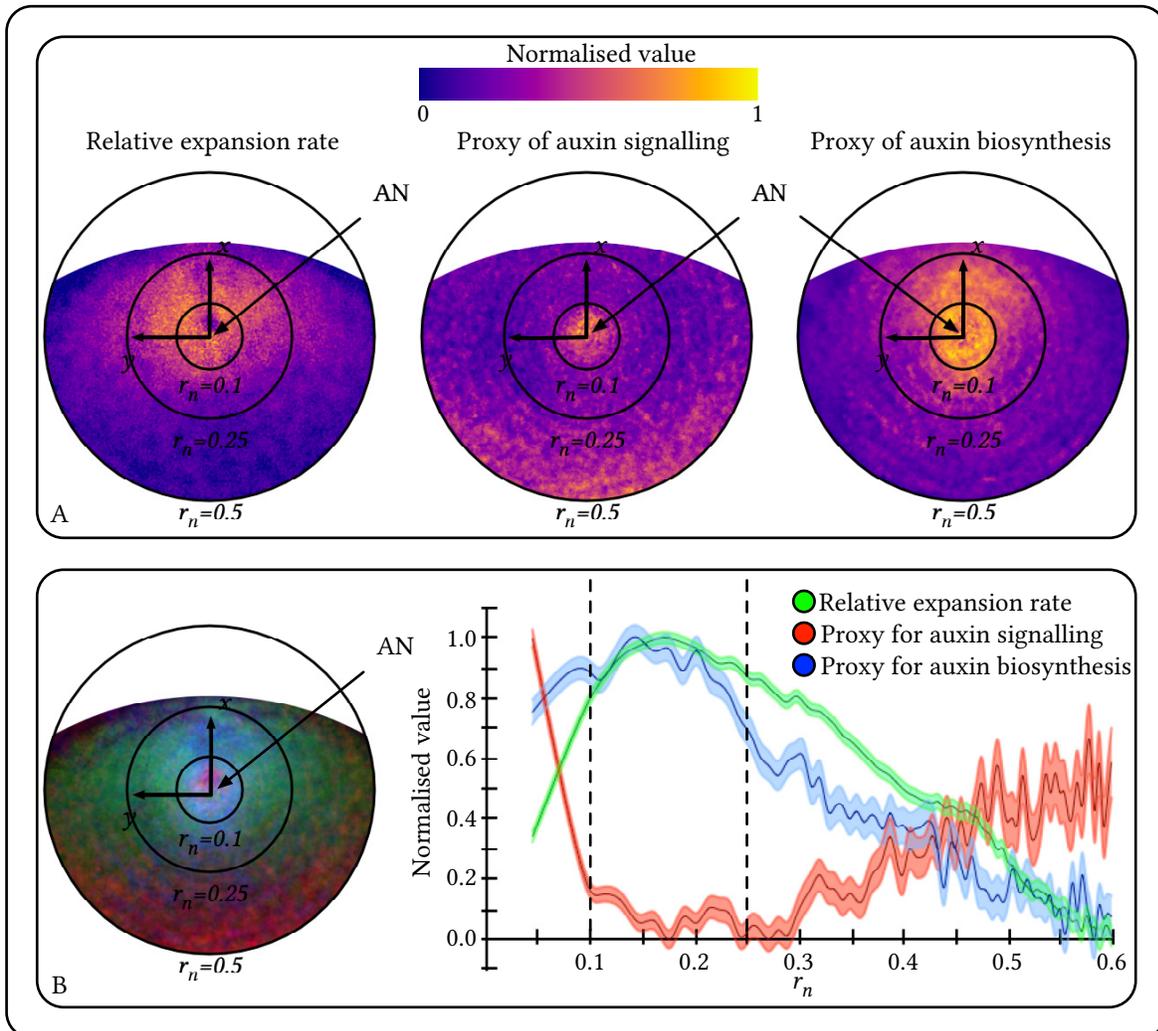


Figure 6.7: **Average patterns of estimated auxin biosynthesis and signalling.** **A.** Visualisation of the average normalised patterns in relation to the closest apical notch (AN). The r_n values denote distance from the closest apical notch, normalised by the distance between the two apical notches of a gemma. Signal is normalised by the highest value recorded. **B.** Overlay of the normalised patterns, highlighting the relationship between the relative tissue expansion rates, rates of auxin biosynthesis and rates of auxin signalling. The traces of the graph denote the mean values. The shaded regions denote 95% confidence interval of the mean estimation, i.e. $\text{mean} \pm 1.96 \text{ SE}$, where SE is a standard error. The vertical dashed lines correspond to the borders of the proximal apical region ($r_n < 0.1$) and apical region proper ($r_n < 0.25$).

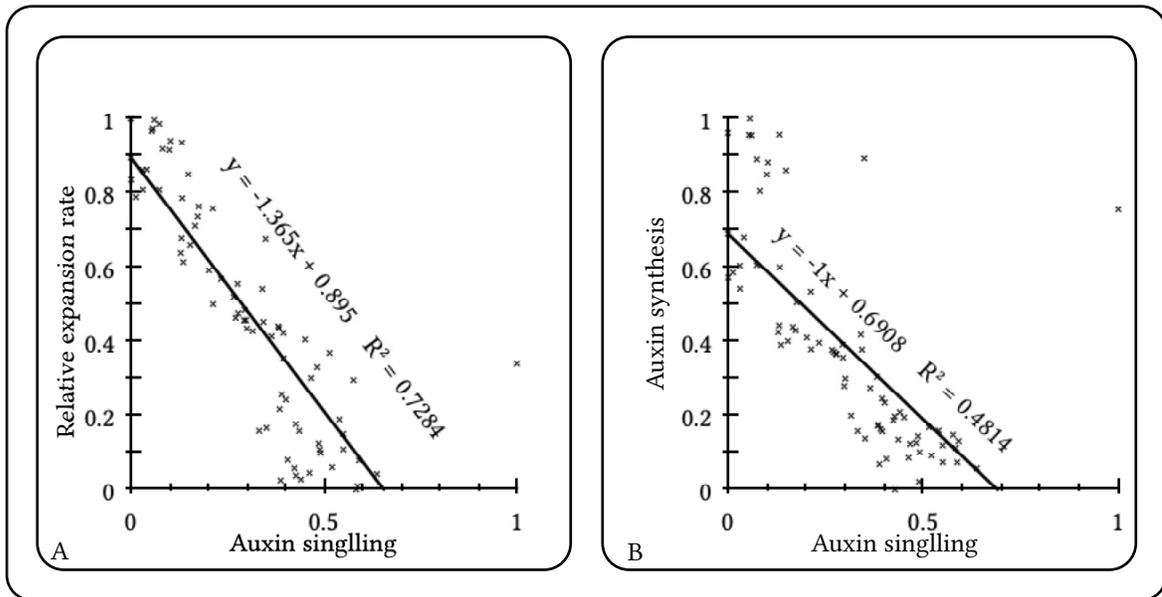


Figure 6.8: **Correlation between growth parameters.** Auxin signalling versus auxin synthesis in A, auxin signalling versus relative tissue expansion rate in B.

variability of the signal is captured in figure 6.9 for auxin biosynthesis reporter and figure 6.11 for auxin signalling reporter. The time variability if the signal is captured in figure 6.10 for auxin biosynthesis reporter and figure 6.12 for auxin signalling reporter.

6.4 Reestablishment of auxin signalling and biosynthesis in the isolated central region

In order to study the signal of the reporters under conditions of the surgical assay (Chapter 4), the apical regions were excised from the *mDYUC2* and *mDDII* gemmae ($n = 2$), placed into a growth chamber and imaged for two days. The readings of the auxin-signalling reporter dropped during the first stage of the regeneration (day 0 – day 1), coinciding with active cell proliferation (Figure 6.13). At the second stage of the regeneration, the signal recovered in most of the cells to levels comparable to those at day 0 after the surgery, but remained significantly lower in the apical precursor regions (5.13 rel.u vs 5.78 rel.u $p\text{-value} < 0.01$). The signal from the auxin biosynthesis reporter, on the contrary, increased constitutively during the first stage of the regeneration, while localising to the apical precursor regions during the second stage of the regeneration (Figure 6.14).

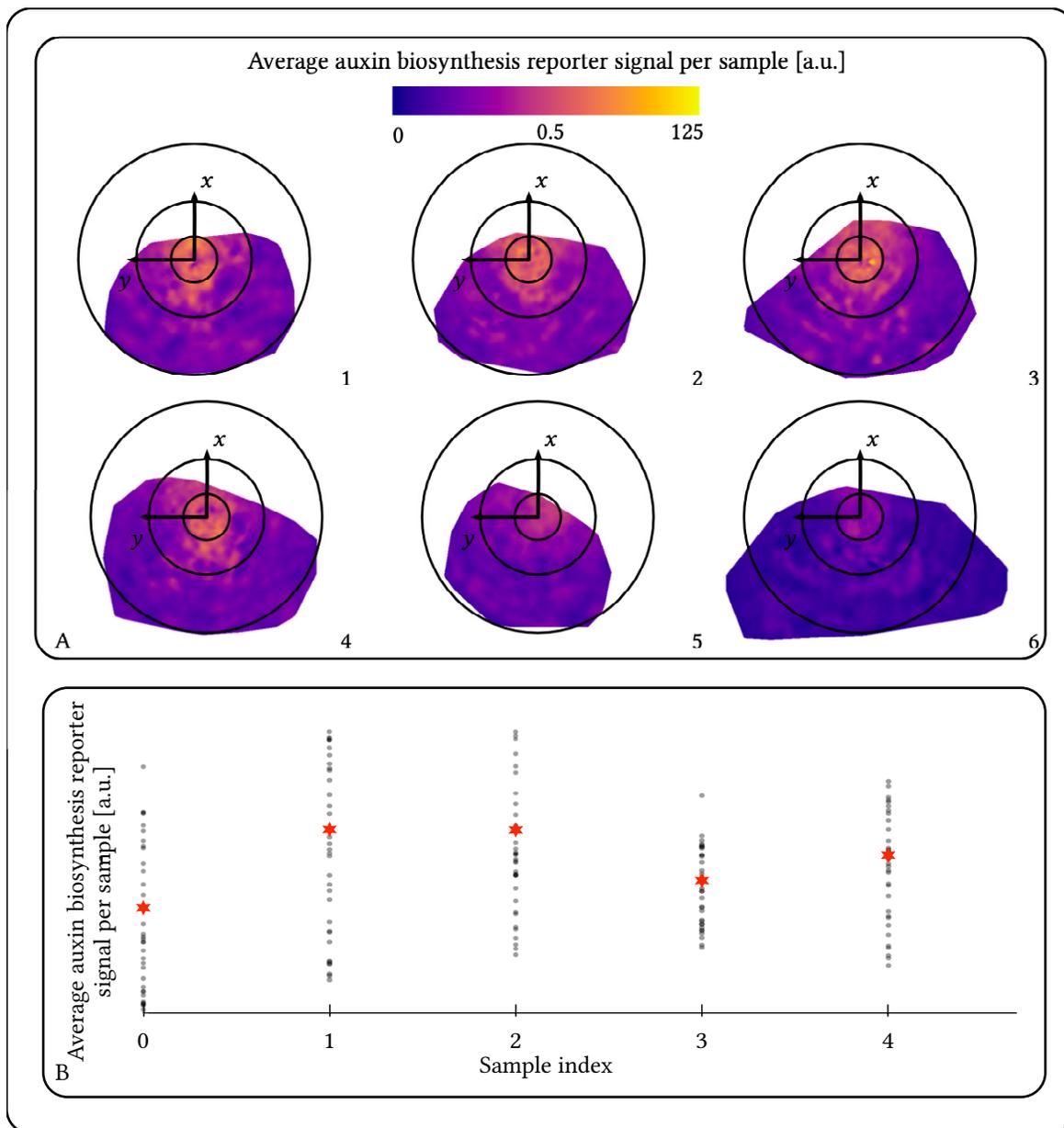


Figure 6.9: **Average auxin biosynthesis reporter signal per sample.** **A.** Normalised patterns of auxin biosynthesis signal averaged over all 35 time frames for each of the six gemma samples. The circles represent $r_n = 0.1, r_n = 0.25$ and $r_n = 0.5$. Each diagram represents a cumulative data from both of the gemma notches. The numbers in the bottom right corner represent gemma index. **B.** Average signal for each gemma. Grey circles represent average signal per time frame, red stars denote the average signal across all time frames.

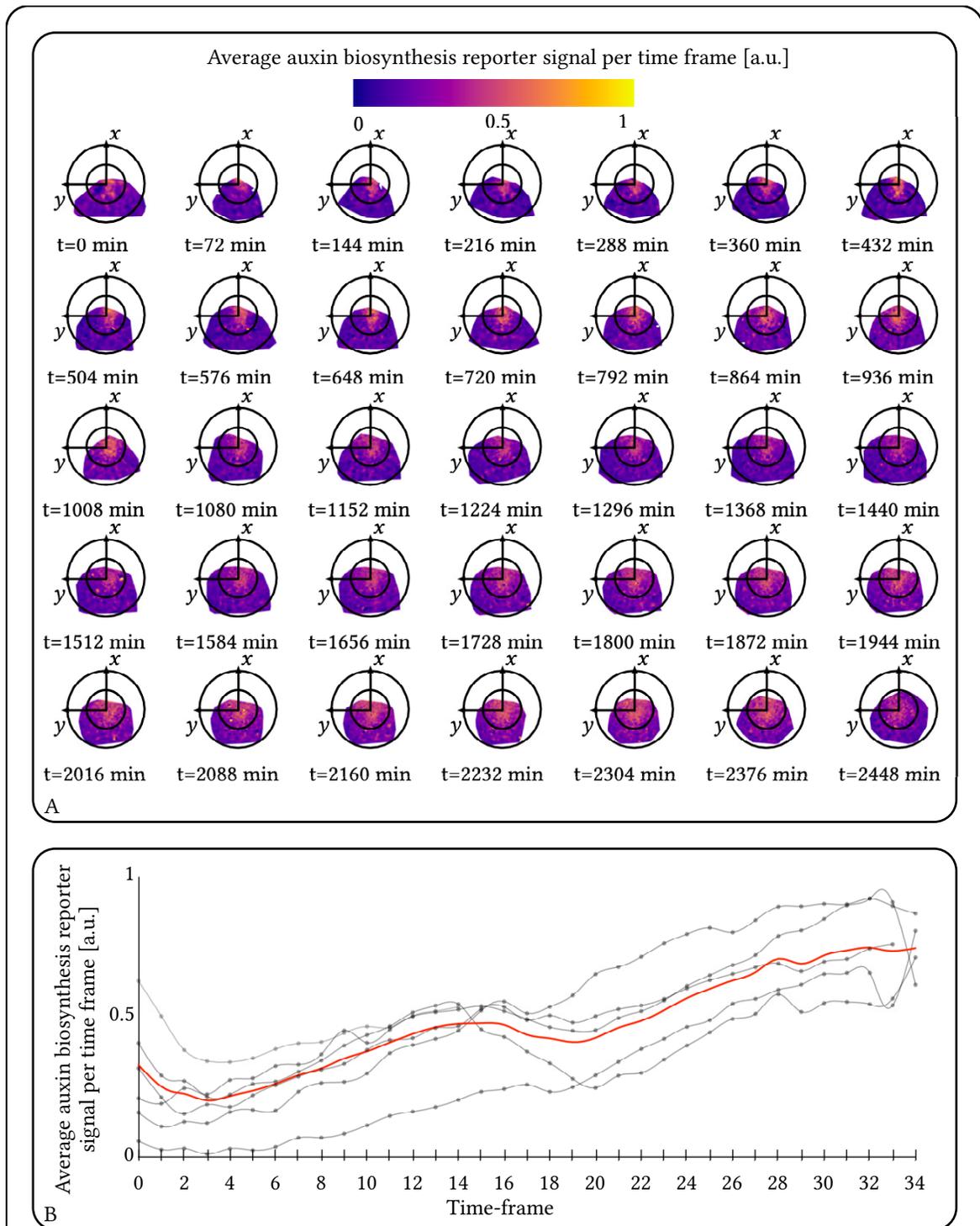


Figure 6.10: **Average auxin biosynthesis reporter signal per time frame.** **A.** Normalised patterns of auxin biosynthesis reporter signal averaged over all gemma samples for each of the time frame. The circles represent $r_n = 0.25$ and $r_n = 0.5$. Each diagram represents a cumulative data from both of the gemma notches. **B.** Average signal for each time frame. Grey traces represent signals for each of the individual gemma samples, red trace is the average across all samples.

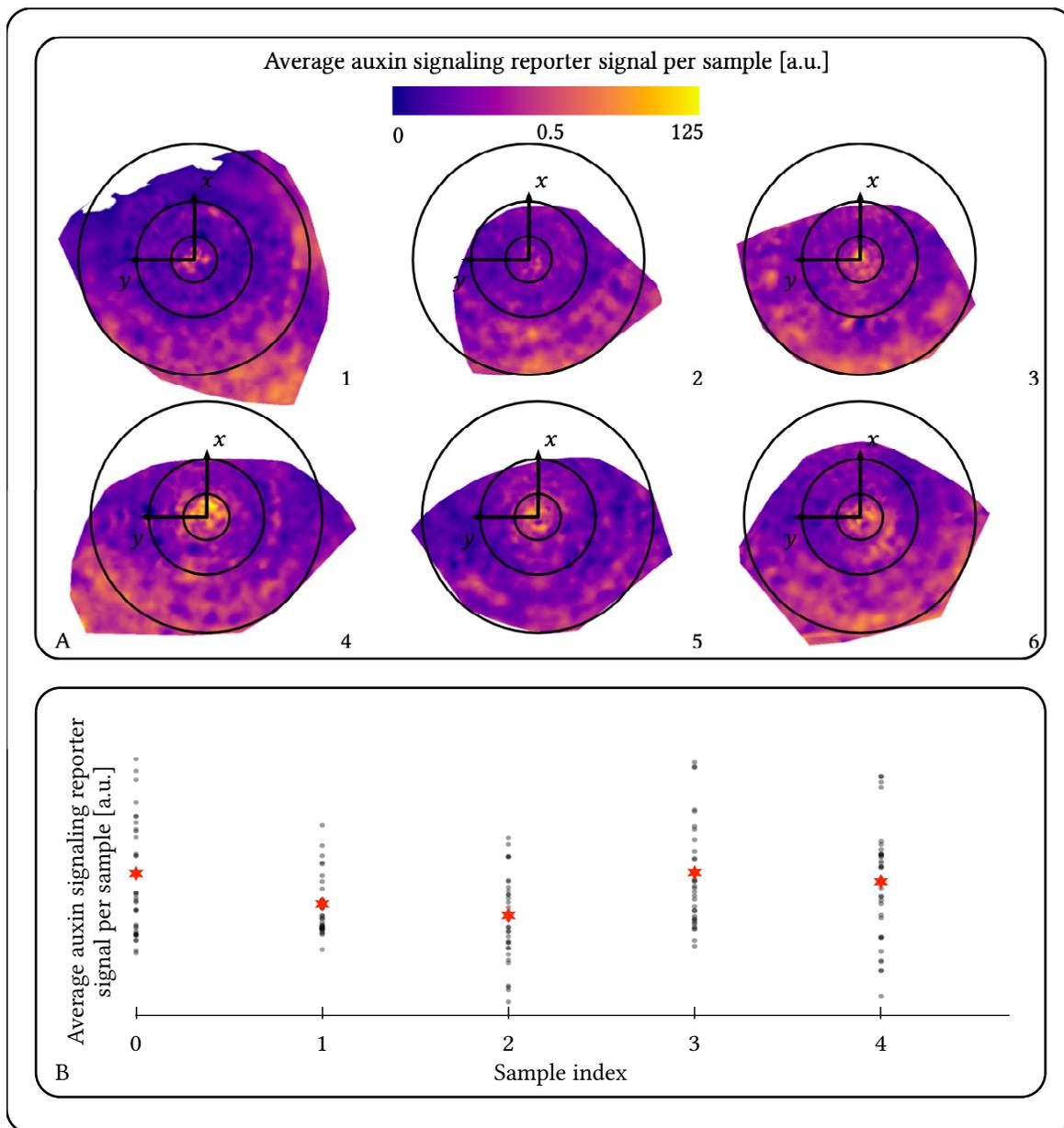


Figure 6.11: **Average auxin signalling reporter signal per sample.** **A.** Normalised patterns of auxin signalling signal averaged over all 35 time frames for each of the six gemma samples. The circles represent $r_n = 0.1, r_n = 0.25$ and $r_n = 0.5$. Each diagram represents a cumulative data from both of the gemma notches. The numbers in the bottom right corner represent gemma index. **B.** Average signal for each gemma. Grey circles represent average signal per time frame, red stars denote the average signal across all time frames.

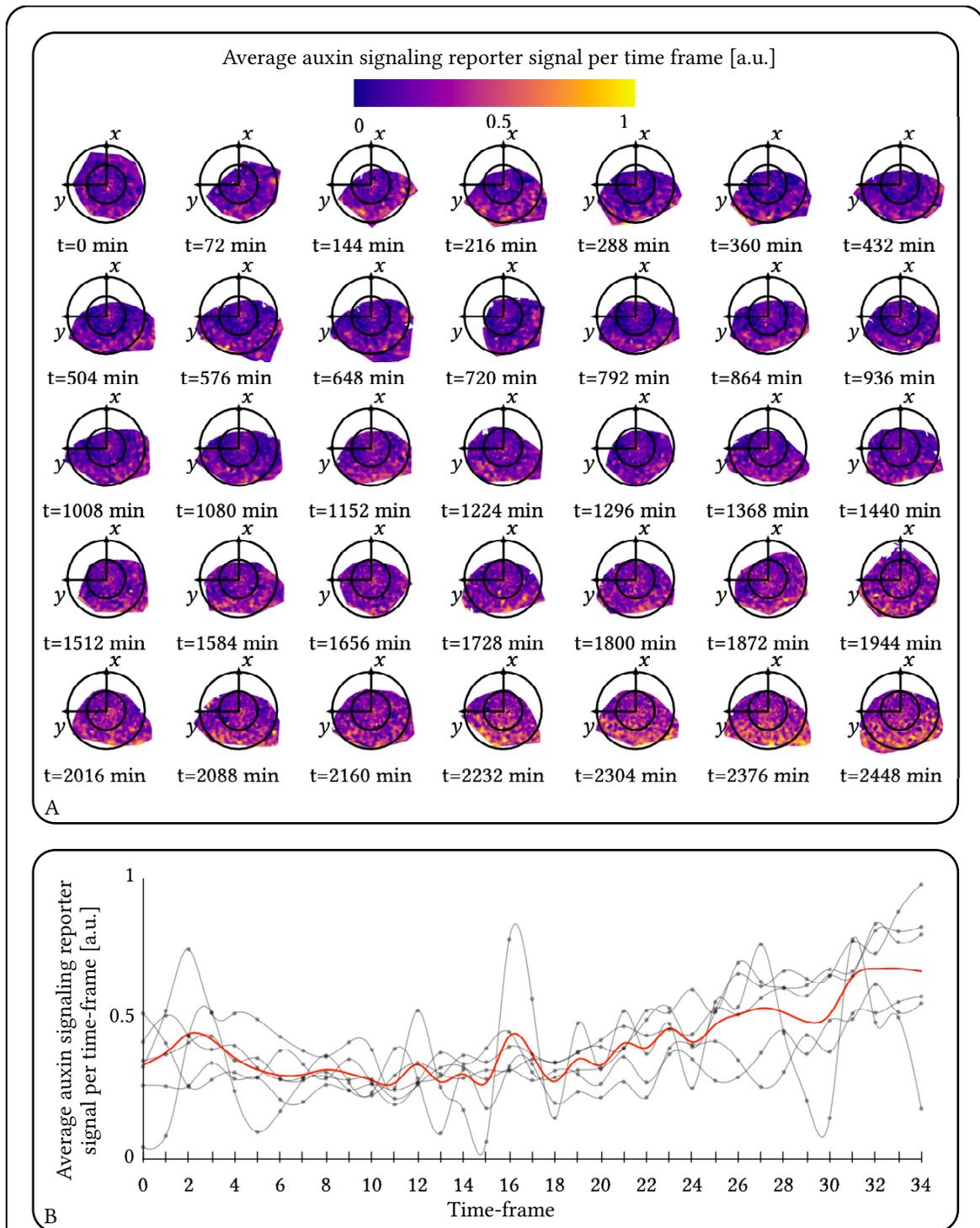


Figure 6.12: **Average auxin signalling reporter signal per time frame.** **A.** Normalised patterns of auxin signalling reporter signal averaged over all gemma samples for each of the time frame. The circles represent $r_n = 0.25$ and $r_n = 0.5$. Each diagram represents a cumulative data from both of the gemma notches. **B.** Average signal for each time frame. Grey traces represent signals for each of the individual gemma samples, red trace is the average across all samples.

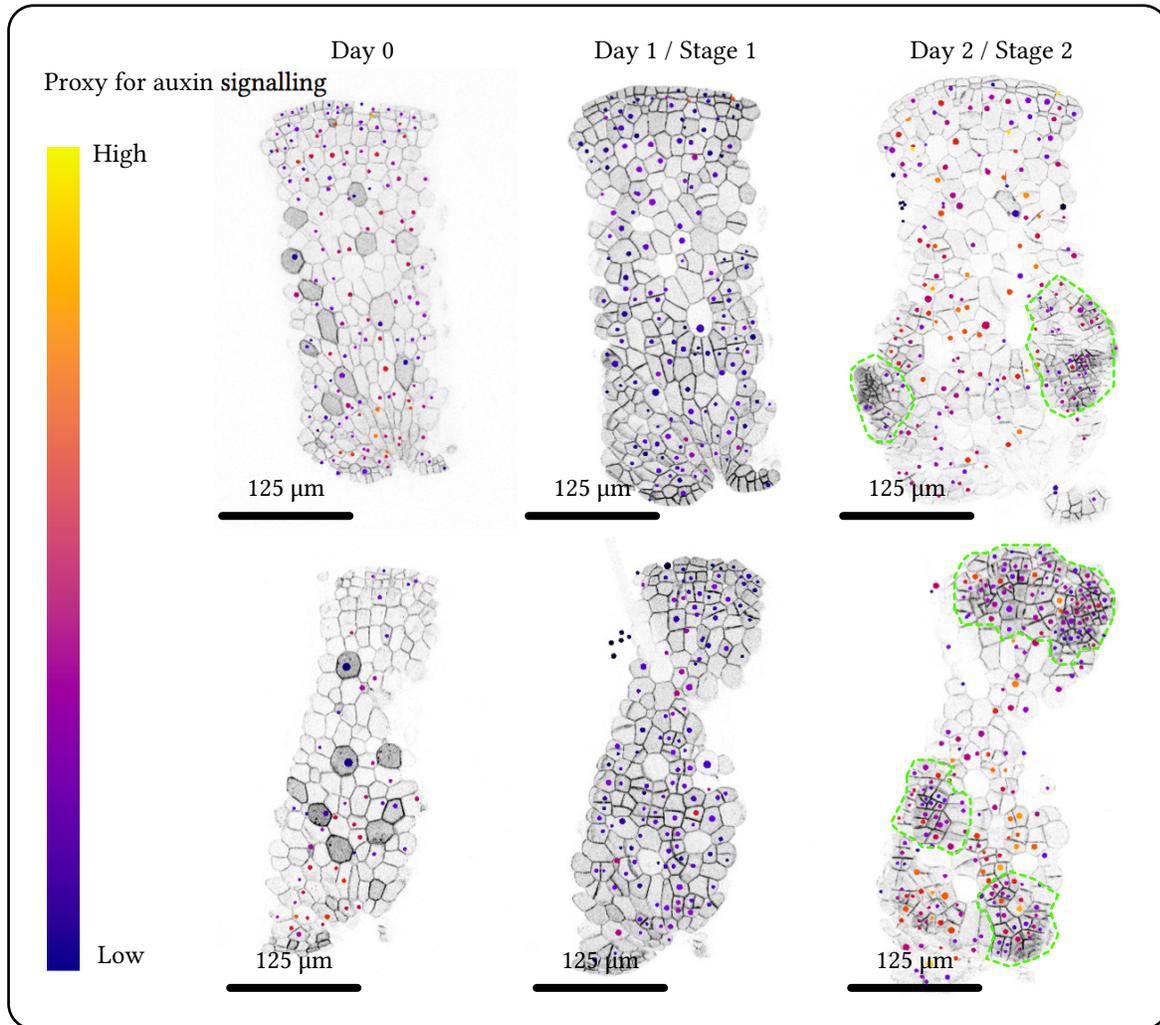


Figure 6.13: **Reestablishment of the auxin-signalling pattern.** Reconstructed nuclear signal of the auxin-signalling reporter overlaid on the inverted signal from the membrane channel of the two excised central regions from *mDDII*. The dashed green lines denote the borders of the apical precursor regions.

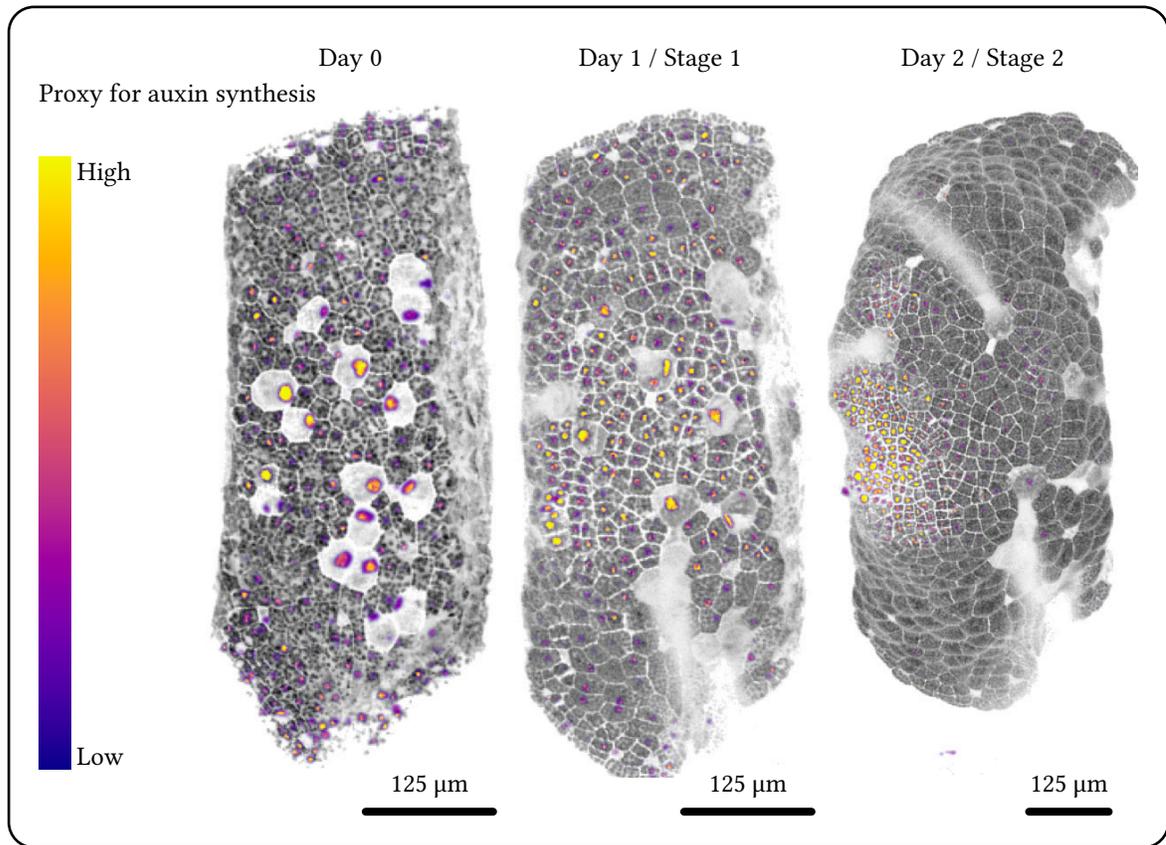


Figure 6.14: **Reestablishment of the auxin-biosynthesis pattern.** Pseudo-colour signal from the CFP channel of the *mDYUC2* line overlaid on the sum of the membrane channel and the inverted mask channel of the excised central regions.

6.5 Continuous model of auxin transport

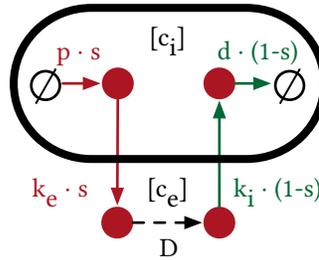
According to the collected patterns of auxin reporters, cells in the apical regions were estimated to have high rates of auxin production, while cells of the central region were estimated to have high rates of auxin signalling. In other words, both auxin biosynthesis and tissue expansion rates were in negative correlation to auxin signalling. The negative correlation between auxin signalling and auxin biosynthesis were also observed in the regenerating cells of the isolated central regions.

These observations support the transport model established in chapter 4: high levels of auxin production in the apical regions, predicted by the model, were confirmed by the pattern of the auxin biosynthesis reporter, while the predicted accumulation of auxin in the central region was supported by the pattern of the auxin-signalling reporter. In addition, the model predicted that the cell proliferation in the isolated central regions was caused by the disturbed auxin influx and resultant drop of auxin concentration. This prediction was supported by the drop in intensity of the auxin-signalling reporter in the isolated central regions one day after the surgery.

In order to visualise the dynamics of the proposed auxin transport mechanism, a mathematical source–sink model was developed based on the results of chapters 3, 4 and 6 (Model 6.4). The model was composed of a pair of reaction–diffusion equations describing the transport mechanism of auxin in the tissue. As in the source–sink model in chapter 4, auxin existed in two phases: intracellular, with concentration c_i , and extracellular, with concentration c_e . The production and degradation of auxin were allowed only for the intracellular phase, while its diffusion only for the extracellular phase. The conversion between the phases was allowed to happen through import and export processes, mimicking the action of the auxin influx and efflux transporters^[207,208]. Following the assumptions of the proposed mechanism and observations of the gemmae development, the parameters of the model depended on the proximity to the apical notches. In order to establish this dependence, a concept of a tissue *state* was introduced. It took values from 0 to 1, where 0 denoted ideal sink tissue, while 1 ideal source. The notch-centred gaussian shape of the state function was chosen to represent the gradual transition from the apical region (source) to the central region (sink), in agreement with the readings of the auxin-biosynthesis reporter (Figure 6.15A, figure 6.7 A). The parameters of the model were proportional to the value of the state function, with a positive proportionality constant for production and export rates, and a negative proportionality constant for import and degradation rates (Figure 6.15B). Therefore, the apical regions were predominantly producing and exporting auxin, while the central region was predominantly importing and degrading it.

The reaction–diffusion equations were solved on a three-dimensional mesh, which resembles the shape of a gemma. The parameters of the model were chosen semi-arbitrarily: auxin degradation rate (d) was chosen to be much smaller than the auxin export rate (k_e) so that the auxin accumulation happened in the sink, based on the analysis of the source–sink model in chapter 4. The auxin import rate (k_i) was chosen to be the same as the auxin export rate. The auxin diffusion and production rates were chosen empirically to ensure stable simulation under the chosen time-step. The simulation was run for 400 arbitrary time units, starting with a homogeneous initial condition $c_i = c_e = 0$ (Figure 6.15 C). In the first 200 time units, the intracellular auxin concentration profile was established, transitioning from intracellular accumulation in the apical regions at $t = 70$, through equilibration between the apical

Model 6.4 Continuous model of auxin transport

**Description**

The model describes the source–sink transport of auxin in plant tissue. The tissue is represented as a cellular material, where intracellular and extracellular spaces are separated by a membrane that constrains passive transport of auxin from one cell to another. Therefore in the model, auxin exists in two phases: intracellular, with concentration $c_i(\mathbf{x})$, and extracellular, with concentration $c_e(\mathbf{x})$. Auxin is produced and degraded only in the intracellular space, while its transport by diffusion only happens in the extracellular space. Auxin can transition from the intracellular space to the extracellular space, and vice versa, through the export and import processes.

The source–sink relationship is established by *state function* $s(\mathbf{x})$ with a range $[0, 1]$, where 0 corresponds to a perfect *sink*, while 1 to a perfect *source*. Production and export of auxin are set to be predominant in the source tissue with the rates equal to ps and $k_e s$, where p and k_e are maximal production and export rates. On the contrary, import and degradation are set to be predominant in the sink tissue with rates $k_i(1-s)$ and $d(1-s)$, where k_i and d are maximal import and degradation rates. The complete model is captured by the following set of equations:

$$\begin{cases} \frac{dc_i(\mathbf{x})}{dt} = ps(\mathbf{x}) - s(\mathbf{x})k_e c_i(\mathbf{x}) + (1-s(\mathbf{x}))k_i c_e(\mathbf{x}) - (1-s(\mathbf{x}))dc_i(\mathbf{x}) \\ \frac{dc_e(\mathbf{x})}{dt} = D\frac{d^2 c_e(\mathbf{x})}{d\mathbf{x}^2} c_e + s(\mathbf{x})k_e c_i(\mathbf{x}) - (1-s(\mathbf{x}))k_i c_e(\mathbf{x}) \end{cases}$$

where D is a diffusion coefficient.

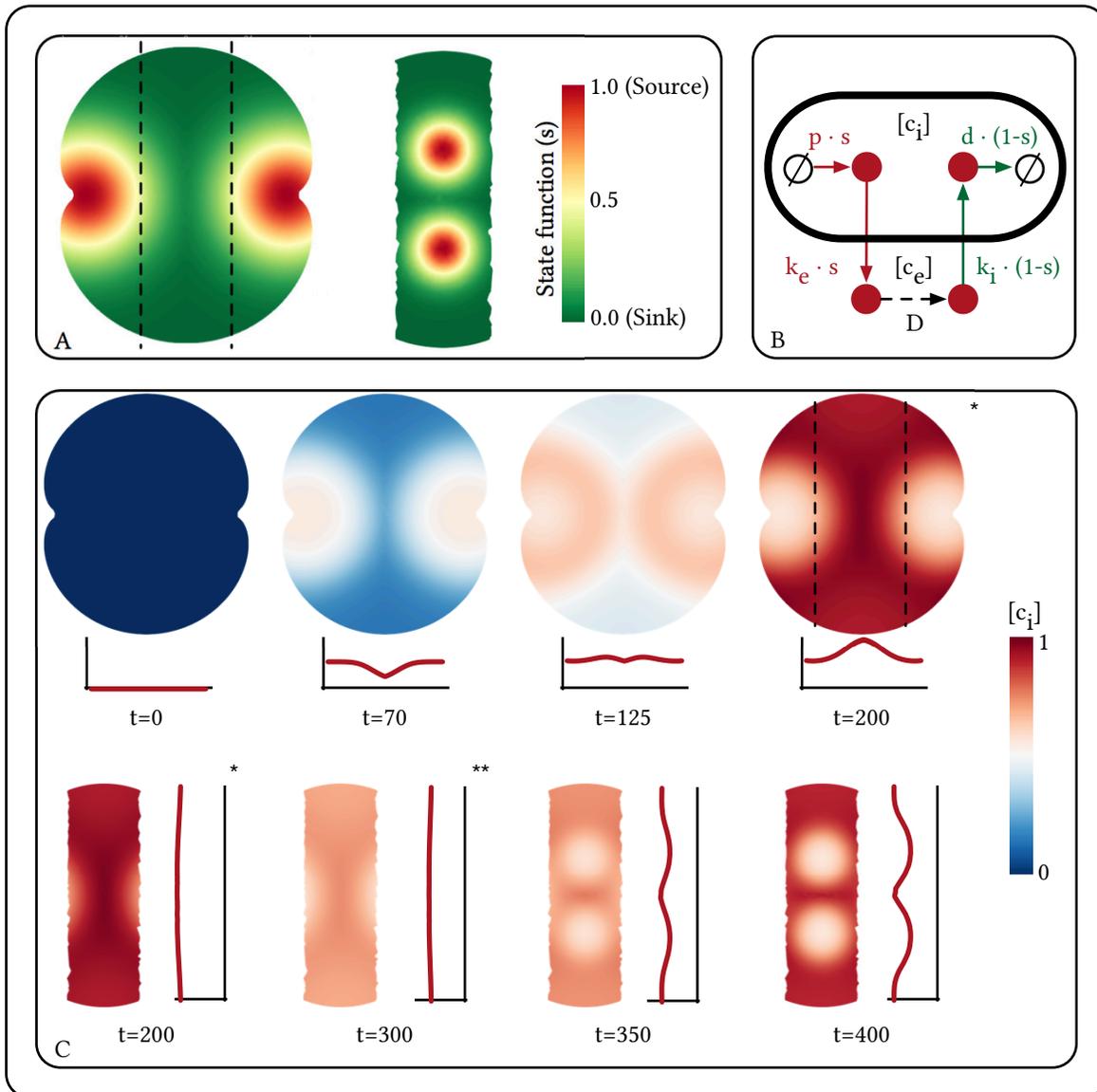


Figure 6.15: **Simulation of the auxin transport model.** **A.** The state function $s(x)$, marking the apical fate, was prescribed on a three-dimensional mesh that resembled the shape of a young gemma. For simulating the surgical assay, the domain of the mesh marked by the dashed lines was extracted, and a state function that mimicked the apical region precursors was prescribed. **B.** Schematic representation of the model. The parameters of the production, degradation, import and export were scaled by the value of the state function. Production and export (red arrows) were predominant in the apical region (high state function values), while import and degradation (green arrows) were predominant in the central region (low state function values). Dashed line denotes diffusion in the intracellular space. **C.** Extracts of the simulation frames, capturing the dynamics of the intracellular morphogen concentration. The graphs show a line profile of the concentration through a mid-section of the mesh. *At time $t = 200$ the apical regions of the mesh were deleted, simulating the surgical assay. **At time $t = 300$ the state function was changed, simulating the reestablishment of the apical regions. The simulation was run with the following parameters: $D = 1.0$, $p = 0.05$, $k_e = 0.5$, $k_i = 0.5$, $d = 0.05$. In total, the simulation was run for 4000 iterations with a time step $\delta t = 0.1$.

and the central region at $t = 125$ to accumulation in the central region at $t = 200$. The pattern was in agreement with the readings of the auxin-signalling reporter, apart from the recorded accumulation of auxin in the proximal apical region.

In the second stage of the simulation, the central region of the mesh was extracted, mimicking the surgical assay. The concentration of the intracellular auxin dropped to levels comparable to those in the apical regions at the earlier stages of the simulation by $t = 300$, thus recapitulating the recorded decrease of auxin signalling in the isolated central region during the first stage of regeneration (Figure 6.13: Stage 1). At $t = 300$ a new state function was prescribed on the isolated mesh fragment, simulating the emergence of the apical region precursors and the localisation of *proMpYUCCA2* activity (Figure 6.14: Stage 2). The simulation continued by reestablishing the border between the apical regions (low c_i) and the central regions (high c_i) by $t = 400$, in agreement with the experimental observations in the second stage of the regeneration (Figure 6.13: Stage 2).

6.6 Discussion

In this chapter, two DNA constructs *mDYUC2* and *mDDII*, featuring fluorescent reporters of auxin biosynthesis and auxin signalling, were introduced into the *Marchantia* genome. Six gemmae of each line were placed inside a growth chamber and imaged for two days with $\sim 1 h$ interval recording signals from the fluorescent reporters, along with chlorophyll autofluorescence and signal from the membrane-localised fluorescent protein. The nuclear-localised signal was sampled using a nuclear detection algorithm. The coordinates of the sample points were transformed relative to the position of the closest apical notch and normalised by the distance between the two notches. Finally, the spatial distribution of signal from the six gemmae for each line was averaged and compared to the pattern of tissue expansion recorded in chapter 3. In addition, the patterns of the reporters of auxin biosynthesis and signalling were studied in relation to the patterns of cell proliferation in the isolated central regions. The results of the estimation confirmed the hypothesis of the auxin transport mechanism established earlier in chapter 4.

The major findings of this thesis were summarised in a simulation of a computational model, capturing the proposed transport mechanism of auxin in *Marchantia* gemma. The model demonstrated that the mechanism can explain the maintenance of the border between the apical and the central regions. By no means does the model capture the complete system that controls tissue expansion in *Marchantia* gemmae. As discussed in section 1.2, the control of morphogenesis spans several levels of scale: from molecular processes inside the cells to physical aspects at the level of cells and further to the transport events at the level of the whole organism. The model presented in this thesis does not account for intracellular processes regulating cell expansion, effects of other morphogens or physiological aspects of tissue expansion, e.g. water and sugar transport. This model is only valid in the context of the experiments presented in chapters 3, 4 and 6.

Although the fluorescent reporters supported the model of auxin transport mechanism, their readings should not be taken as true representation of auxin concentration or auxin production rates. For example, the validity of the auxin synthesis rate measurement relies on the assumption that the transport of the fluorescent protein and its transcript is similar to that of *MpYUCCA2*. Furthermore, introduction of the additional copy of the promoter

proMpYUCCA2 creates a competition for the transcription-regulation machinery of the original promoter downstream of the *MpYUCCA2* coding sequence. This fact might produce an effect on the total *MpYUCCA2* concentration in the cell, and hence the normal rate of auxin synthesis. Therefore, the introduction of a fluorescent reporter to measure a particular biological signal inevitably introduces biases to the signal being measured.

The linear model explained well the correlations between the measured signals of auxin concentration and tissue expansion, and the signals of auxin concentration and auxin biosynthesis. However, it might not represent the actual biological correlation and causation between the signals. In general, the transfer function between two causal processes caused by ligand binding can be modelled by a sigmoid Hill function^[209]. Therefore, the fixed difference in *YUCCA2* concentration between two cells might produce a biologically high difference in auxin synthesis or no difference at all, depending on the nominal *YUCCA2* concentration.

In addition to the above concerns, the validity of the auxin concentration measurement relied on the assumption that the DII-enabled degradation of the fluorescent protein was homogeneous across the tissue. For example, non-homogeneous distribution of the TIR1/AFB protein, which is responsible for the degradation of the DII-tagged proteins, could have introduced artefacts into the readings.

Furthermore, in the model, the establishment of the border between the central and the apical regions relied on the manual specification of the state function that determined the fate of the cells in the virtual gemmae.

In order to increase the confidence of the auxin biosynthesis pattern, the expression patterns of genes encoding other enzymes of auxin biosynthesis pathways should be collected. To increase the confidence of the auxin-signalling pattern, the assumption about the homogeneous distribution of other enzymes involved in auxin-sensitive degradation should be verified. .

Another advancement to the model can be achieved by introducing the effects of other morphogens on tissue expansion and/or regulation of auxin transport. As mentioned in the discussion section of chapter 4, including cytokinin (antagonist of auxin) into the model of *Marchantia* growth is a next logical step. An attractive model to test would include the concentration-dependent production of auxin and cytokinin, similar to Turing's models of morphogenesis^[43].

Chapter 7

Discussion and Conclusions

7.1 Summary of results

The work presented in this thesis followed the line of research conducted by several previous members of the Haseloff Lab who worked towards a common goal of being able to manipulate the development of plants. Fernan Federici and Lionel Dupuy provided insights into the hormonal regulation of cell division and introduced an *in planta* cytometry method^[112] for the simultaneous recording of the physical and genetic properties of plant cells. Nuri Purswani introduced and characterised *Marchantia* as a model system for studying morphogenesis, and proposed using image registration for measuring tissue strain^[130]. Bernardo Pollak developed frameworks for reprogramming *Marchantia* on the genetic level as well as introducing design standards for fluorescent constructs and the image-processing of their signals in *Marchantia*^[131].

The main contribution of this dissertation is the advancement and standardisation of the methods for the time-lapse imaging of *Marchantia* gemmae; development of the image-processing algorithms for extracting the mean pattern of relative tissue expansion and signals of the fluorescent markers; development of a novel data model for representing genomic information; introduction of a surgical assay for accessible observation of cell proliferation and differentiation; and proposition of the mechanism of auxin-regulated cell proliferation in *Marchantia*.

7.1.1 Genome representation for synthetic biology

Since the engineering of biological organisms happens at the gene level, the future of programmable biology relies on the ability to organise genomic data in a way that facilitates the fast prototyping and assembly of synthetic DNA constructs.

MarpoDB, introduced in chapter 5, is a genomic resource with a unique data model that is ready for downstream synthetic biology applications. MarpoDB represents the *Marchantia* genome as a registry of functional elements, similar to the Registry of Standard Biological parts^[210,211], and allows for fast access to the *Marchantia* genetic parts.

Each gene in MarpoDB is presented as if it was artificially assembled from functional gene elements. An integrated sequence domestication tool, *Recode2s*^[200], aids in preparing the *Marchantia* parts for use in Type IIS DNA assembly systems, in particular, the Loop assembly^[147]. The modular gene model together with the tools available at the web interface allow sequences of functional *Marchantia* DNA elements to be extracted with a few clicks of a mouse.

In addition, MarpoDB is written in Python and is released open-source. This allows for and encourages the development of third-party integrations, e.g. for

high-throughput construct design, batch extraction of promoter sequences, analysis of the annotation data and multi-scale computational models of *Marchantia* development.

Similarly, the bioinformatics pipeline that allowed the conversion of raw genomic and transcriptomic sequencing data into a collection of modular genes is distributed along with MarpoDB. This pipeline can be used to transform the genomic data of any other plant genome into a MarpoDB-like resource. By doing so, the plant synthetic biology community can establish a standard way to represent plant genomes and corresponding genetic parts, thereby promoting the exchange of knowledge and accessibility for comparative plant genome studies.

7.1.2 Methods for recording development of a gemma

Making links between tissue kinematics, cellular processes and genetic dynamics requires the collection of microscopic observations. *Marchantia* gemmae are an attractive target for this purpose: they are small in size and resilient to environmental conditions, and their development can be studied with cellular resolution using a low-power air objective. Growing plants directly under a microscope allows a reduction in time-step between consecutive time frames to minutes and the collection of data with high temporal resolution, making it possible to capture fine details of tissue movement.

In order to image *Marchantia* gemmae continuously for up to four days under a microscope, a 3D-printable growth chamber was designed (Subsection 2.3.1). The chamber is small in size, and can be adhered to a standard $25 \times 75 \text{ mm}$ microscope slide. It can hold up to 24 gemmae in the individual wells, allowing for separate treatment conditions. The chamber is designed to be sealed by a coverslip, which prevents the evaporation of the solid media inside. Preventing the media from evaporating is highly beneficial for obtaining time-lapse recordings; as media evaporates, its volume decreases, causing any sample that is placed on top of it to shift downwards and out from the imaging plane. An alternative approach is to trap a sample in-between the microscopy slide and the coverglass in a drop of liquid media. If the coverglass is sealed, this approach provides a very stable setup with virtually zero sample drift, however, the conditions of the gemma sample in this case are extremely unnatural, i.e. being submerged in liquid and under dorsal-ventral stress exposed by the coverglass. The design of the 3D-printable growth chamber presented in this thesis solves the problem of surface drift while maintaining the gemma in less unnatural conditions and, more importantly, creates a standardised and reproducible micro-environment for the observation of gemmae development.

In chapters 3 and 6, fluorescent reporters along with imaging and image-processing methods were developed. These methods allowed both the tissue kinetics and the dynamics of the gene expression to be tracked simultaneously.

An image-registration-based pipeline, introduced in the earlier work^[130], was advanced to enable the estimation of tissue expansion of *Marchantia gemmae* with sub-cellular resolution using the signal of a membrane-localised fluorescent protein (Chapter 3). The same image registration technique was applied to track the positions of the apical notches, which was an enabling step for the downstream normalisation and averaging of the spatial patterns of tissue expansion and fluorescent-construct output (Section 3.4). The normalised and averaged spatial patterns allow one to account for the natural variation in gemma size and enable the study of correlations between signals obtained from different *Marchantia* lines.

The strategy for recording the expression pattern of the MpYUCCA2 gene, introduced in chapter 6, can be applied to the study of expression patterns for any other *Marchantia* gene. The promoter regions of *Marchantia* genes can be extracted from MarpoDB or alternative genomic resources and assembled into a fluorescent construct upstream of the coding region of a fluorescent protein. When introduced to *Marchantia*, these constructs would create a fluorescent signal proportional to the activity of the cloned promoter. These fluorescent reporters combined with confocal microscopy allow the pattern of gene expression to be recorded with high spatial resolution ($< 1 \mu m$) of cellular signals.

The image processing pipeline introduced in this thesis provides a standard for measuring the expression patterns in *Marchantia gemmae* independent of their overall geometry. The benefit of the normalisation procedure is that the patterns of expression from different experiments can be compared directly; the measurements of the gene expression patterns can be conducted by introducing one fluorescent reporter at a time, and data from experiments separated by both time and space can be analysed in relation to each other. This standard has the potential to play a critical role in the way biological findings are shared and combined between people working on *Marchantia* in different labs across the world.

7.1.3 Role of auxin in gemma development

The methods and tools developed in this thesis were used to study the role of phytohormone auxin in controlling cell proliferation in *Marchantia gemmae*. In chapter 3, the imaging and image-processing pipeline was applied to record patterns of relative tissue expansion from a sample of gemmae for the first two days of development. As a result of the quantification, two cell-fate-specific regions of a gemma were observed: the apical regions, where cells proliferate, and the differentiated central region, where cells are in a non-mitotic state and do not expand.

In chapter 4, gemmae were surgically separated into central and apical regions, and their development was studied in isolation from each other. After the surgical isolation, the cells of the central regions re-entered the cell cycle and formed new apical regions. This observation led to a hypothesis that a hypothetical morphogen, synthesised in the apical regions, is transported and accumulated in the central regions of a gemma, where it represses the proliferation of cells. Upon excision of the apical regions, the flow of the morphogen into the central region is disrupted and the cells of the central region enter proliferation fate.

The application of the exogenous auxin suppressed the proliferation of cells in the isolated central regions, as well as the cells in both the apical and the central regions in the intact gemmae. This led to the conclusion that auxin

might play a role the hypothetical proliferation-repressing morphogen.

In order to support this conclusion, the approximate patterns of auxin concentration and biosynthesis rate were collected chapter 6 by the means of the two fluorescent reporters

The first fluorescent reporter captured the expression pattern of the MpYUCCA2 gene, encoding an enzyme that catalyses the final step in auxin biosynthesis. The second contained a fluorescent protein fused with the auxin-sensitive DII domain of the MpAux/IAA gene. The reporters were introduced into the Marchantia genome, producing two separate transgenic lines. Six gemmae of each transgenic line were imaged for two days, and the average spatial pattern of fluorescent-reporter expression was calculated. Based on the mathematical models of protein dynamics, the signal from the first reporter was assumed to be proportional to the rate of auxin biosynthesis, the second to the cytosolic auxin concentration. The collected observations supplied evidence of high levels of auxin biosynthesis in the apical regions, and elevated levels of auxin signalling in the central region. Furthermore, the expression of the auxin-signalling reporter dropped in the isolated central regions, coinciding with the active cell proliferation, and further reestablished high expression outside of the newly-formed apical regions. On the contrary, the expression of the auxin-biosynthesis reporter in the isolated central regions increased during the first day following the surgery, and later localised to the newly-formed apical regions, thus phenocopying the expression pattern in the intact gemmae.

These observations led to the formulation of the following model of auxin-controlled cell proliferation. Auxin plays the role of the inhibitor of cell proliferation in Marchantia. It is produced in the apical regions before being transported to and accumulated in the central region. The proliferation of cells is inhibited in the central region due to the high concentration of auxin, while the low concentration of auxin in the apical regions allows cell division and expansion. Upon excision of the apical regions, the influx of the auxin is disrupted, leading to a drop in auxin concentration in the cells of the central region and their re-entry into the cell cycle. A negative feedback of auxin on its own biosynthesis induces the production of auxin in the cells of the isolated central region and the establishment of new apical regions.

This model was summarised and visualised by means of computer simulation. An implication of this model suggests that auxin plays a major role in controlling cell proliferation in Marchantia. Therefore, auxin transport is an attractive system for the purpose of engineering the shape of Marchantia.

7.2 Future work

7.2.1 Model development

The research performed in this thesis laid out a model for the auxin-dependent regulation of Marchantia tissue expansion rates. As discussed in chapter 6, the model suggests that the repression of tissue expansion happens in areas of auxin accumulation. The significance of this model from the perspective of the engineering of plant morphogenesis is that it enables auxin to be used as a means for establishing novel patterns of Marchantia tissue growth. In order to proceed with this line of research, one has to understand how Marchantia regulates the formation of auxin sources and sinks, i.e. how the regulation of auxin transport works. An attractive explanation, supported

by earlier experiments and used as a foundation for many auxin transport models^[212], is that auxin forms a feedback loop on its own transport.

For example, it has been previously demonstrated that the activity of the YUC gene is transcriptionally down-regulated upon external application of auxin in Arabidopsis^[61] as well as other species, including *Marchantia*^[62]. The negative correlation of the *proMpYUCCA2* reporter signal with the signal of the auxin-signalling reporter, demonstrated in chapter 6, also supports the existence of the negative feedback auxin has on its own production.

It has also been previously shown that intracellular auxin inhibits endosomal compartmentalisation of the auxin efflux protein PIN1 in Arabidopsis, promoting the localisation of the PIN1 at the plasma membrane and increasing the rate of auxin export^[54]. Polarisation of the PIN1 proteins at the plasma membrane is attributed as a key mechanism behind auxin pattern establishment, and therefore control of developmental processes. Development and analysis of the auxin transport models where production and export rates depend on auxin concentration will provide new insights into how auxin transport might be regulated, and hence what would be the best strategy to rewire and exploit it for the formation of novel auxin source and sink patterns.

The computational models of auxin transport should be implemented on growing domains, where the expansion rate is governed by the concentration of auxin. Not only will this allow the effects of auxin on tissue expansion to be better visualised, it would also add additional terms to the transport model, namely the dilution of auxin due to tissue expansion and the advection of auxin due to tissue movement. Experimental methods

The range of methods for measuring transport of auxin in *Marchantia gemmae* should be expanded in order to test the mathematical models. The fluorescent reporters, based on the promoter fusion strategy, should be designed for the auxin exporters (PINs) and auxin importers (LAX). Alternative methods for recording patterns of auxin concentration should also be developed in *Marchantia*. Fluorescent reporters based on synthetic DR5^[213] and DR5v2^[214] promoters, containing several auxin-response-factor-binding motifs, are a popular choice for measuring auxin signalling. A DR5v2::Venus-containing construct was constructed using Loop assembly as part of the work presented in this thesis, and introduced into *Marchantia* spores together with a hygromycin resistance gene. The growth of the sporelings on the selective media indicated a successful integration of the construct into the *Marchantia* genome, however, no fluorescent signal was observed in the selected plants up to the mature stage $\sim 2 - 3$ weeks after the transformation. The lack of observed fluorescent signal might be explained by the fact that the auxin-transcription-binding motifs employed in DR5v2 were based on the crystal structure of the Arabidopsis ARF5 protein. These binding motifs might not be compatible with *Marchantia* ARFs. Hence, an alternative synthetic promoter contacting *Marchantia*-specific ARF-binding motifs might be explored.

Similarly, the lack of fluorescent signal was noted in lines containing *proMpPIN1:mVenus* and *proMpPIN1:MpPIN1-mVenus* with sequence of *proMpPIN1* obtained as a 3kb region upstream of the *MpPIN1* transcription start site. The lack of fluorescent signals observed from these constructs may be explained by low activity among the promoters or insufficient promoter length. To address the latter possibility, a fluorescent reporter using a longer extract of the *MpPIN1* promoter might be used.

In chapter 6, the promoter of the *MpYUCCA2* gene was used to build a fluorescent reporter, providing a proxy signal for auxin biosynthesis. A similar approach could be applied to study the expression patterns of other genes in

the *Marchantia* genome related to the regulation of tissue expansion in *Marchantia*. For example, expression patterns of aquaporins, cellulose synthases and expansins can be used to draw further conclusions about auxin transport and the cellular processes connected with water transport and cell-wall remodelling.

Inferring correlations between these signals still would not be sufficient to conclude that auxin is causally responsible for the regulation of tissue expansion. To gain strong evidence about such causality one would have to demonstrate the link between changes in auxin concentration and the resultant tissue expansion rate at the level of single cells. One experimental technique that would allow such a demonstration relies on inducible promoters. For example, a fluorescent reporter based on an endogenous *Marchantia* promoter proMpHSP17.8A1 has been shown to produce a signal in response to a mild heat-shock of 37° [146]. It has been demonstrated that the induction of the promoter can be achieved with a cellular resolution by heating individual cells with an infrared laser. The later development of the heat-induction system in *Marchantia* showed that by fusing a glucocorticoid receptor (GR domain) to a fluorescent protein under the control of proMpHSP17.8A1 it is possible to achieve a high signal-to-noise ratio during combined dexamethasone and laser-heating treatment [131]. In addition, it was shown that this system can be activated by heat generated by the laser of a confocal microscope, thus allowing induction between the image acquisition cycles [131]. By introducing proMpHSP17.8A1:MpYUCCA2-GR together with the fluorescent reporters of auxin signalling and hypothetical downstream targets of auxin, one would be able to record causal links between the increase of auxin concentration and auxin response. Similar strategy might be employed to create a heat-inducible auxin degradation construct based on the bacterial auxin conjugation enzyme *iaaL* [215], which is capable of deactivating auxin molecules by binding them to lysine.

Complementary approaches can be employed to identify the links between a cell's epigenetic state and its function. For example, single-cell RNA sequencing (SCS) has been tested in *Marchantia* gemmae as part of the work presented in this thesis, showing the segregation of cells by their epigenetic state (Figure 7.1). However, the link between the cell clusters, identified in the SCS experiment, and actual cell types is yet to be determined.

The annotation data stored in MarpoDB can provide additional insight into the functional segregation of *Marchantia* cells. For example, Gene Ontology (GO) annotations were extracted from MarpoDB for each gene and recorded in an annotation matrix, where each row corresponded to an individual gene and each column to a single GO term (Figure 7.1 B). The elements of the matrix were assigned to 1 if the gene contained the corresponding GO annotation, and 0 otherwise. The multiplication of the single-cell matrix with the gene expression data by the annotation matrix resulted in a functional single-cell matrix the elements of which summarise the number of sequencing reads that support a corresponding cellular function (Figure 7.1 C). Analysis of the matrix produced additional insight into the organisation of the SCS data, which could be used for the identification of cell types in *Marchantia* (Figure 7.1 D).

7.3 General discussion

The work presented in this thesis covered a broad range of topics united under a general flag of engineering morphogenesis. The breadth of this work is justified by that of the various scientific disciplines and tools required to address plant morphogenesis. The ability to find a solution to the problem relies on knowledge from physical disciplines

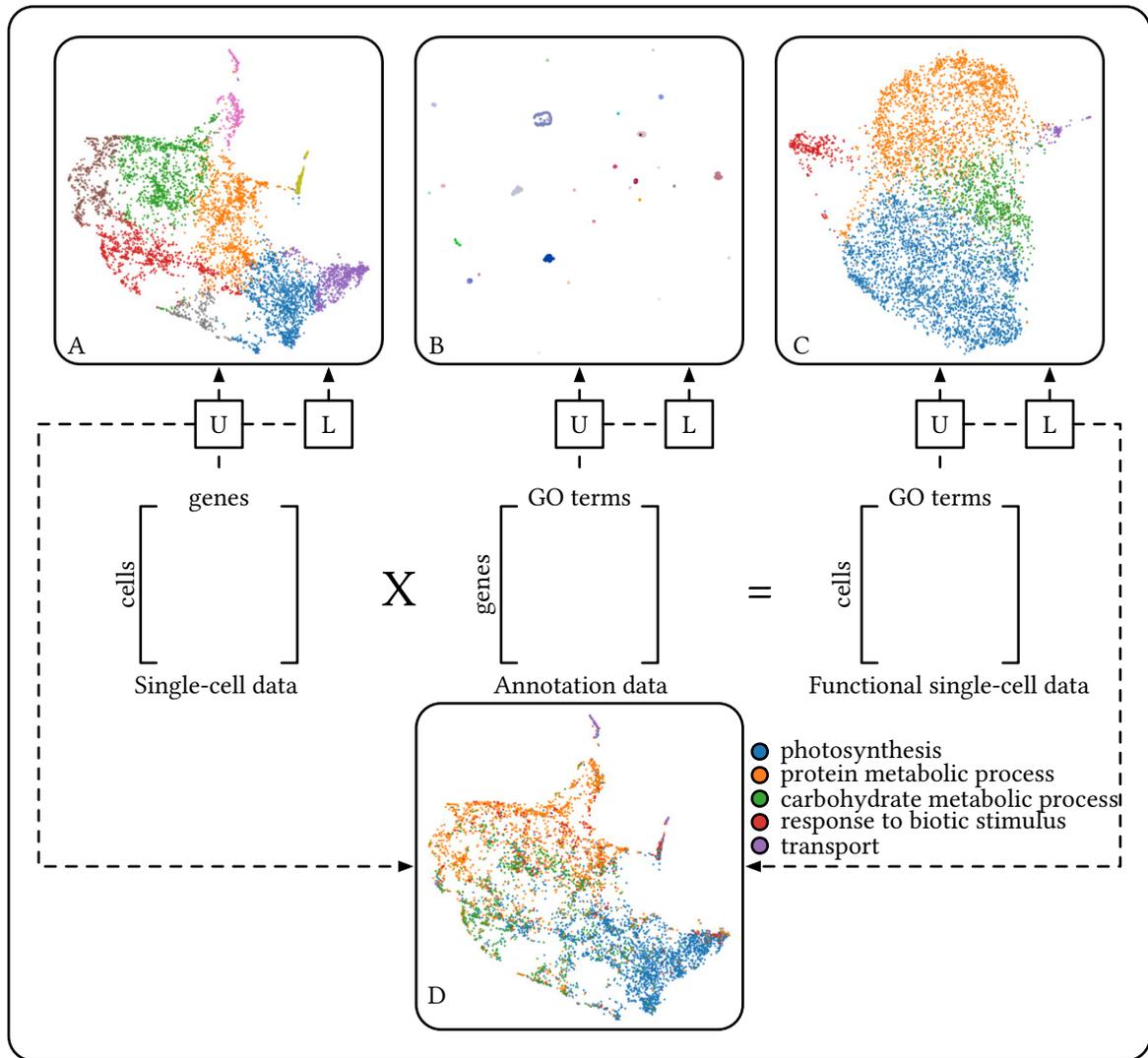


Figure 7.1: **Summary of the single-cell RNA-sequencing data.** **A.** Visualisation of the single-cell RNA-sequencing data. **B.** Visualisation of the gene clustering by biological function. **C.** Visualisation of the cell clustering by biological function. **D.** Projection of the biological function clusters on the single-cell RNA-sequencing data. One of the top 5 representative Gene Ontology terms for each cluster is shown on the right. The pairs of the square brackets correspond to the matrices, left to right: single-cell data (number of unique mRNA molecules in each cell), gene annotation data (presence of the corresponding Gene Ontology term for each gene), functional single-cell data (weighted summary of the Gene Ontology terms for each cell). The axes of the scatter plots correspond to the first two components of the uniform manifold approximation and projection^[216] (U) of the given matrices. Each dot corresponds to the individual cell in the single-cell and functional single-cell data, or individual gene in the annotation data. Dots of the same colour belong to the same cluster, identified using the Louvain method (L)^[217].

like fluid dynamics and morphoelasticity, and biological disciplines like plant physiology and cell biology as well as myriad practical skills encompassing molecular biology, microscopy, image processing and computational simulation.

Although some find it more appealing to research a particular topic in-depth rather than in-breadth, the broad multidisciplinary approach allows knowledge from multiple fields to be brought together into a united view of the problem. Therefore, arguably the greatest value of this work is not in the main results outlined in the previous section, but rather in illustrating an approach that might be taken when addressing the expansiveness of the problem of engineering morphogenesis.

Following this principle, the approach taken in this thesis was aimed at gaining general biological and physical intuition about the regulation of *Marchantia gemmae* growth rather than going deeply into the fine details of the underlying biological processes. Throughout the whole work, priority was given to practicality and transparency rather than comprehensiveness. For example, during the normalisation of patterns of gemma tissue expansion in chapter 3, the deformation of the original coordinate frames and the resultant loss of the true positional information was chosen in favour of obtaining a biologically inaccurate yet theoretically clear result; in chapter 4, a simplistic yet easily testable transport model was chosen to explain the growth pattern; in chapter 5, large sections of *Marchantia* genome, i.e. its non-coding regions, were ignored in the data model in favour of producing a clean and usable dataset of genes ready for the design of synthetic constructs; in chapter 6, a set of imprecise *in vivo* fluorescent reporters was chosen instead of the in-depth biological and analytical analysis of auxin transport in the gemmae.

Particular care was taken when proposing the model of regulation of tissue expansion in *Marchantia gemmae*. Most biological models suffer from being introduced following an inductive approach, i.e. generalising observations into the formation of new theories. The danger of using inductivism in biology is that there usually exist many different ways to collect indirect observations of a phenomenon of interest, e.g. concentration of auxin can be estimated by fluorescent or other reporters of several different designs, and each of the sets of observations due to their imperfections might be interpreted and generalised into principally different theories. An alternative is to follow a problem-solving approach, where theories play the primary role, while the role of experiments is to attempt to disprove or test those theories.

For example, the state of the art model for *Marchantia* growth regulation, discussed in chapter 4, was derived using the inductive approach. The authors observed that the rate of tissue expansion decreases as a function of distance from the closest apical notch. They then formed a hypothesis that there exists a growth-inducing morphogen, which is produced in proximity to the apical notches, and the pattern of tissue expansion can be explained by the drop of the morphogen concentration as it diffuses away from the apical regions. At this point in time, the authors did not consider any alternative hypothesis and continued experimentation. They further acknowledged that their initial model does not explain the behaviour of the tissue when subject to the surgical excision of the apical notches. Instead of taking a step back to explore alternative equally simplistic hypotheses, they ‘patched’ their model by introducing a differential response to the growth-inducing morphogen in different tissue regions. In doing so, they upgraded their initial model in a manner that gave it greater predictive power, but at the same time explained the phenomenon of *Marchantia* growth in a less clear way; not only did it have to address the nature of the hypothetical morphogen, but also the mechanism behind the differential response of the tissue to that morphogen. Although by following such an

approach, where each step opens up more and more questions, one can garner vast amounts of experimental data about the finer details of *Marchantia* cellular response, genetic regulation, etc., one can argue this does not increase the general understanding of how the tissue expansion of *Marchantia* is regulated.

A problem-solving approach was instead employed to study the same question in chapter 4. Instead of diving deep into an exploration of a particular hypothesis, several alternative hypotheses fitting the data were explored. The follow-up experiment was designed to eliminate one or several competing hypotheses. As a result, three out of four hypotheses were eliminated. In doing so, no extra details were introduced, no additional questions were opened, and the mechanism of the regulation of tissue expansion was explained in a clear and simplistic model. To sum up, both studies collected similar data and performed similar surgical experiments, yet come up with very different models and explanations of the growth regulation. The only difference was in the approach; the first study took a depth-first approach, where hypotheses are tested experimentally one by one until they break, while the second took a breadth-first approach, where several hypotheses are considered at once and experiments are designed to eliminate a number of possible explanations. Since biology is inherently complex and full of details, the breadth-first approach might be better-suited for gaining conceptual understanding of biological systems required for solving problems like engineering plant morphogenesis.

7.4 Conclusion

The work presented in this thesis contributed to the methods for studying multi-scale processes of development of *Marchantia* gemmae.

The novel representation of the *Marchantia* genome will facilitate the extraction of functional elements and the creation of fluorescent reporters for studying spatial patterns of cellular and genetic processes. The imaging and image-processing tools will allow for the observation and correlation of these patterns with differential expansion of the gemma tissue. The developed model and observations of the role of auxin in cell expansion expanded the understanding of the regulation of growth in *Marchantia*, and provided insight into possible ways of engineering *Marchantia* shape through genetic manipulations.

Bibliography

- [1] Maya (Autodesk, 2019). <<https://www.autodesk.co.uk/products/maya/>>.
- [2] 3ds max (Autodesk, 2019). <<https://www.autodesk.co.uk/products/3ds-max/>>.
- [3] Blender (Blender Foundation, 2019). <<https://www.blender.org/about/>>.
- [4] ZBrush (Pixologic, 2019).
- [5] Thompson, D. W. *On Growth and Form* (Cambridge University Press, 1917).
- [6] Goriely, A. *The Mathematics and Mechanics of Biological Growth*. Interdisciplinary Applied Mathematics (Springer-Verlag, 2017).
- [7] Goriely, A. & Ben Amar, M. On the definition and modeling of incremental, cumulative, and continuous growth laws in morphoelasticity. *Biomechanics and Modeling in Mechanobiology* **6**, 289–296 (2007). doi:10.1007/s10237-006-0065-7.
- [8] Goriely, A. & Moulton, D. *Morphoelasticity: A theory of elastic growth* (Oxford University Press, 2011).
- [9] Silk, W. K. & Erickson, R. O. Kinematics of plant growth. *Journal of Theoretical Biology* **76**, 481–501 (1979). doi:10.1016/0022-5193(79)90014-6.
- [10] Richards, O. W. & Kavanagh, A. J. The analysis of the relative growth gradients and changing form of growing organisms: Illustrated by the tobacco leaf. *The American Naturalist* **77**, 385–399 (1943).
- [11] Dumais, J. & Kwiatkowska, D. Analysis of surface growth in shoot apices. *The Plant Journal* **31**, 229–241 (2002). doi:10.1046/j.1365-3113.2001.01350.x.
- [12] Wolf, S. D., Silk, W. K. & Plant, R. E. Quantitative patterns of leaf expansion: Comparison of normal and malformed leaf growth in *vitis vinifera* cv. ruby red. *American Journal of Botany* **73**, 832–846 (1986). doi:10.2307/2444294.
- [13] Rolland-Lagan, A.-G., Remmler, L. & Girard-Bock, C. Quantifying shape changes and tissue deformation in leaf development. *Plant Physiology* **165**, 496–505 (2014). doi:10.1104/pp.113.231258.

- [14] Gupta, M. D. & Nath, U. Divergence in patterns of leaf growth polarity is associated with the expression divergence of miR396. *The Plant Cell* **27**, 2785–2799 (2015). doi:10.1105/tpc.15.00196.
- [15] Hamant, O. & Traas, J. The mechanics behind plant development. *New Phytologist* **185**, 369–385 (2010). doi:10.1111/j.1469-8137.2009.03100.x.
- [16] Lockhart, J. A. An analysis of irreversible plant cell elongation. *Journal of Theoretical Biology* **8**, 264 – 275 (1965). doi:10.1016/0022-5193(65)90077-9.
- [17] Cosgrove, D. J., Van Volkenburgh, E. & Cleland, R. E. Stress relaxation of cell walls and the yield threshold for growth: demonstration and measurement by micro-pressure probe and psychrometer techniques. *Planta* **162**, 46–54 (1984).
- [18] Ray, P. M., Green, P. B. & Cleland, R. Role of turgor in plant cell growth. *Nature* **239**, 163–164 (1972). doi:10.1038/239163a0.
- [19] Ortega, J. K. E. Dimensionless number is central to stress relaxation and expansive growth of the cell wall. *Scientific Reports* **7**, 3016 (2017). doi:10.1038/s41598-017-03002-6.
- [20] Cosgrove, D. J. How do plant cell walls extend? *Plant Physiology* **102**, 1–6 (1993). doi:10.1104/pp.102.1.1.
- [21] Cosgrove, D. J. Wall extensibility: its nature, measurement and relationship to plant cell growth. *The New Phytologist* **124**, 1–23 (1993). doi:10.1111/j.1469-8137.1993.tb03795.x.
- [22] Cosgrove, D. J. Diffuse growth of plant cell walls. *Plant Physiology* **176**, 16–27 (2018).
- [23] Cleland, R. E., Cosgrove, D. & Tepfer, M. Long-term acid-induced wall extension in an in-vitro system. *Planta* **170**, 379–385 (1987). doi:10.1007/BF00395030.
- [24] Rayle, D. L. & Cleland, R. Enhancement of wall loosening and elongation by acid solutions. *Plant Physiology* **46**, 250–253 (1970). doi:10.1104/pp.46.2.250.
- [25] Taiz, L. Plant cell expansion: Regulation of cell wall mechanical properties. *Annual Review of Plant Physiology* **35**, 585–657 (1984). doi:10.1146/annurev.pp.35.060184.003101.
- [26] McQueen-Mason, S., Durachko, D. M. & Cosgrove, D. J. Two endogenous proteins that induce cell wall extension in plants. *The Plant Cell* **4**, 1425–1433 (1992). doi:10.1105/tpc.4.11.1425.
- [27] Georgelis, N., Tabuchi, A., Nikolaidis, N. & Cosgrove, D. J. Structure-function analysis of the bacterial expansin EXLX1. *The Journal of Biological Chemistry* **286**, 16814–16823 (2011). doi:10.1074/jbc.M111.225037.
- [28] Brummell, D. A., Harpster, M. H. & Dunsmuir, P. Differential expression of expansin gene family members during growth and ripening of tomato fruit. *Plant Molecular Biology* **39**, 161–169 (1999). doi:10.1023/A:1006130018931.

- [29] Cho, H. T. & Kende, H. Expression of expansin genes is correlated with growth in deepwater rice. *The Plant Cell* **9**, 1661–1671 (1997). doi:10.1105/tpc.9.9.1661.
- [30] Fleming, A. J., McQueen-Mason, S., Mandel, T. & Kuhlemeier, C. Induction of leaf primordia by the cell wall protein expansin. *Science* **276**, 1415–1418 (1997). doi:10.1126/science.276.5317.1415.
- [31] Link, B. M. & Cosgrove, D. J. Acid-growth response and α -expansins in suspension cultures of bright yellow 2 tobacco. *Plant Physiology* **118**, 907–916 (1998).
- [32] Kimura, S. *et al.* Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant vigna angularis. *The Plant Cell* **11**, 2075–2085 (1999). doi:10.1105/tpc.11.11.2075.
- [33] Baskin, T. I. Anisotropic expansion of the plant cell wall. *Annual Review of Cell and Developmental Biology* **21**, 203–222 (2005). doi:10.1146/annurev.cellbio.20.082503.103053.
- [34] Green, P. B. Mechanism for plant cellular morphogenesis. *Science* **138**, 1404–1405 (1962). doi:10.1126/science.138.3548.1404.
- [35] Paredez, A. R., Somerville, C. R. & Ehrhardt, D. W. Visualization of cellulose synthase demonstrates functional association with microtubules. *Science* **312**, 1491–1495 (2006). doi:10.1126/science.1126551.
- [36] Hamant, O. *et al.* Developmental patterning by mechanical signals in arabidopsis. *Science* **322**, 1650–1655 (2008). doi:10.1126/science.1165594.
- [37] Uyttewaal, M. *et al.* Mechanical stress acts via katanin to amplify differences in growth rate between adjacent cells in arabidopsis. *Cell* **149**, 439–451 (2012). doi:10.1016/j.cell.2012.02.048.
- [38] Liu, D. *et al.* Characterization and expression of plasma and tonoplast membrane aquaporins in elongating cotton fibers. *Plant Cell Reports* **27**, 1385–1394 (2008). doi:10.1007/s00299-008-0545-6.
- [39] Chen, W. *et al.* Involvement of rose aquaporin RhPIP1;1 in ethylene-regulated petal expansion through interaction with RhPIP2;1. *Plant Molecular Biology* **83**, 219–233 (2013). doi:10.1007/s11103-013-0084-6.
- [40] Ma, N. *et al.* Rh-PIP2;1, a rose aquaporin gene, is involved in ethylene-regulated petal expansion. *Plant Physiology* **148**, 894–907 (2008). doi:10.1104/pp.108.120154.
- [41] Maurel, C. *et al.* Molecular physiology of aquaporins in plants. In *International Review of Cytology*, vol. 215 of *Molecular Mechanisms of Water Transport Across Biological Membranes*, 105–148 (Academic Press, 2002). <<http://www.sciencedirect.com/science/article/pii/S0074769602150078>>.
- [42] Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology* **25**, 1–47 (1969). doi:10.1016/S0022-5193(69)80016-0.

- [43] Turing, A. M. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond. B* **237**, 37–72 (1952). doi:10.1098/rstb.1952.0012.
- [44] Hager, A. Role of the plasma membrane H⁺-ATPase in auxin-induced elongation growth: historical and new aspects. *Journal of Plant Research* **116**, 483–505 (2003). doi:10.1007/s10265-003-0110-x.
- [45] Barbez, E., Dünser, K., Gaidora, A., Lendl, T. & Busch, W. Auxin steers root cell expansion via apoplastic pH regulation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **114**, E4884–E4893 (2017). doi:10.1073/pnas.1613499114.
- [46] Majda, M. & Robert, S. The role of auxin in cell wall expansion. *International Journal of Molecular Sciences* **19** (2018). doi:10.3390/ijms19040951.
- [47] Nemhauser, J. L., Hong, F. & Chory, J. Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. *Cell* **126**, 467–475 (2006). doi:10.1016/j.cell.2006.05.050.
- [48] Heisler, M. G. *et al.* Alignment between PIN1 polarity and microtubule orientation in the shoot apical meristem reveals a tight coupling between morphogenesis and auxin transport. *PLOS Biology* **8**, e1000516 (2010). doi:10.1371/journal.pbio.1000516.
- [49] Lavy, M. & Estelle, M. Mechanisms of auxin signaling. *Development (Cambridge, England)* **143**, 3226–3229 (2016). doi:10.1242/dev.131870.
- [50] Kagale, S. & Rozwadowski, K. EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression. *Epigenetics* **6**, 141–146 (2011). doi:10.4161/epi.6.2.13627.
- [51] Guilfoyle, T. J. The PB1 domain in auxin response factor and Aux/IAA proteins: A versatile protein interaction module in the auxin response. *The Plant Cell* **27**, 33–43 (2015). doi:10.1105/tpc.114.132753.
- [52] Ulmasov, T., Liu, Z. B., Hagen, G. & Guilfoyle, T. J. Composite structure of auxin response elements. *The Plant Cell* **7**, 1611–1623 (1995). doi:10.1105/tpc.7.10.1611.
- [53] Boer, D. R. *et al.* Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell* **156**, 577–589 (2014). doi:10.1016/j.cell.2013.12.027.
- [54] Paciorek, T. *et al.* Auxin inhibits endocytosis and promotes its own efflux from cells. *Nature* **435**, 1251–1256 (2005). doi:10.1038/nature03633.
- [55] Krupinski, P. & Jönsson, H. Modeling auxin-regulated development. *Cold Spring Harb Perspect Biol* a001560 (2010).

- [56] Jönsson, H., Heisler, M. G., Shapiro, B. E., Meyerowitz, E. M. & Mjolsness, E. An auxin-driven polarized transport model for phyllotaxis. *Proceedings of the National Academy of Sciences* **103**, 1633–1638 (2006). doi:10.1073/pnas.0509839103.
- [57] Smith, R. S. *et al.* A plausible model of phyllotaxis. *Proceedings of the National Academy of Sciences* **103**, 1301–1306 (2006). doi:10.1073/pnas.0510457103.
- [58] Mitchison, G. J. A model for vein formation in higher plants. *Proceedings of the Royal Society B: Biological Sciences* **207**, 79–109 (1980). doi:10.1098/rspb.1980.0015.
- [59] Fujita, H. & Mochizuki, A. Pattern formation of leaf veins by the positive feedback regulation between auxin flow and auxin efflux carrier. *Journal of Theoretical Biology* **241**, 541–551 (2006). doi:10.1016/j.jtbi.2005.12.016.
- [60] Swarup, K. *et al.* The auxin influx carrier LAX3 promotes lateral root emergence. *Nature Cell Biology* **10**, 946–954 (2008). doi:10.1038/ncb1754.
- [61] Suzuki, M. *et al.* Transcriptional feedback regulation of YUCCA genes in response to auxin levels in arabidopsis. *Plant Cell Reports* **34**, 1343–1352 (2015). doi:10.1007/s00299-015-1791-z.
- [62] Mutte, S. K. *et al.* Origin and evolution of the nuclear auxin response system. *eLife* **7**, e33399 (2018). doi:10.7554/eLife.33399.
- [63] Harrison, C. J. The evolution of plant form. *Annals of Botany* **112**, vii–viii (2013). doi:10.1093/aob/mct116.
- [64] Purugganan, M. D. & Fuller, D. Q. The nature of selection during plant domestication. *Nature* **457**, 843–848 (2009). doi:10.1038/nature07895.
- [65] Kelly, T. J. & Smith, H. O. A restriction enzyme from hemophilus influenzae: II. base sequence of the recognition site. *Journal of Molecular Biology* **51**, 393 – 409 (1970).
- [66] Smith, H. O. & Welcox, K. A restriction enzyme from hemophilus influenzae: I. purification and general properties. *Journal of Molecular Biology* **51**, 379 – 391 (1970).
- [67] Weiss, B. & Richardson, C. C. Enzymatic breakage and joining of deoxyribonucleic acid, i. repair of single-strand breaks in dna by an enzyme system from escherichia coli infected with t4 bacteriophage. *Proceedings of the National Academy of Sciences* **57**, 1021–1028 (1967). doi:10.1073/pnas.57.4.1021. <https://www.pnas.org/content/57/4/1021.full.pdf>.
- [68] Hernalsteens, J.-P. The Agrobacterium tumefaciens Ti plasmid as a host vector system for introducing foreign DNA in plant cells. **287**, 654–656 (1980). doi:10.1038/287654a0.

- [69] Fromm, M., Taylor, L. P. & Walbot, V. Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proceedings of the National Academy of Sciences* **82**, 5824–5828 (1985). doi:10.1073/pnas.82.17.5824. <https://www.pnas.org/content/82/17/5824.full.pdf>.
- [70] Shillito, R. D., Saul, M. W., Paszkowski, J., Müller, M. & Potrykus, I. High efficiency direct gene transfer to plants. *Bio/Technology* **3**, 1099–1103 (1985). doi:10.1038/nbt1285-1099.
- [71] Aly, M. A. M. & Owens, L. D. A simple system for plant cell microinjection and culture. *Plant Cell, Tissue and Organ Culture* **10**, 159–174 (1987). doi:10.1007/BF00037301.
- [72] Sanford, J. C. Biolistic plant transformation. *Physiologia Plantarum* **79**, 206–209 (1990). doi:10.1111/j.1399-3054.1990.tb05888.x. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1399-3054.1990.tb05888.x>.
- [73] Chaleff, R. S. & Ray, T. B. Herbicide-resistant mutants from tobacco cell cultures. *Science* **223**, 1148–1151 (1984). doi:10.1126/science.223.4641.1148. <https://science.sciencemag.org/content/223/4641/1148.full.pdf>.
- [74] Vaeck, M. *et al.* Transgenic plants protected from insect attack. *Nature* **328**, 33–37 (1987). doi:10.1038/328033a0.
- [75] Bruening, G. & Lyons, J. The case of the FLAVR SAVR tomato. *California Agriculture* **54**, 6–7 (2000-07-01).
- [76] Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000). doi:10.1038/35002131.
- [77] Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000). doi:10.1038/35002125.
- [78] Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012). doi:10.1038/nature11516.
- [79] Anderson, J. C., Voigt, C. A. & Arkin, A. P. Environmental signal integration by a modular AND gate. *Molecular systems biology* **3**, 133–133 (2007). doi:10.1038/msb4100173. 17700541.
- [80] Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nature methods* **11**, 1261–1266 (2014). doi:10.1038/nmeth.3147. 25344638.
- [81] Daniel, R., Rubens, J. R., Sarpeshkar, R. & Lu, T. K. Synthetic analog computation in living cells. *Nature* **497**, 619–623 (2013). doi:10.1038/nature12148.
- [82] Tomazou, M., Barahona, M., Polizzi, K. M. & Stan, G.-B. Computational re-design of synthetic genetic oscillators for independent amplitude and frequency modulation. *Cell Systems* **6**, 508–520.e5 (2018). doi:10.1016/j.cels.2018.03.013.

- [83] Basu, S., Mehreja, R., Thiberge, S., Chen, M.-T. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6355 (2004). doi:10.1073/pnas.0307571101.
- [84] Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005). doi:10.1038/nature03461.
- [85] Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352** (2016). doi:10.1126/science.aac7341. <https://science.sciencemag.org/content/352/6281/aac7341.full.pdf>.
- [86] Cohen, S. N., Chang, A. C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **70**, 3240–3244 (1973-11). doi:10.1073/pnas.70.11.3240.4594039.
- [87] Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345 (2009). doi:10.1038/nmeth.1318.
- [88] Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA cloning using in vitro site-specific recombination. *Genome research* **10**, 1788–1795 (2000). doi:10.1101/gr.143000.11076863.
- [89] Engler, C. *et al.* A golden gate modular cloning toolbox for plants. *ACS Synthetic Biology* **3**, 839–843 (2014). doi:10.1021/sb4001504.
- [90] Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS one* **3**, e3647–e3647 (2008). doi:10.1371/journal.pone.0003647.18985154.
- [91] Knight, T. Idempotent vector design for standard assembly of biobricks. *MIT Artificial Intelligence Laboratory, MIT Synthetic Biology Working Group* (2003).
- [92] Patron, N. J. *et al.* Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *The New Phytologist* **208**, 13–19 (2015). doi:10.1111/nph.13532.
- [93] Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**, D1178–D1186 (2012). doi:10.1093/nar/gkr944.
- [94] Twyford, A. D. The road to 10,000 plant genomes. *Nature Plants* **4**, 312 (2018). doi:10.1038/s41477-018-0165-2.
- [95] Lemmon, Z. H. *et al.* Rapid improvement of domestication traits in an orphan crop by genome editing. *Nature Plants* **4**, 766 (2018). doi:10.1038/s41477-018-0259-x.
- [96] Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq: reference generation and analysis with trinity. *Nature protocols* **8** (2013). doi:10.1038/nprot.2013.084.

- [97] Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Research* **40**, D290–D301 (2012). doi:10.1093/nar/gkr1065.
- [98] Gough, J., Karplus, K., Hughey, R. & Chothia, C. Assignment of homology to genome sequences using a library of hidden markov models that represent all proteins of known structure11edited by g. von heijne. *Journal of Molecular Biology* **313**, 903–919 (2001). doi:10.1006/jmbi.2001.5080.
- [99] Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B. & Gough, J. The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Research* **47**, D490–D494 (2018). doi:10.1093/nar/gky1130. <http://oup.prod.sis.lan/nar/article-pdf/47/D1/D490/27437286/gky1130.pdf>.
- [100] Sigrist, C. J. A. *et al.* PROSITE, a protein domain database for functional characterization and annotation. *Nucleic acids research* **38**, D161–D166 (2010). doi:10.1093/nar/gkp885. 19858104.
- [101] Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)* **30**, 1236–1240 (2014). doi:10.1093/bioinformatics/btu031.
- [102] Finn, R. D. *et al.* InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Research* **45**, D190–D199 (2017). doi:10.1093/nar/gkw1107.
- [103] Kapustin, Y., Souvorov, A., Tatusova, T. & Lipman, D. Splign: algorithms for computing spliced alignments with identification of paralogs. *Biology Direct* **3**, 20 (2008). doi:10.1186/1745-6150-3-20.
- [104] Huang, S. *et al.* SOAPsplice: Genome-wide ab initio detection of splice junctions from RNA-seq data. *Frontiers in genetics* **2**, 46–46 (2011). doi:10.3389/fgene.2011.00046. 22303342.
- [105] Jammali, S., Aguilar, J.-D., Kuitche, E. & Ouangraoua, A. SplicedFamAlign: CDS-to-gene spliced alignment and identification of transcript orthology groups. *BMC Bioinformatics* **20**, 133 (2019). doi:10.1186/s12859-019-2647-2.
- [106] Bu, J., Chi, X. & Jin, Z. HSA: a heuristic splice alignment tool. *BMC systems biology* **7 Suppl 2**, S10–S10 (2013). doi:10.1186/1752-0509-7-S2-S10. 24564867.
- [107] Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, aequorea. *Journal of Cellular and Comparative Physiology* **59**, 223–239 (1962). doi:10.1002/jcp.1030590302. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcp.1030590302>.
- [108] Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. Primary structure of the aequorea victoria green-fluorescent protein. *Gene* **111**, 229 – 233 (1992). doi:[https://doi.org/10.1016/0378-1119\(92\)90691-H](https://doi.org/10.1016/0378-1119(92)90691-H).

- [109] Rodriguez, E. A. *et al.* The growing and glowing toolbox of fluorescent and photoactive proteins. *Trends in Biochemical Sciences* **42**, 111 – 129 (2017). doi:<https://doi.org/10.1016/j.tibs.2016.09.010>.
- [110] Ulmasov, T., Murfett, J., Hagen, G. & Guilfoyle, T. J. Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *The Plant Cell* **9**, 1963–1971 (1997). doi:10.1105/tpc.9.11.1963.
- [111] Laskowski, M. *et al.* Root system architecture from coupling cell shape to auxin transport. *PLOS Biology* **6**, 1–15 (2008). doi:10.1371/journal.pbio.0060307.
- [112] Federici, F., Dupuy, L., Laplaze, L., Heisler, M. & Haseloff, J. Integrated genetic and computation methods for *in planta* cytometry. *Nature Methods* **9**, 483–485 (2012). doi:10.1038/nmeth.1940.
- [113] Wong, J. H. & Hashimoto, T. Novel arabidopsis microtubule-associated proteins track growing microtubule plus ends. *BMC Plant Biology* **17** (2017). doi:10.1186/s12870-017-0987-5.
- [114] Benková, E. *et al.* Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* **115**, 591–602 (2003). doi:10.1016/S0092-8674(03)00924-3.
- [115] Burton, A. L., Williams, M., Lynch, J. P. & Brown, K. M. Rootscan: Software for high-throughput analysis of root anatomical traits. *Plant and Soil* **357**, 189–203 (2012). doi:10.1007/s11104-012-1138-2.
- [116] von Wangenheim, D. *et al.* Live tracking of moving samples in confocal microscopy for vertically grown roots. *eLife* **6**, e26792 (2017). doi:10.7554/eLife.26792.
- [117] Prusinkiewicz, P. & Runions, A. Computational models of plant development and form. *New Phytologist* **193**, 549–569 (2012). doi:10.1111/j.1469-8137.2011.04009.x. <https://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8137.2011.04009.x>.
- [118] Bozorg, B., Krupinski, P. & Jönsson, H. A continuous growth model for plant tissue. *Physical Biology* **13**, 065002 (2016).
- [119] Kennaway, R., Coen, E., Green, A. & Bangham, A. Generation of diverse biological forms through combinatorial interactions between tissue polarity and growth. *PLoS Computational Biology* **7**, e1002071 (2011). doi:10.1371/journal.pcbi.1002071.
- [120] Richardson, A., Rebocho, A. B. & Coen, E. Ectopic *KNOX* expression affects plant development by altering tissue cell polarity and identity. *The Plant Cell* **28**, 2079–2096 (2016). doi:10.1105/tpc.16.00284.
- [121] Sauret-Güeto, S., Schiessl, K., Bangham, A., Sablowski, R. & Coen, E. JAGGED controls arabidopsis petal growth and shape by interacting with a divergent polarity field. *PLoS Biology* **11**, e1001550 (2013). doi:10.1371/journal.pbio.1001550.

- [122] Fayant, P. *et al.* Finite element model of polar growth in pollen tubes. *The Plant Cell* **22**, 2579–2593 (2010). doi:10.1105/tpc.110.075754.
- [123] Cieslak, M. *et al.* Integrating physiology and architecture in models of fruit expansion. *Frontiers in Plant Science* **7** (2016). doi:10.3389/fpls.2016.01739.
- [124] Dupuy, L., Mackenzie, J., Rudge, T. & Haseloff, J. A system for modelling cell–cell interactions during plant morphogenesis. *Annals of Botany* **101**, 1255–1265 (2008). doi:10.1093/aob/mcm235.
- [125] Rudge, T. & Haseloff, J. A computational model of cellular morphogenesis in plants. In Capcarrère, M. S., Freitas, A. A., Bentley, P. J., Johnson, C. G. & Timmis, J. (eds.) *Advances in Artificial Life*, Lecture Notes in Computer Science, 78–87 (Springer Berlin Heidelberg, 2005).
- [126] Fozard, J. A., Bennett, M. J., King, J. R. & Jensen, O. E. Hybrid vertex-midline modelling of elongated plant organs. *Interface Focus* **6** (2016). doi:10.1098/rsfs.2016.0043.
- [127] Merks, R. M., Guravage, M., Inzé, D. & Beemster, G. T. Virtualleaf: An open-source framework for cell-based modeling of plant tissue growth and development. *Plant Physiology* **155**, 656–666 (2011). doi:10.1104/pp.110.167619. <http://www.plantphysiol.org/content/155/2/656.full.pdf>.
- [128] Brisson, E. Representing geometric structures in dimensions: Topology and order. *Discrete & Computational Geometry* **9**, 387–426 (1993). doi:10.1007/BF02189330.
- [129] van Opheusden, J. H. J. & Molenaar, J. Algorithm for a particle-based growth model for plant tissues. *Royal Society Open Science* **5**, 181127 (2018). doi:10.1098/rsos.181127.
- [130] Purswani, N. *Introducing the gemma of the liverwort Marchantia polymorpha L. as a simple morphogenetic system*. Ph.D. thesis, University of Cambridge (2014).
- [131] Pollak Williamson, B. *Frameworks for reprogramming lower plant systems*. Ph.D. thesis, University of Cambridge (2017).
- [132] Bowman, J. L. *et al.* Insights into land plant evolution garnered from the marchantia polymorpha genome. *Cell* **171**, 287–304.e15 (2017). doi:10.1016/j.cell.2017.09.030.
- [133] Flores-Sandoval, E., Eklund, D. M. & Bowman, J. L. A simple auxin transcriptional response system regulates multiple morphogenetic processes in the liverwort marchantia polymorpha. *PLOS Genetics* **11**, e1005207 (2015). doi:10.1371/journal.pgen.1005207.
- [134] Kato, H. *et al.* Auxin-mediated transcriptional system with a minimal set of components is critical for morphogenesis through the life cycle in marchantia polymorpha. *PLoS Genetics* **11** (2015). doi:10.1371/journal.pgen.1005084.

- [135] Ishizaki, K., Nishihama, R., Yamato, K. T. & Kohchi, T. Molecular genetic tools and techniques for marchantia polymorpha research. *Plant and Cell Physiology* **57**, 262–270 (2016). doi:10.1093/pcp/pcv097.
- [136] Nasu, M. *et al.* Efficient transformation of marchantia polymorpha that is haploid and has very small genome DNA. *Journal of Fermentation and Bioengineering* **84**, 519–523 (1997). doi:10.1016/S0922-338X(97)81904-6.
- [137] Ishizaki, K., Chiyoda, S., Yamato, K. T. & Kohchi, T. Agrobacterium-mediated transformation of the haploid liverwort marchantia polymorpha l., an emerging model for plant biology. *Plant and Cell Physiology* **49**, 1084–1091 (2008). doi:10.1093/pcp/pcn085.
- [138] Tsuboyama, S., Nonaka, S., Ezura, H. & Kodama, Y. Improved g-AgarTrap: A highly efficient transformation method for intact gemmalings of the liverwort marchantia polymorpha. *Scientific Reports* **8**, 10800 (2018). doi:10.1038/s41598-018-28947-0.
- [139] Tsuboyama, S. & Kodama, Y. AgarTrap: a simplified agrobacterium-mediated transformation method for sporelings of the liverwort marchantia polymorpha l. *Plant & Cell Physiology* **55**, 229–236 (2014). doi:10.1093/pcp/pct168.
- [140] Kubota, A., Ishizaki, K., Hosaka, M. & Kohchi, T. Efficient agrobacterium-mediated transformation of the liverwort marchantia polymorpha using regenerating thalli. *Bioscience, Biotechnology, and Biochemistry* **77**, 167–172 (2013). doi:10.1271/bbb.120700.
- [141] Tsuboyama-Tanaka, S. & Kodama, Y. AgarTrap-mediated genetic transformation using intact gemmae/gemmalings of the liverwort marchantia polymorpha l. *Journal of Plant Research* **128**, 337–344 (2015). doi:10.1007/s10265-014-0695-2.
- [142] Takenaka, M. *et al.* Direct transformation and plant regeneration of the haploid liverwort marchantia polymorpha l. *Transgenic Research* **9**, 179–185 (2000). doi:10.1023/A:1008963410465.
- [143] Chiyoda, S., Yamato, K. T. & Kohchi, T. Plastid transformation of sporelings and suspension-cultured cells from the liverwort marchantia polymorpha l. *Methods in Molecular Biology (Clifton, N.J.)* **1132**, 439–447 (2014). doi:10.1007/978-1-62703-995-6_30.
- [144] Chiyoda, S. *et al.* Simple and efficient plastid transformation system for the liverwort marchantia polymorpha l. suspension-culture cells. *Transgenic Research* **16**, 41–49 (2007). doi:10.1007/s11248-006-9027-1.
- [145] Sugano, S. S. *et al.* CRISPR/cas9-mediated targeted mutagenesis in the liverwort marchantia polymorpha l. *Plant & Cell Physiology* **55**, 475–481 (2014). doi:10.1093/pcp/pcu014.
- [146] Nishihama, R., Ishida, S., Urawa, H., Kamei, Y. & Kohchi, T. Conditional gene expression/deletion systems for marchantia polymorpha using its own heat-shock promoter and Cre/loxP-mediated site-specific recombination. *Plant and Cell Physiology* **57**, 271–280 (2016). doi:10.1093/pcp/pcv102.

- [147] Pollak, B. *et al.* Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytologist* **Accepted Author Manuscript** (2018). doi:10.1111/nph.15625.
- [148] Delmans, M. Supplementary files and information for the Ph.D. thesis of Mihails Delmans on Engineering morphogenesis of *Marchantia polymorpha gemmae*. <https://github.com/HaseloffLab/MihailsDelmansThesis>.
- [149] Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S. & Mullineaux, P. M. pGreen: a versatile and flexible binary ti vector for agrobacterium-mediated plant transformation. *Plant Molecular Biology* **42**, 819–832 (2000). doi:10.1023/A:1006496308160.
- [150] Delmans, M. Supplementary plasmid sequences for the Ph.D. thesis of Mihails Delmans on Engineering morphogenesis of *Marchantia polymorpha gemmae*. <https://benchling.com/mdelmans>.
- [151] Nagaya, S., Takemura, M. & Ohyama, K. Endogenous promoter, 5'-UTR and transcriptional terminator enhance transient gene expression in a liverwort, *Marchantia polymorpha* l. *Plant Biotechnology* **28**, 493–496 (2011). doi:10.5511/plantbiotechnology.11.1025a.
- [152] Goedhart, J. *et al.* Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature Communications* **3**, 751 (2012). doi:10.1038/ncomms1738.
- [153] Kremers, G.-J., Goedhart, J., van Munster, E. B. & Gadella, T. W. J. Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius. *Biochemistry* **45**, 6570–6580 (2006). doi:10.1021/bi0516273.
- [154] Cutler, S. R., Ehrhardt, D. W., Griffiths, J. S. & Somerville, C. R. Random GFP::cDNA fusions enable visualization of subcellular structures in cells of *Arabidopsis* at a high frequency. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 3718–3723 (2000).
- [155] Engler, C. *et al.* A golden gate modular cloning toolbox for plants. *ACS Synthetic Biology* **3**, 839–843 (2014). doi:10.1021/sb4001504.
- [156] Kurup, S. *et al.* Marking cell lineages in living tissues. *The Plant Journal* **42**, 444–453 (2005). doi:10.1111/j.1365-3113.2005.02386.x.
- [157] Boissard-Lorig, C. *et al.* Dynamic analyses of the expression of the HISTONE::YFP fusion protein in *Arabidopsis* show that syncytial endosperm is divided in mitotic domains. *The Plant Cell* **13**, 495–509 (2001). doi:10.1105/tpc.13.3.495.
- [158] Sarrion-Perdigones, A. *et al.* GoldenBraid 2.0: A comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiology* **162**, 1618–1631 (2013). doi:10.1104/pp.113.217661.

- [159] van Rossum, G. Python tutorial, technical report CS-R9526. Tech. Rep., Centrum voor Wiskunde en Informatica (CWI) (1995).
- [160] Lowekamp, B. C., Chen, D. T., Ibanez, L. & Blezek, D. The design of SimpleITK. *Frontiers in Neuroinformatics* **7** (2013). doi:10.3389/fninf.2013.00045.
- [161] Marstal, K., Berendsen, F., Staring, M. & Klein, S. SimpleElastix: A user-friendly, multi-lingual library for medical image registration. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 574–582 (2016).
- [162] Otsu, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9**, 62–66 (1979). doi:10.1109/TSMC.1979.4310076.
- [163] ECMAScript 2018 language specification (ECMA International, 2018). <<https://www.ecma-international.org>>.
- [164] Delmans, M. & Pollak, B. MarpoDB: An open registry of marchantia polymorpha genetic parts (HaseloffLab, 2016). <<https://github.com/HaseloffLab/MarpoDB>>.
- [165] PostgreSQL (2018). <<https://www.postgresql.org>>.
- [166] PostgreSQL database modeler (The PostgreSQL Global Development Group, 2018). <<https://pgmodeler.io/>>.
- [167] SQLAlchemy (2018). <https://www.sqlalchemy.org/>.
- [168] Chapman, B. & Chang, J. Biopython: Python tools for computational biology. *SIGBIO Newsl.* **20**, 15–19 (2000). doi:10.1145/360262.360268.
- [169] Miller, C. A., Anthony, J., Meyer, M. M. & Marth, G. Scribl: an HTML5 canvas-based graphics library for visualizing genomic data over the web. *Bioinformatics* **29**, 381–383 (2013). doi:10.1093/bioinformatics/bts677.
- [170] Lane, L. *et al.* neXtProt: a knowledge platform for human proteins. *Nucleic Acids Research* **40**, D76–D83 (2012). doi:10.1093/nar/gkr1179.
- [171] Rocha, Z. clipboard.js (2018). <https://clipboardjs.com/>.
- [172] Ronacher, A. Flask (2018). <http://flask.pocoo.org/>.
- [173] Chapman, J. A. *et al.* Meraculous: De novo genome assembly with short paired-end reads. *PLoS ONE* **6** (2011). doi:10.1371/journal.pone.0023501.
- [174] Chang, Z. *et al.* Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biology* **16**, 30 (2015). doi:10.1186/s13059-015-0596-2.

- [175] Bateman, A. *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **45**, D158–D169 (2017). doi:10.1093/nar/gkw1099.
- [176] Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990). doi:10.1016/S0022-2836(05)80360-2.
- [177] Li, H. & Durbin, R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754–1760 (2009). doi:10.1093/bioinformatics/btp324.
- [178] Picard (Broad Institute, 2018). <<http://broadinstitute.github.io/picard>>.
- [179] McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303 (2010). doi:10.1101/gr.107524.110.
- [180] De Wit, P. *et al.* The simple fool's guide to population genomics via RNA-seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources* **12**, 1058–1067 (2012). doi:10.1111/1755-0998.12003.
- [181] Logg, A., Wells, G. N. & Hake, J. *DOLFIN: a C++/Python Finite Element Library*, chap. 10 (Springer, 2012).
- [182] Logg, A., Mardal, K.-A., Wells, G. N. *et al.* *Automated Solution of Differential Equations by the Finite Element Method* (Springer, 2012).
- [183] Solly, J. E., Cunniffe, N. J. & Harrison, C. J. Regional growth rate differences specified by apical notch activities regulate liverwort thallus shape. *Current Biology* **27**, 16–26 (2017). doi:10.1016/j.cub.2016.10.056.
- [184] Arganda-Carreras, I. *et al.* Consistent and elastic registration of histological sections using vector-spline regularization. In Beichel, R. R. & Sonka, M. (eds.) *Computer Vision Approaches to Medical Image Analysis*, vol. 4241, 85–95 (Springer Berlin Heidelberg, 2006).
- [185] Barton, M. K. Twenty years on: The inner workings of the shoot apical meristem, a developmental dynamo. *Developmental Biology* **341**, 95–113 (2010). doi:10.1016/j.ydbio.2009.11.029.
- [186] Maravolo, N. C. & Voth, P. D. Morphogenic effects of three growth substances on marchantia gemmalings. *Botanical Gazette* **127**, 79–86 (1966).
- [187] Dickson, H. Polarity and the production of adventitious growing points in marchantia polymorpha. *Annals of Botany* **46**, 683–701 (1932).
- [188] Nishihama, R. *et al.* Phytochrome-mediated regulation of cell division and growth during regeneration and sporeling development in the liverwort marchantia polymorpha. *Journal of Plant Research* **128**, 407–421 (2015). doi:10.1007/s10265-015-0724-9.

- [189] Eklund, D. M. *et al.* Auxin produced by the indole-3-pyruvic acid pathway regulates development and gemmae dormancy in the liverwort *Marchantia polymorpha*. *The Plant Cell* **27**, 1650–1669 (2015). doi:10.1105/tpc.15.00065.
- [190] Aki, S. S. *et al.* Cytokinin Signaling Is Essential for Organ Formation in *Marchantia polymorpha*. *Plant and Cell Physiology* **60**, 1842–1854 (2019). doi:10.1093/pcp/pcz100. <http://oup.prod.sis.lan/pcp/article-pdf/60/8/1842/29103176/pcz100.pdf>.
- [191] Liu, W. & Stewart, C. N. Plant synthetic biology. *Trends in Plant Science* **20**, 309–317 (2015). doi:10.1016/j.tplants.2015.02.004.
- [192] Boehm, C. R., Pollak, B., Purswani, N., Patron, N. & Haseloff, J. Synthetic botany. *Cold Spring Harbor Perspectives in Biology* a023887 (2017). doi:10.1101/cshperspect.a023887.
- [193] Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLOS ONE* **3**, e3647 (2008). doi:10.1371/journal.pone.0003647.
- [194] Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* **6** (2011). doi:10.1371/journal.pone.0016765.
- [195] Sarrion-Perdigones, A. *et al.* GoldenBraid: An iterative cloning system for standardized assembly of reusable genetic modules. *PLOS ONE* **6**, e21622 (2011). doi:10.1371/journal.pone.0021622.
- [196] Alberts, B. *et al.* Analyzing protein structure and function. *Molecular Biology of the Cell*. 4th edition (2002).
- [197] The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi:10.1038/nature11247.
- [198] Pollak, B. & Delmans, M. marpodb.io. <http://marpodb.io>.
- [199] Eddy, S. R. Accelerated profile HMM searches. *PLOS Computational Biology* **7**, e1002195 (2011). doi:10.1371/journal.pcbi.1002195.
- [200] Pollak, B. Recode2s - DNA recoding for type IIS cloning. <https://github.com/bpollakw/recode2s>.
- [201] Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Research* **33**, D34–D38 (2005). doi:10.1093/nar/gki063.
- [202] Pollak, B., Delmans, M. & Haseloff, J. Analysis of Cambridge isolates of *Marchantia polymorpha*. *Supplementary material for Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome*, *Cell* **171**, 287–304 (2017).

- [203] MarpolBase (Marchantia Working Group and Genome Informatics Laboratory, 2018). <http://marchantia.info>.
- [204] Jin, J. *et al.* PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research* **45**, D1040–D1045 (2017). doi:10.1093/nar/gkw982.
- [205] Technologies, O. N. Advantages of long reads for genome assembly. *Oxford Nanopore Technologies* (2019).
- [206] Brunoud, G. *et al.* A novel sensor to map auxin response and distribution at high spatio-temporal resolution. *Nature* **482**, 103–106 (2012). doi:10.1038/nature10791.
- [207] Michniewicz, M., Brewer, P. B. & Friml, J. Polar auxin transport and asymmetric auxin distribution. *The Arabidopsis Book / American Society of Plant Biologists* **5** (2007). doi:10.1199/tab.0108.
- [208] Petrášek, J. & Friml, J. Auxin transport routes in plant development. *Development* **136**, 2675–2688 (2009). doi:10.1242/dev.030353.
- [209] HILL, A. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol (Lond)* **40**, 4–7 (1910).
- [210] Registry of standard biological parts (iGEM Foundation, 2018). <http://parts.igem.org>.
- [211] Peccoud, J. *et al.* Targeted development of registries of biological parts. *PLOS ONE* **3**, e2671 (2008). doi:10.1371/journal.pone.0002671.
- [212] van Berkel, K., de Boer, R. J., Scheres, B. & ten Tusscher, K. Polar auxin transport: models and mechanisms. *Development* **140**, 2253–2268 (2013). doi:10.1242/dev.079111. <https://dev.biologists.org/content/140/11/2253.full.pdf>.
- [213] Ulmasov, T., Murfett, J., Hagen, G. & Guilfoyle, T. J. Aux/iaa proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *The Plant Cell* **9**, 1963–1971 (1997). doi:10.1105/tpc.9.11.1963. <http://www.plantcell.org/content/9/11/1963.full.pdf>.
- [214] Liao, C.-Y. *et al.* Reporters for sensitive and quantitative measurement of auxin response. *Nature methods* **12**, 207–210 (2015). doi:10.1038/nmeth.3279. 25643149.
- [215] Glass, N. L. & Kosuge, T. Cloning of the gene for indoleacetic acid-lysine synthetase from *Pseudomonas syringae* subsp. *savastanoi*. *Journal of Bacteriology* **166**, 598–603 (1986).
- [216] McInnes, L. & Healy, J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv:1802.03426 [cs, stat]* (2018). 1802.03426.

- [217] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008). doi:10.1088/1742-5468/2008/10/P10008.0803.0476.
- [218] Kaul, K. & Sabharwal, P. S. Effects of sucrose and kinetin on growth and chlorophyll synthesis in tobacco tissue cultures. *Plant Physiology* **47**, 691-695 (1971). doi:10.1104/pp.47.5.691. <http://www.plantphysiol.org/content/47/5/691.full.pdf>.
- [219] Okazawa, Y., Katsura, N. & Tagawa, T. Effects of auxin and kinetin on the development and differentiation of potato tissue cultured in vitro. *Physiologia Plantarum* **20**, 862-869 (1967). doi:10.1111/j.1399-3054.1967.tb08373.x. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1399-3054.1967.tb08373.x>.
- [220] Wan, Y., Sorensen, E. L. & Liang, G. H. The effects of kinetin on callus characters in alfalfa (*medicago sativa* l.). *Euphytica* **39**, 249-254 (1988). doi:10.1007/BF00037103.

Appendices

Appendix A

Marchantia protocols

Protocol 1 Growth media**1.1 1/2 Gamborg B5****Materials**

- 0.158 % (w/v) Gamborg B5 medium including vitamins (Duchefa)
- 1M NaOH
- RO water
- Stock of antibiotic(s) (optional)
- 1 – 3 % (w/v) agar (Sigma) for solid medium

Procedure

1. Mix Gamborg B5 medium in water using magnetic stirrer.
2. Optionally add agar to the mixture for the solid medium.
3. Adjust pH with NaOH to 5.8.
4. Autoclave in a glass bottle.
5. Add 1 : 1000 dilution of the required stock antibiotic (optional).
6. For solid medium, pour into 100 × 15 mm petri dishes, ~ 20 mL per plate, or store at room temperature.

1.2 Antibiotics

Antibiotic	[Stock]	[Working]	Solvent
hygromycin	20 mg/mL	20 µg/mL	Water
spectinomycin	100 mg/mL	100 µg/mL	Water
rifampicilin	10 mg/mL	10 µg/mL	DMSO
tetracyclin	5 mg/mL	5 µg/mL	Ethanol
carbenicillin	50 mg/mL	50 µg/mL	Water
kanamycin	50 mg/mL	50 µg/mL	Water
cefotaxime	100 mg/mL	100 µg/mL	Water

1.3 1/2 Gamborg B5 Supplemented**Materials**

- 0.158 % (w/v) Gamborg B5 medium including vitamins (Duchefa)
- 1M NaOH
- RO water
- 5 % (w/v) sucrose (Fisher Scientific)
- 0.1 % (w/v) N-Z amine A (Sigma-Aldrich)
- 0.03 % (w/v) L-glutamine (Sigma-Aldrich)

Procedure

1. Mix Gamborg B5 medium in water using magnetic stirrer.
2. Add sucrose, N-Z amine A, L-glutamine.
3. Autoclave in a glass bottle.

1.4 LB medium**Materials**

- 20 g/L LB broth Lennox (Formedium)
- RO water
- 1 % (w/v) agar (Sigma)
- Antibiotics stock solution (optional)

Protocol

1. Pour water into an autoclave bottle.
2. Add LB broth.
3. Optionally add agar to the mixture for the solid medium.
4. Autoclave the broth.
5. Add 1 : 1000 dilution of the required stock antibiotic (optional).
6. For solid medium, pour into 100 × 15 mm petri dishes, ~ 20 mL per plate, or store at room temperature.

Protocol 2 Preparation of the Marchantia spores

2.1 Sterilisation of Marchantia spores
Materials

- Dried Marchantia spore-heads
- Milton Mini Sterilising Tablets 3g (Laboratoire Rivadis), active component Troclosen sodium 19.5 % (w/v)
- Sterile RO water
- 100 × 15 mm 1 % (w/v) agar 1/2 Gamborg B5 medium plate

Procedure

1. Prepare Milton solution by dissolving 1 tablet in 25 mL of sterile RO water.
 2. Crush dried Marchantia spore-heads in a 1.5 mL Eppendorf tube using metallic tweezers.
 3. Add 1 mL of the Milton solution to the Eppendorf tube with the crushed spore-heads and filter the spores into a 50 mL Falcon tube through a 40 µL cell strainer (greiner bio-one EASYstrainer™).
 4. Dilute the spores with the Milton solution, so that the total volume in mL is equal to the number of the spore-heads used. Aliquot the spores into 1.5 mL Eppendorf tubes, 1 mL per tube and let them sterilise for 20 min.
 5. Spin the tubes for 1 min at 13000 g, remove supernatant and resuspend the spores in sterile water, 1 mL per tube.
 6. Repeat the previous step.
 7. To plate the spores, transfer 1 mL of the suspension to the 1 % (w/v) agar 1/2 Gamborg B5 medium plate, let the water to evaporate, seal the plate with 1.25 cm 3M Micropore tape.
-

2.2 Collection of Marchantia spores from agar plates
Materials

- Marchantia spores grown on solid medium plate
- Liquid medium of choice / sterile water

Procedure

1. Grow Marchantia spores in a Petri dish in an inverted orientation.
2. Remove the lid of the plate and add 1 – 5 mL of the liquid medium of choice / water on top of the solid medium.
3. Detach the spores from the solid medium by gently scratching the surface of the agar with a plastic spreader.
4. Use a wide-bore pipet to transfer the spores into a 50 mL Falcon tube.
5. If necessary, add more liquid into the plate and repeat the transfer.

Protocol 3 Bacterial transformation

3.1 Preparation of electrocompetent *Agrobacterium* cells**Materials**

- Sterile LB medium
- 10% Glycerol chilled on ice

Procedure

1. Inoculate 3 mL of liquid LB with *Agrobacterium* seed stock in 14 mL round-bottom Falcon tube and incubate overnight at 28° C shaking at 140 rpm.
2. Inoculate 300 mL of liquid LB with the overnight culture in a sterile 500 mL conical flask. Incubate for ~ 9 h at 28° C shaking at 140 rpm.
3. Periodically check the optical density of the culture until the OD 600 reading reaches 0.5.
4. Aliquot the culture into six 50 mL round bottom Falcon tubes and centrifuge at 4° C, 4000 g for 5 min
5. Discard supernatant and resuspend the pellets in 25 mL of glycerol. Repeat the centrifugation.
6. Discard supernatant, resuspend the pellets in 4 mL of glycerol, combine and aliquot the cells into the two tubes. Repeat the centrifugation.
7. Discard supernatant, resuspend the pellets in 2 mL of glycerol, combine the cells into one tube. Repeat the centrifugation.
8. Discard supernatant, resuspend the pellet in 2 mL of glycerol.
9. Aliquot the cells into 1.5 mL Eppendorf tubes, 50 µL each. Store at -80° C.

3.2 *Agrobacterium* transformation by electroporation**Materials**

- 50 µL of electrocompetent GV2260 *Agrobacterium* cells
- Liquid LB medium
- LB plate with added rifampicin, carbenicillin and tetracycline + antibiotic for the plasmid selection

Procedure

1. Thaw tubes with electrocompetent *Agrobacterium* cells on ice.
2. Add 20 ng of DNA and transfer the cells to the 2 mm electroporation cuvette (VWR). Keep the cuvette on ice.
3. Apply a 2.5 kV 5.5 ms pulse, add 1 mL of liquid LB medium to the cuvette, transfer the cells into a 1.5 mL Eppendorf tube and cool on ice.
4. Incubate the cells at 28° C incubator shaking at 120 rpm for 2 h.
5. Inoculate the LB plate with 100 µL of the cells and incubate upside-down at 28° C for two days.

3.3 Heat-shock transformation of *E. coli* cells**Materials**

- TOP10 chemically-competent *E. coli* cells (Thermo Fisher Scientific)
- Assembly reaction
- Sterile LB medium
- LB plates with antibiotics for selection

Procedure

1. Thaw 25 µL of the cells on ice.
 2. Add 1 µL of the assembly reaction.
 3. Incubate the cells at 42° C water bath for 45 sec, add 250 µL of the LB medium and cool on ice.
 4. Incubate the cells at 37° C incubator, shaking at 120 rpm for 1 h.
 5. Inoculate the LB plate with 100 µL of the cells and incubate them upside-down at 37° C overnight.
-

Protocol 4 Amplification from genomic DNA

CTAB extraction buffer

Mix the following materials to a final volume of 10 mL per one extraction:

- 20 mM EDTA (Sigma)
- 2 % (w/v) PVP-40 (Sigma-Aldrich)
- 1.42 M NaCl (Fisher Scientific)
- 100 mM Tris-HCl pH 8.0 (Melford Laboratories)
- 2 % (w/v) CTAB (Acros Organics)
- 10 mM β -mercaptoethanol (Merck)
- 1 mg/mL ascorbic acid (Sigma-Aldrich)

Phusion Master Mix

For 20 reactions mix the following materials in the order they are listed in a 1.5 mL Eppendorf tube:

- 268 μ L HPLC grade water
- 80 μ L 5x HF Buffer
- 8 μ L dNTPs
- 4 μ L Phusion DNA Polymerase

4.1 PCR Amplification
Materials

- 2 μ L of 10 μ M forward and 2 μ L of 10 μ M reverse primer
- 1 μ L of extracted genomic DNA
- 35 μ L Phusion master mix

Procedure

1. Mix the materials in the order listed in a PCR tube, one tube per reaction.
2. Load the tubes into a thermal cycler and run using the following program:

Step	Temperature	Time
1x	98 °C	30 sec
35x	98 °C	10 sec
	T_m °C	20 sec
	72 °C	30 sec per kb
Hold	4 °C	

T_m °C - lowest melting temperature of the two primers.

4.2 Genomic DNA extraction from Marchantia
Materials

- CTAB extraction buffer
- 100 mg/mL RNase A (Qiagen)
- Chloroform
- Isoamyl alcohol
- isopropanol (Acros Organic)
- 70 % ethanol (Fisher Scientific)

Procedure

1. Freeze two grams of 3 week-old Marchantia gemmae with liquid nitrogen and grind into a fine powder using mortar and pestle. Collect the powder into 50 mL Falcon tube.
2. Add 10 mL of CTAB heated to 65° C into the tube with the tissue powder and vortex for 30 sec.
3. Add 10 μ L of RNase A to the mixture and incubate at 65° C for 15 min, vortexing for 30 sec every 5 min.
4. Combine 0.4 mL of isoamyl alcohol with 9.6 mL of chloroform and add to the mixture. Mix the contents of the tube by gentle shaking for 30 sec.
5. Centrifuge the mixture at 4500 g for 10 min. Collect the upper layer into a new 50 mL Falcon tube.
6. Add 0.7 volumes of isopropanol, and mix by inverting the tube four times. Centrifuge at 4500 g for 10 min and discard the supernatant.
7. Resuspend the pellet in 1 mL of 70 % ethanol and transfer to a 1.5 mL Eppendorf tube. Centrifuge at 13000 g for 10 min.
8. Discard the supernatant, resuspend 1 mL of 70 % ethanol and centrifuge at 13000 g for 10 min. Make this step twice.
9. Discard supernatant, let the pellet dry for 10 min and resuspend in 50 μ L of TRIS-EDTA buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA).
10. Process the sample using the PowerClean Pro DNA Cleanup Kit (MO BIO Laboratories).

Protocol 5 DNA assembly**5x isothermal buffer**

Mix the following materials in HPLC water:

- 25 %*(w/v)* PEG-8000 (Sigma-Aldrich)
- 500 *mM* Tris-HCl pH 7.5 (Sigma-Aldrich)
- 50 *mM* MgCl₂ (Fisher Scientific)
- 50 *mM* dithiothreitol (Thermo Fisher Scientific)
- 1 *mM* dNTPs (Bioline)
- 5 *mM* NAD⁺ (New England Biolabs)

1.33x Gibson assembly master mix

Mix the following components in a 1.5 *mL* Eppendorf tube:

- 320 μ L 5x isothermal buffer
- 0.64 μ L T5 exonuclease, 10 *U*/ μ L (New England Biolabs)
- 20 μ L Phusion DNA Polymerase, 2 *U*/ μ L (Thermo Fisher Scientific)
- 160 μ L Taq DNA Ligase, 2 *U*/ μ L (New England Biolabs)
- 860 μ L RO water

5.1 Gibson assembly**Materials**

- 1.33x Gibson assembly master mix
- Linear DNA fragments with 20 – 40 *bp* overlapping overhangs

Procedure

1. Mix equimolar amounts of the linear DNA fragments in a PCR tube to total volume of 1 μ L. Add 3 μ L of the Gibson assembly master mix and mix by pipetting.
2. Transfer the tube to a thermocycler preheated to 50° C immediately after adding the master mix.
3. Incubate at 50° C for 1 *h*.

5.2 Loop assembly master mix

Mix the following materials in a 1.5 *mL* Eppendorf tube per one reaction:

- 2 μ L HPLC water
- 1 μ L T4 DNA ligase buffer (NEB)
- 0.5 μ L of 1 *mg/mL* purified Bovine Serum Albumin (NEB)
- 0.25 μ L T4 DNA Ligase, 400 *U*/ μ L (NEB)
- 0.25 μ L BsaI (NEB) for odd levels, SapI (NEB) for even levels

Loop assembly**Materials per reaction**

- 4 μ L of Loop assembly master mix
- 1 μ L of the receiver plasmid (7.5 *f**mol*/ μ L)
- 1 μ L of each of the DNA parts (15 *f**mol*/ μ L)

Procedure

1. Mix DNA parts and the receiver plasmid in a PCR tube. Add HPLC water to total volume of 6 μ L.
2. Add 4 μ L of the master mix into the PCR tube. Mix by pipetting.
3. Transfer the samples into a thermocycler and run the following program:

Step	Temperature	Time
26x	37° C	3 <i>min</i>
	16° C	4 <i>min</i>
1x	50° C	5 <i>min</i>
	80° C	10 <i>min</i>

Appendix B

Effect of kinetin on *Marchantia gemma* growth and regeneration

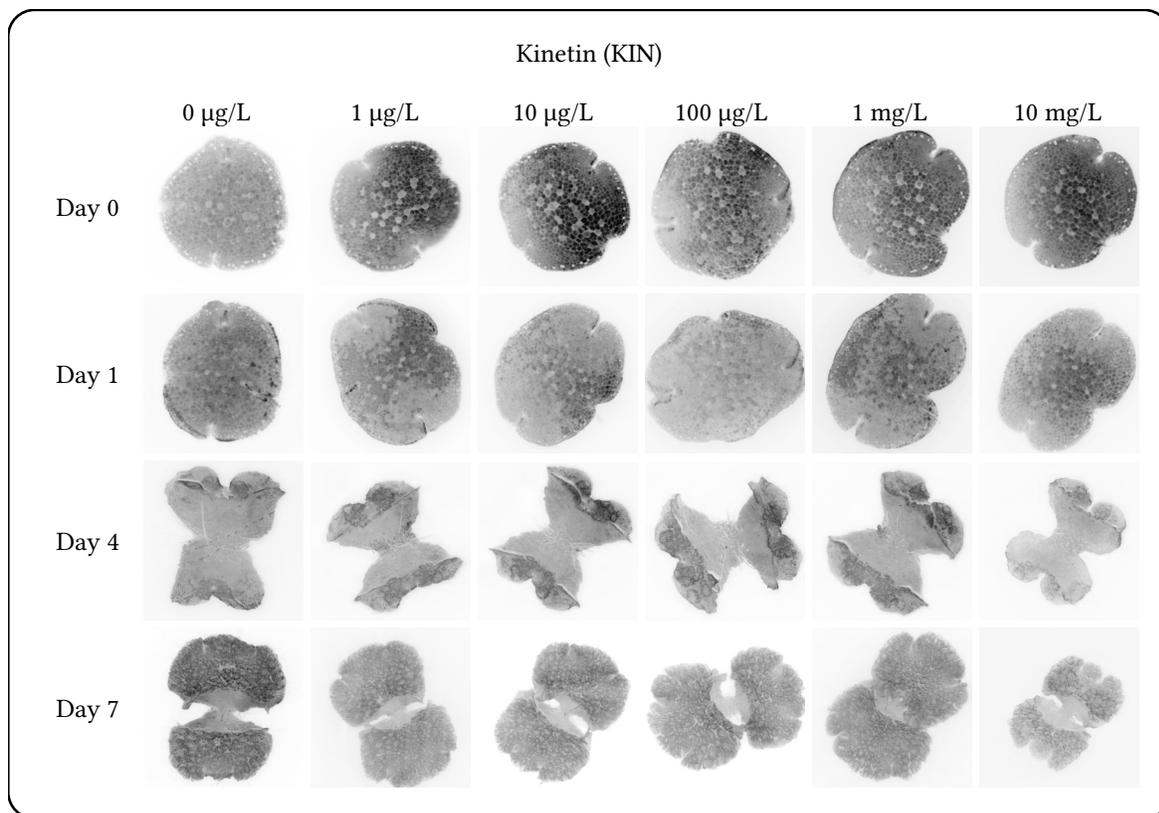


Figure B.1: **Dose-response to kinetin.** The images presented show a characteristic response ($n = 3$ for each concentration) of gemmae to various concentrations of kinetin. The maximal concentration in the assay was chosen to exceed a typical concentration of kinetin ($0.5 - 8 \text{ mg/L}$) in tissue cultures based on the earlier studies^[218-220].

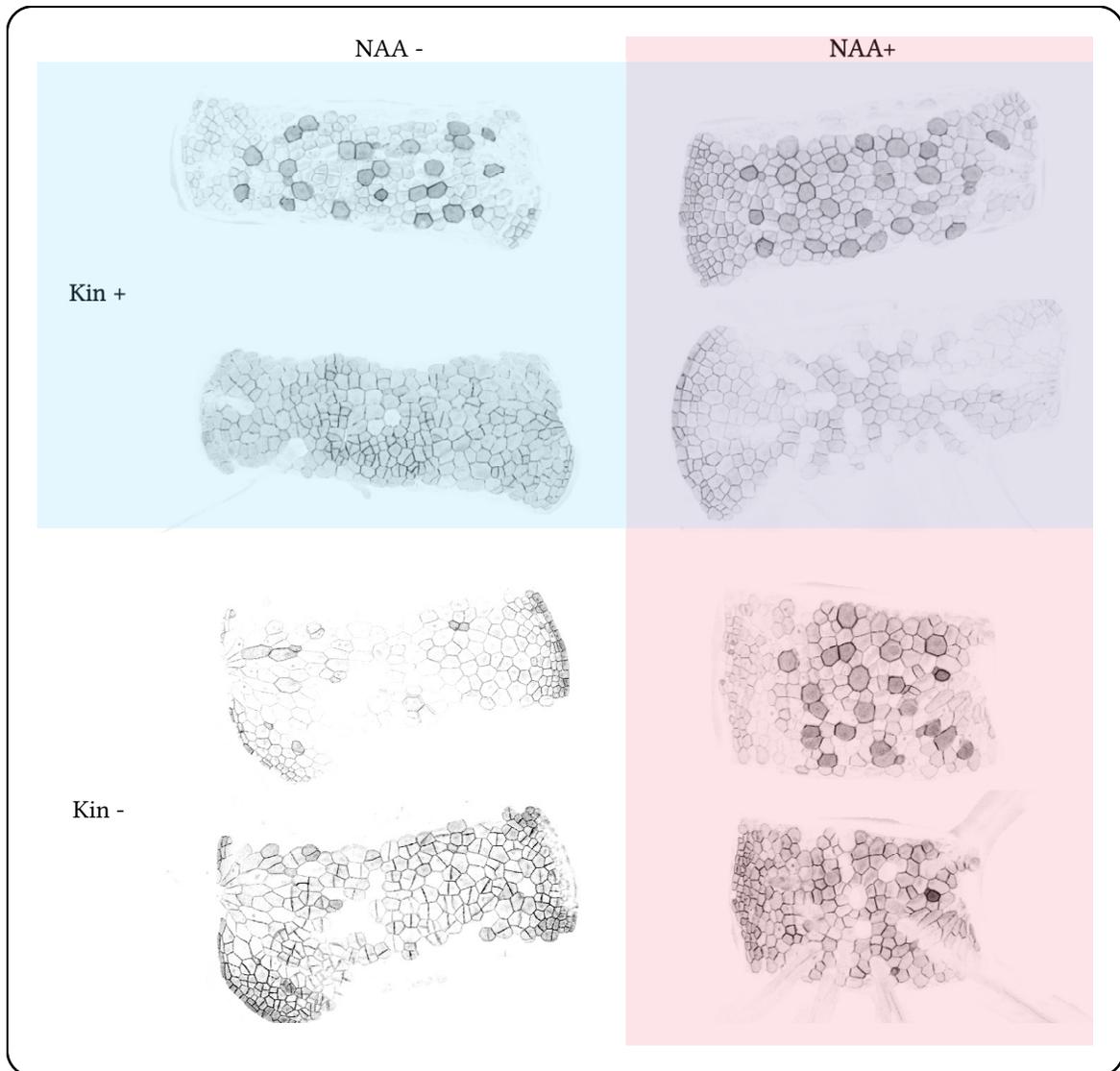


Figure B.2: **Effect of kinetin and auxin on regeneration of *Marchantia* gemma.** Microscopy images of the excised central region of gemma at day 0 (top image for each quadrant) and day 2 (bottom image for each quadrant) supplemented with combination of synthetic auxin ($100 \mu\text{g}/\text{L}$ NAA, highlighted in blue) and synthetic cytokinin ($10 \text{mg}/\text{L}$ kinetin, highlighted in red). The experiment ($n = 4$ for each out of four treatments) suggests independence of auxin-repressed regeneration from application of cytokinin.

Appendix C

Continuous model of pressure-driven growth

Pressure-driven tissue expansion

Assuming a viscous model of growth, the velocity (\mathbf{v}) of material points of a tissue are proportional to the applied force (\mathbf{f}):

$$\mathbf{v} = \phi \mathbf{f}$$

where ϕ is tissue extensibility. Excluding the external forces acting on the tissue, the force \mathbf{f} results from the gradient of the hydrostatic pressure between the cells:

$$\mathbf{f} = -\nabla p$$

Then, according to the model 3.1, the volumetric growth rate (γ) is equal to

$$\gamma = \nabla \cdot \mathbf{v} = -\phi \nabla^2 p \quad (1)$$

Water transport model

The movement of the water through a cellular tissue can be modelled by a flow through a porous medium, governed by Darcy's law:

$$\mathbf{u} = -K_w \nabla P \quad (2)$$

where \mathbf{u} is water velocity, K_w is water permeability and P is water potential equal to the difference between the

Protocol 6 Continuous model of pressure-driven tissue expansion

osmotic pressure (Π) and hydrostatic pressure (p):

$$P = \Pi - p \quad (3)$$

Expansion due to water accumulation

Assuming that the volume of a cell is approximately equal to the volume of the water inside the cell, the volumetric growth rate must be equal to the rate of water accumulation, i.e.:

$$\gamma = \nabla \cdot \mathbf{v} = -\nabla \cdot \mathbf{u}$$

Substituting equations (2) and (3) results in

$$\gamma = K_w \nabla^2 p - K_w \nabla^2 \Pi$$

Eliminating the hydrostatic pressure term from the above equation by substituting equation (1) gives an equation of the osmotic pressure-driven tissue expansion:

$$\gamma = -\frac{K_w \phi}{K_w + \phi} \nabla^2 \Pi$$