



HAL
open science

Toward a multi-scale understanding of flower development - from auxin networks to dynamic cellular patterns

Jonathan Legrand

► **To cite this version:**

Jonathan Legrand. Toward a multi-scale understanding of flower development - from auxin networks to dynamic cellular patterns. *Development Biology*. Ecole normale supérieure de lyon - ENS LYON, 2014. English. NNT: 2014ENSL0947. tel-01191471

HAL Id: tel-01191471

<https://theses.hal.science/tel-01191471>

Submitted on 4 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



ÉCOLE NORMALE SUPÉRIEURE DE LYON

Thèse soutenue en vue de l'obtention du grade de
**Docteur de l'Université de Lyon, délivré par l'École
Normale Supérieure de Lyon.**

Toward a multi-scale understanding of flower development: from auxin networks to dynamic cellular patterns.

Présentée et soutenue publiquement le 7 Novembre 2014
par Monsieur Jonathan LEGRAND

Discipline: Sciences de la vie.

Laboratoire: Reproduction et Développement des Plantes.

École Doctorale: Biologie Moléculaire Intégrative et Cellulaire.

Directeur de thèse:

Pr. Arezki BOUDAUD, ENS Lyon

Co-directeur de thèse:

Dr. Yann GUÉDON, CIRAD Montpellier

Devant la commission d'examen formée de :

Dr. Dorota KWIATKOWSKA, Université de Silesia (Pologne), Rapporteur

Dr. Marie-Laure MARTIN-MAGNIETTE, AgroParisTech/INRA, Rapporteur

Dr. Grégoire MALANDAIN, INRIA Sophia Antipolis, Examineur

Dr. Gaël YVERT, ENS Lyon, Examineur

Pr. Arezki BOUDAUD, ENS Lyon, Examineur

Dr. Yann GUÉDON, CIRAD Montpellier, Examineur

Résumé:

Dans le domaine de la biologie développementale, un des principaux défis est de comprendre comment des tissus multicellulaires, à l'origine indifférenciés, peuvent engendrer des formes aussi complexes que celles d'une fleur.

De part son implication dans l'organogenèse florale, l'auxine est une phytohormone majeure. Nous avons donc déterminé son réseau binaire potentiel, puis y avons appliqué des modèles de clustering de graphes s'appuyant sur les profils de connexion présentés par ces 52 facteurs de transcription (FT). Nous avons ainsi pu identifier trois groupes, proches des groupes biologiques putatifs: les facteurs de réponse à l'auxine activateurs (ARF+), répresseurs (ARF-) et les Aux/IAAs. Nous avons détecté l'auto-interaction des ARF+ et des Aux/IAA, ainsi que leur interaction, alors que les ARF- en présentent un nombre restreint. Ainsi, nous proposons un mode de compétition auxine indépendant entre ARF+ et ARF- pour la régulation transcriptionnelle.

Deuxièmement, nous avons modélisé l'influence des séquences de dimérisation des FT sur la structure de l'interactome en utilisant des modèles de mélange Gaussien pour graphes aléatoires. Les groupes obtenus sont proches des précédents, et les paramètres estimés nous ont conduit à conclure que chaque sous-domaine peut jouer un rôle différent en fonction de leur proximité phylogénétique.

Enfin, nous sommes passés à l'échelle multi-cellulaire ou, par un graphe spatio-temporel, nous avons modélisé les premiers stades du développement floral d'*A. thaliana*. Nous avons pu extraire des caractéristiques cellulaires (3D+t) de reconstruction d'imagerie confocale, et avons démontré la possibilité de caractériser l'identité cellulaire en utilisant des méthodes de classification hiérarchique et des arbres de Markov cachés.

Abstract:

A striking aspect of flowering plants is that, although they seem to display a great diversity of size and shape, they are made of the same basic constituents, that is the cells. The major challenge is then to understand how multicellular tissues, originally undifferentiated, can give rise to such complex shapes.

We first investigated the uncharacterised signalling network of auxin since it is a major phytohormone involved in flower organogenesis. We started by determining the potential binary network, then applied model-based graph clustering methods relying on connectivity profiles. We demonstrated that it could be summarised in three groups, closely related to putative biological groups. The characterisation of the network function was made using ordinary differential equation modelling, which was later confirmed by experimental observations.

In a second time, we modelled the influence of the protein dimerisation sequences on the auxin interactome structure using a mixture of linear models for random graphs. This model led us to conclude that these groups behave differently, depending on their dimerisation sequence similarities, and that each dimerisation domain might play different roles.

Finally, we changed scale to represent the observed early stages of *A. thaliana* flower development as a spatio-temporal property graph. Using recent improvements in imaging techniques, we could extract 3D+t cellular features, and demonstrated the possibility of identifying and characterising cellular identity on this basis. In that respect, hierarchical clustering methods and hidden Markov tree have proven successful in grouping cells depending on their feature similarities.

Mots-clefs:

Arabidopsis thaliana; biologie du développement; signalisation auxinique; morphogenèse florale; réseaux de facteurs de transcription; graphes spatio-temporel avec attributs; détection de patrons cellulaire; partitionnement non-supervisé; attributs cellulaires; imagerie sur tissu vivant.

Keywords:

Arabidopsis thaliana; developmental biology; auxin signalling pathways; floral morphogenesis; transcription factors network; attributed spatio-temporal graphs; cellular patterns detection; unsupervised clustering; cellular features; tissue live-imaging.

Table of contents

Résumé	i
Abstract	ii
Mots-clefs	iii
Keywords	iii
Table of contents	viii
General introduction	1
State of the art	9
1 The Shoot Apical Meristem: a self-maintained structure initiating new organs like flowers.	9
1.1 Structural and functional zonation of the SAM.	9
1.2 Main regulators of the SAM.	10
2 Auxin control organ initiation and patterning.	12
2.1 Auxin levels are spatially controlled by efflux and influx carriers.	12
2.2 Polar auxin transport regulation or self-organisation?	13
3 Auxin signal transduction is ensured by a TF network.	16
3.1 ARFs - Aux/IAA molecular and functional structures.	16
3.2 Auxin transduction pathway model.	17
3.3 Auxin perception is spatially regulated.	18
4 Stereotypic shape of flowers	19
4.1 Morphological description of early flower development.	20
4.2 Induction of floral meristem fate.	20
4.3 Floral meristem patterning by homeotic genes: the ABC model.	21
4.4 Structural and functional zonation of the early FM.	23
4.5 Regulation of flower size.	24
5 Fluorescence live-imaging enable genes, shape and growth quantification at cellular scale.	26
5.1 From staining to labelling.	26
5.2 What to label ?	27
5.3 A 5D universe to compare -omic tools and live-imaging.	29
5.4 Examples of live-imaging in animals morphogenesis.	31
5.5 Imaging cell geometry to quantify shape and growth at cellular scale.	32
5.6 Image processing: from intensity-based to object-based representations.	33
Conclusions	36
1 The auxin signalling network translates dynamic input into robust patterning at the shoot apex	39
Abstract	40
1 Introduction	41

2	Result and Discussion	43
2.1	Auxin signalling is spatially regulated in the shoot apical meristem	43
2.2	The Aux/IAA-ARF interaction network has a simple structure	46
2.3	The structure of the Aux/IAA-ARF network provides a plausible model for auxin signalling	48
2.4	The auxin signalling model allows predicting differential sensitivity to auxin and buffering capacities for the Aux/IAA-ARF pathway in the meristem	51
2.5	A novel signalling sensor, DII-VENUS, reports on the auxin signalling input in the meristem	53
2.6	Auxin signalling sensors distribution and dynamics confirm the model predictions	55
3	Discussion	56
4	Materials and methods	60
4.1	Plant material, growth conditions and plant treatments	60
4.2	Analysis of gene expression	60
4.3	Confocal microscopy and live imaging	61
4.4	Generation of plasmids and transgenic plants	61
4.5	<i>In vitro</i> transcription/translation of TIR1/AFB tagged proteins, immunoprecipitations and pull-down assays	62
4.6	Protein interaction analyses	63
4.7	Cluster analysis of Aux/IAA-ARF interaction network	65
4.7.1	Construction of the interaction network on the basis of the yeast-2-hybrid data	65
4.7.2	Cluster analysis	65
4.7.3	Analysis of the adequacy of the clustering	66
4.7.4	Graphical representation of proteins preserving pairwise distances between them	67
4.8	Modelling of the Aux/IAA-ARF signalling pathway	68
4.8.1	Model description	68
4.8.2	Numerical simulations	69
	Acknowledgements	70
	Author contributions	70
	Supplementary Materials	72

2 Modelling the influence of phylogenetic distances between dimerisation sequences on the auxin signalling interactome 93

	Abstract	94
1	Introduction	95
2	Result and Discussion	96
2.1	Y2H experiment and data	97
2.2	Building a binary network from Y2H data	98
2.2.1	Determination of interaction threshold per biological test	98
2.2.2	Defining the decision rules	99
2.3	Building a valued network from Y2H data	99
2.4	Network analysis using Bernoulli and Gaussian mixture models	100
2.4.1	Binary or valued graph and associated models	101
2.4.2	Comparing model selection criterion and clusters composition returned by the BM and GM models	101

	2.4.3	Conclusions on the presented models	104
2.5		Influence of the phylogenetic distance between dimerisation sequences on the interaction network	105
	2.5.1	Building the phylogenetic distance matrix	105
	2.5.2	Introduction of phylogenetic distances as an explanatory variable	106
	2.5.3	Interpretation of the patterns found within the auxin interactome	110
3		Conclusions	111
4		Methods	113
	4.1	Construction of the Aux/IAA-ARF interaction network on the basis of the yeast-2-hybrid data.	113
	4.1.1	Yeast-2-hybrid protein interaction testing:	113
	4.1.2	Data description:	113
	4.2	Network binarisation.	114
	4.2.1	Mixture model over OD ratios under the X-Gal ‘+’ mark:	114
	4.2.2	Decision rules:	114
	4.3	Phylogenetic data	115
	4.4	Response distance matrix: standardized distances between transcriptional factors	116
	4.5	Phylogenetic distance matrix	117
	4.6	Assessing the adequacy of the clustering	117
	4.7	Availability of supporting data	118
	4.7.1	Original Yeast-2-Hybrid data for Aux/IAA - ARFs interaction tests	118
	4.7.2	Aux/IAA - ARFs protein sequences	118
	4.7.3	Domains III and IV sub-sequences	119
		Author contributions	119
		Supplementary Materials	120
3		An attributed spatio-temporal graph framework for characterizing floral morphogenetic patterns in <i>Arabidopsis thaliana</i>	131
		Abstract	132
1		Introduction	133
	1.1	Background and requirements	135
	1.1.1	Morphological description of the early FM development	135
	1.1.2	Quick overview of segmentation and lineage tools.	136
	1.1.3	Obtaining the segmented and lineaged data.	137
	1.2	Relation to previous works	137
	1.3	Principal contribution of our methods	139
2		Methods	139
	2.1	Spatio-temporal graph provides a natural framework for 3D+t data.	140
	2.2	Quantifying cellular features of a 3D segmented tissue and enabling graph analysis.	143
	2.2.1	Cellular features data	143
	2.2.2	Spatial differentiation functions	152
	2.2.3	Temporal differentiation functions	153
	2.2.4	Features transformation functions	156
	2.3	Adaptating the clustering framework to spatio-temporal property data.	162

2.3.1	Creating a standardised pairwise distance matrix.	162
2.3.2	Data standardisation.	164
2.3.3	Re-weighting and handling missing data.	165
2.3.4	Clustering a pairwise distance matrix.	166
2.3.5	Managing outliers for the standardisation step.	167
2.3.6	Clusters assessment and inspection.	168
2.3.7	Clusterings comparison with bipartite matching of cluster composition.	171
2.3.8	Supplementary function definitions.	171
3	Results	172
3.1	Exploratory analysis of the temporal property graph	172
3.1.1	Taking the “universal rule for symmetric division” presented in Besson and Dumais [2011] to the third dimension	172
3.1.2	Mathematical relevance of user defined regions.	177
3.2	Cellular pattern identification on the basis of geometrical features cluster- ing	180
3.2.1	Relating genetic spatio-temporal patterns to quantitative fea- ture changes by clustering AHP6 signal.	180
3.2.2	Univariate clusterings on the basis of geometrical features.	183
3.2.3	Multi-variate clusterings analysis.	187
3.3	Cellular pattern identification on the basis of Markov tree analysis of the lineage and geometrical features	192
3.3.1	Definition of hidden multi-type branching processes	193
3.3.2	Multivariate hidden multi-type branching process model out- puts.	194
4	Discussion and perspectives	198
	Author contributions	205
	Supplementary Materials	206
	General perspectives	217
	Bibliography	234
	List of figures	246
	List of tables	248

General introduction

Developmental biology is the study of the processes by which multi-cellular organisms grow and develop. Therefore organogenesis and morphogenesis are the two main processes studied. Over a large variety of organisms studied, it has been repeatedly demonstrated that shaping a tissue or an organ requires multiple regulatory mechanisms such as hormonal signalling, homeotic/positional genes, cell-cell communication and mechanical cues. These diverse regulatory systems often present redundancies, multiple feedback loops and levels of controls to ensure the robustness of the shaping process and its outcome. Being fixed in their location by a root system and pursuing most of their development after embryonic development, plants are totally dependant on their faculty to adapt and respond to environmental cues. By continuously maintaining stem cells at the Shoot Apical Meristem (SAM) and Root Apical Meristem (RAM), plants possess the ability to initiate and produce different organs. This faculty of continuous organogenesis during their whole life is an important difference between plants and animals. For animals, organogenesis indeed takes place during embryogenesis and post-embryonic development then follows the previously established plan, rarely presenting *de novo* organ formation (except for animals presenting a larval stage) as seen in plants.

Understanding morphogenesis, the coordinated cell behaviour shaping an organism, thus requires uncovering, piece by piece, the complex mechanisms at work and their various levels of interplay. Historically, the reductionist paradigm has prevailed in biology whereby a complex system is the sum of its parts and understanding each fundamental parts and their relations is sufficient to explain causality. However, considering the issued questions in developmental biology, it has been shown that this paradigm fails to explain some high-order processes, notably emergent properties undefined in each fundamental processes or

relations. These problems thus require an holistic approach, often called *Systems Biology*, which presents a shift in paradigm where the goal is to understand how properties of the system emerge by taking into account the pluralism of both cause and effect into the model, plus the intertwined complexities in between. One way to achieve this is through the quantitative measurement of the multiple processes under study and a rigorous mathematical formalisation, leading to a comprehensive view of the problem and the identification of patterns and emergent properties.

In *Arabidopsis thaliana*, as for in angiosperms, the male and female reproductive organs, stamens and carpels are set up during flower development. It is known that organogenesis initiation is controlled by a spatio-temporal pattern of phytohormones, that lead to the definition of a group of founder cells (initium), which will later grow into the lateral organ. The local accumulation of on such hormone, auxin, by polar transporters triggers a developmental response, initiating the formation of a new organ. Auxin is therefore an important morphogenetic regulator triggering developmental responses. Explaining its downstream signalling network is a necessary step to further understanding the observed cellular patterns that lead to shape changes at the organ scale. As we can see, we are facing a multi-scale problem where proteins can be organized as a network of interactions, locally present in each cell and where their output is finally integrated at tissue or even organ scale. Naturally many other levels of interplay exist quantifying the inputs and outputs of the system should help us to identify the rules regulating these phenomena.

Auxin signal transduction (perception, integration and transmission) is believed to be achieved by a transcription factor (TF) family made up of 52 interacting proteins (ARFs and Aux/IAA) (reviewed in Chapman and Estelle [2009]). Before our extensive analyses, little was known about the structure of this interaction network since only few components had been characterised. Moreover, only hypotheses were made regarding its function and no integrative views or testable models were available. Furthermore, the previous biological framework presenting the auxin signalisation mechanisms, as reviewed by [Leyser, 2006;

Chapman and Estelle, 2009], did not account for nor explain (observed) homo-dimerisation within the Aux/IAA subfamily.

Using genetical modification of the GAL-4 protein, Chien *et al.* [1991] proposed a high-throughout synthetic method to test for protein-protein interaction in Yeast. To create this synthetic system, they separated the DNA binding domain (BD) and the activation domain (AD) of the protein, which normally activate *GAL-4* expression after binding to its promoter. In this configuration the transcription of the reporter gene placed downstream the targeted DNA sequence is not possible. Testing the interaction capacity of two proteins thus requires to fused each to one domain, AD or BD, hence the name two-hybrid (or yeast-two-hybrid if achieved in Yeast). Finally in the case that the tested protein are interacting, by bringing together the AD and BD domains, they activate the expression of the reporter gene whose transcripts are detected.

To analyse the structure of such interaction network (undirected graph with loops), a standard solution consists of applying graph clustering methods. These methods often focus on community detection whose aim is to find groups of vertices that are highly intra-connected and poorly inter-connected [Schaeffer, 2007; Fortunato, 2010]. Interaction networks exhibit connectivity patterns that differ strongly from community structures and we looked for graph clustering methods that do not make strong *a priori* on the connectivity patterns. We selected a model-based clustering method relying on the stochastic block structure model introduced by Nowicki and Snijders [2001]. The aim of this mixture model for random graphs is to group vertices with similar connectivity profiles at the cluster scale and it has been shown by Daudin *et al.* [2007] that this can be used to model various types of graph structures (e.g. community structures but also scale-free networks or star patterns). This mixture model has been extended to valued graphs and explanatory variables have been incorporated in such models [Mariadassou *et al.*, 2010]. We chose to apply this modelling framework to test the influence of phylogenetic distances on the connectivity patterns of transcription factors involved in auxin signaling.

Switching scales from protein network, we now address the problem of auxin response at scale of a developing organ: the flower. In the last 15 years, a huge effort has been made to

identify the genes patterns implicated in the floral morphogenesis. This has led to identification of gene networks, and the introduction of a model of floral organ patterning, known as the ABC model. However, these models are mostly descriptive, meaning they are not mathematically or computationally testable since they are not defined as a set of equations. Therefore these models cannot be used to identify patterns or emergence of properties, nor to quantitatively link different scales, for instance genes and organ shape. Furthermore, it is neither possible to account for intermediary scales (e.g. cells and tissues), nor to define the rules and hierarchy in between them. Nevertheless they are an interesting basis to develop quantitative model, often by providing simplifying assumptions or higher order rules. In a quantitative modelling paradigm, hormones, genes and mechanical cues can be considered as inputs of the morphogenetic process, then cell and tissue shape can be viewed as the output. Being able to quantify some of the inputs and the output simultaneously should greatly facilitate understanding of what is happening “in-between”. Our goal is therefore to enable a quantitative linking between genes and shape by characterising dynamic geometrical cell features. We are confident that it will permit us to identify cellular patterns, and shifts in those patterns related to changes of genetic identity or external cues. In addition, comparing flower displaying phenotypic differences

Recently, promising developments have been made regarding the quantification of plant morphogenesis. A surface reconstruction of *Anagallis arvensis* developing SAM by Dumais and Kwiatkowska [2002] focused on enabling the extraction of surfacic cellular features with a non-destructive moulding method. The repeated moulding of the same SAM permitted, for the first time in plant, the extraction of live surfacic quantitative data at cellular resolution. Applying this technique to *A. thaliana* SAM, the authors could observe and quantify growth rate differences in the different SAM zones, coherent with the putative dynamical patterns of the SAM [Kwiatkowska, 2004]. They also detected a previously unsuspected phenomenon with their precise cellular quantification: a crease is forming at the periphery of the meristem before the bulge defining the FM initium appears [Kwiatkowska, 2006].

Clonal analyses of the developing petal of *Antirrhinum* [Rolland-Lagan *et al.*, 2003; Coen

et al., 2004; Green *et al.*, 2010] have led to a better understanding of growth pattern distribution in time and space. By developing a mechanistic *in silico* dynamic growth model they tested the effects on shape change of modifying the values of certain observed parameters such as growth rate, anisotropy and direction of growth. They concluded that the amount of anisotropy was responsible for stretching the petal while petal asymmetry was determined by the principal growth direction [Rolland-Lagan *et al.*, 2003]. However the local variations of the parameters were not implicated in generating petal asymmetry since using their average observed values were sufficient to reproduce it. Finally they proposed the presence of a long range signal maintaining growth direction along the proximodistal axis, quantitatively linking for the first time long-range signal to directional growth and generation of shape. Hence quantitatively linking “genes” and shape. It provides an interesting framework to quantitatively test and assess the interplay between genes and shape for the mature flower. However, this model is based on data acquired at larger scale than cellular resolution, has only been applied to nearly mature flowers (need apparent petal) and was first proposed in 2D. Recently an improved version of the growing tissue framework was proposed to model growth in 3D, although it models the growing tissue as a continuous sheet of material with two surfaces and a thickness, termed a canvas. This model is therefore not yet applicable to non-flat tissue like the SAM or the root.

Live-imaging of the whole tissue is a recent breakthrough in microscopy and it enables us to access the spatio-temporal events taking place during growth at a cellular level. Acquiring cell membrane positions (with a membrane marker) at several time points during tissue development allows for a quantitative description of its growth at cellular level using cell-based segmentation algorithms. Moreover, adding fluorescent genes or hormone reporters in the digitally-acquired organism can give us access to their spatial distribution and dynamics at a cellular level. Finally, even if the effort required to produce these data still limits the number of observed gene or protein expression patterns or hormones signalling pathways that may be simultaneously acquired, when compared compared to other omics tools, live-imaging is non-destructive for the sample and give us access to precise spatio-temporal information.

This therefore provides new opportunities to shed light on morphogenesis and its links to the spatio-temporal activities of genes and hormones at multiple scales.

The need for adequate solution to highly complex pattern recognition problems has emerged with the advent of an increase in biological dataset size and complexity. It has been faced by many scientists working in fields as diverse as biomedical engineering, computer science, applied mathematics and also image analysis or computer vision. Therefore I would like to stress the importance in exchanges between the fields of biology, mathematics, physics and computer science in order to progress toward a more integrative and comprehensive view of developmental biology.

The main drawback in the extensive use of live imaging is that the data generated (voxel-based 3D stacks) are extremely complex, since they represent the biological object under study. Creating algorithmic tools to tackle the time-consuming work of pattern recognition is a true challenge for all engineers and researchers facing those data. Indeed, going through the manual segmentation of thousands of cells for several time-series present too many drawbacks: too time-consuming, unrepeatable, open to user interpretation and so on. Our brain is naturally trained to recognise objects based on inputs (color, shape...) and spotting a cell nucleus, identifying a cell wall or interpreting the visual result of a reaction/test is often no challenge even in the case of partially missing data. However these tasks are extremely complex to encode mathematically and computationally. *Pattern recognition* and more generally *pattern theory* are fields providing concepts and tools that have proven successful in addressing the extraction and identification of highly complex patterns in communication, imaging, simulation (forecasting) and large networks.

As we have seen, the local accumulation of auxin triggers a developmental response leading to organogenesis by regulating its downstream gene networks through a perception network of transcription factors. These genes will impact on cell proliferation and differentiation, leading to changes in cell geometrical patterns, which overall will drive the shape changes at organ scale In the following chapters I will develop the work carried out on the characterisation of the auxin perception network and explain the contribution of my work

in deciphering network structures with graph clustering models. We will then see that even when changing scale to analyse flower morphogenesis, graph representations and clustering techniques still prove useful in organising and formalising the spatio-temporal evolution of the biological object.

Pattern theory is therefore a relevant framework for the two main biological problems we are facing, since it proposes to achieve a mathematical formulation of a small set of ideas through the identification of the patterns explaining both the structure and the variability of our data. Both developing tissues and the transcription network can be viewed as graphs where a cell or a transcription factor is a vertex and a wall or an interaction is an edge. In this way we encode the regularity and structure of the biological object and then aim to partition variability and quantify uncertainty in our data thanks to mathematical formalisations leading to pattern identification. It is therefore a powerful algebraic framework helping to organize a finite number of both local and global combinatory operations.

In the following sections I will present a summary of the literature regarding the auxin signalling pathway and early floral patterning. I will explain some of the genetic models, describe some of the diverse scientific approaches used and , when limited, those available that present the potential to go beyond these limitations. In addition, I will address the insights they have provided as a basis for a more integrative approach of the biological system.

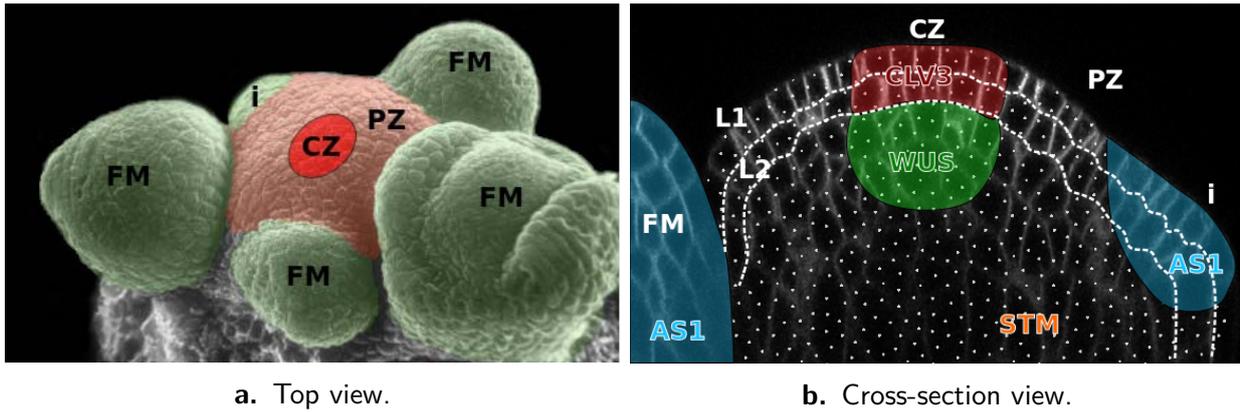
State of the art

1 The Shoot Apical Meristem: a self-maintained structure initiating new organs like flowers.

Flowering plants (Angiosperms) possess two primary meristematic zones, located at the tip of shoots and roots, that are responsible for the development of the aerial and underground parts of the plants, respectively. The apical meristematic zone, Shoot Apical Meristem (SAM), contain a domain of centrally located pluripotent stem cells (central zone), surrounded by a much larger region of multipotent cells that are competent to differentiate (peripheral zone). The meristematic zone is self-maintained and can modulate its developmental activity in response to external cues, providing the necessary flexibility to modulate plant shape and growth in response to varying environmental conditions. The SAM is responsible for a continuous organogenesis of leaves, internodes, axillary meristems and flowers [Traas and Doonan, 2001] throughout the plant life (see figure 1a).

1.1 Structural and functional zonation of the SAM.

One can distinguish two levels of cellular organisation at the SAM, a structural organisation in layers, consisting of the tunica (outer layers) and the corpus (inner layers), as well as a functional organisation in zones, of often specific cellular identity (figure 1b). In addition to this structural organisation, the SAM is partitioned in zones: mainly the central zone, the peripheral zone and the initiums. The central zone comprise slowly dividing cells, a characteristic of the stem cell function of self-maintenance. It is indeed essential to preserve these cells otherwise the organogenesis activity would stop [Laux *et al.*, 1996]. At the periphery of this zone, cells accelerate their growth rate and may enter into a differentiation step, leading



a. Top view.

b. Cross-section view.

Figure 1: SAM functional and structural organisations in *A. thaliana*. **a.** External top view of a –cleaned– SAM displaying the precise spatial disposition of the floral meristems (FM) in green. The SAM is indicated in pink. It is possible to appreciate the three first morphological stages of flower development from the floral initium (i) to a stage 3 FM (largest one at south-east of the picture) by successive counter-clockwise rotations of 137.5° . **b.** Cross-section view of a SAM highlighting the L1 and L2 tunica layers structure. Floral meristems (FM) and initium (i) can be observed on the sides. AS1, STM, WUS and CLV3 are introduced later. In both figures, CZ refers to the Central Zone and PZ to the Peripheral Zone, i to initium

to the formation of new lateral organs. Primordium initiation begin with the definition of a group of founder cells, that will give rise to a lateral organ, just outside the central zone. Several studies, like those from Grandjean *et al.* [2004] and Reddy *et al.* [2004], show that a small group of 4 to 10 founder cells emerge at the margins of the central zone in the first layer (L1), leading to organ formation.

1.2 Main regulators of the SAM.

Some of the main regulator involved in controlling SAM maintenance and function have been identified in the last twenty years. First, the *SHOOTMERISTEMLESS* (*STM*) gene has been shown to be necessary for SAM formation and post-embryonic maintenance. It belongs to the KNOX gene family that encode homeodomain transcription factors, and it is expressed throughout the meristem, except in the early developing primordia where it is down-regulated [Long *et al.*, 1996]. Another important homeodomain transcription factor, *WUSCHEL* (*WUS*) has been shown to be responsible for stem cell niche maintenance [Laux *et al.*, 1996]. The gene is expressed in the organising center, a small area located in the corpus below the central zone (figures 1), preventing the stem cells to differentiate [Mayer

et al., 1998]. A feedback loop between *WUS* and the *CLAVATA* genes (*CLV*) is responsible for controlling the size of the central zone [Schoof *et al.*, 2000; Brand *et al.*, 2000]. It has been demonstrated that the WUS protein moves to the central zone, thus establishing a gradient necessary for proliferation control [Yadav *et al.*, 2011]. When doing so, it is perceived by a *CLV1-2* receptor complex [Fletcher *et al.*, 1999; Ogawa *et al.*, 2008; Nimchuk *et al.*, 2011], activating *CLV3* expression [Kondo *et al.*, 2006]. This peptide likely diffuses broadly and acts non-autonomously to restrict WUSCHEL expression domain to the organising center, thus controlling its size by feedback. Finally, the over-expression of both *WUS* and *STM* is sufficient to induce SAM identity, even in root tissues [Brand *et al.*, 2002].

Discussion.

Phyllotaxy refers to the specific disposition of the newly formed organs along the axis of the stems. Some cells at the periphery of the stem cell niche, enter into a differentiation state and lose their meristematic identity by repressing *STM* function, thus permitting the expression of the transcription factor *ASYMMETRIC LEAVES 1* (*AS1*), which is normally repressed by *STM* in the SAM. This antagonistic mechanism therefore allows for a proper distinction between meristem and organ identity, whose development follows the activation of a set of genes controlling the size, the polarity and the identity of the organ.

However, these genes are downstream from a dynamic regulatory mechanism controlling organ initiation sites and rhythms. Hypothesised in 1868 by Hofmeister, who first observed that the emergence of a new organ was periodic and located in the largest space available between previously formed organs, it was not until 1913 that Schoute proposed the influence of chemical inhibitors produced by surrounding organs as a control mechanism. Since phyllotaxy implies the precise control of organ initiation sites and since auxin is a major inducer of organogenesis, it is highly important to further investigate how auxin signalling occurs.

2 Auxin control organ initiation and patterning.

The first demonstration of a moving signal mediating growth was made by Charles Darwin in 1880, who postulated the involvement of a “transmissible signal” in the regulation of organ growth in plants. Later in 1934, indol-3-acetic acid (IAA) or auxin was characterised as the first phytohormone [Paciorek and Friml, 2006], but it was not until the 1970s that a chemio-osmotic model for polar auxin transport was proposed [Rubery and Sheldrake, 1974; Raven, 1975].

Since, many efforts have been made to determine the physiological effects of auxin. The developmental processes in which auxin plays a role are legions: root and shoot patterning, fruit and embryo development, elongation and tropistic growth, apical dominance and organogenesis [Paciorek and Friml, 2006]. In 2003, Reinhardt *et al.* demonstrated the unique potential of auxin, by triggering organ (leaf) formation after ectopic application of auxin in the

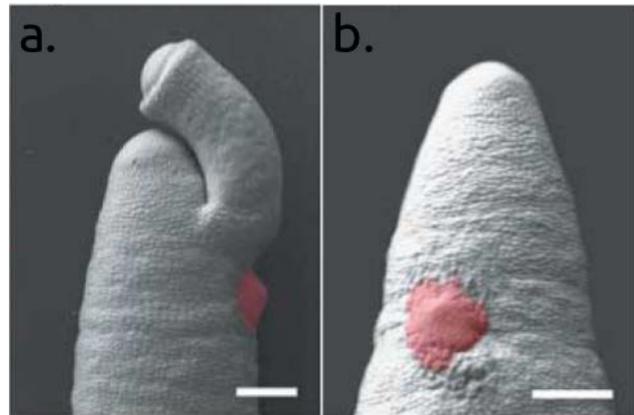


Figure 2: Effects of ectopic auxin application (red) on auxin-related mutants in *A. thaliana*. **a**, auxin induce lateral organ formation on *pid* mutant. **b**, auxin does not induce organ formation on *mp* mutants. Source: Reinhardt *et al.* [2003]

SAM peripheral zone of *pinoid* (*pid*) mutants (normally incapable of generating organs) (figure 2a). Around the same time, the first auxin biosynthesis sites were identified as the young leaves and the young, proliferating part of shoots and roots, as well as developing seeds [Ljung *et al.*, 2001, 2005].

2.1 Auxin levels are spatially controlled by efflux and influx carriers.

Even if the details of auxin perception, regulation and transport remain partially elusive, it is clear that auxin is actively transported through the plant. In fact both polar and non-polar transports of auxin takes place at the same time in plants, and it can be explained by the

chemio-osmotic model of Raven [1975]. Auxin being a weak acid, it can either be protonated or deprotonated depending on the local pH. The extracellular space possess a pH around 5.5, and auxin is then present in its acidic (protonated - uncharged) form, making it free to diffuse through the plasma membrane to enter the cell. Non-polar transport is achieved through diffusion process (along the gradient) in the phloem.

In addition, auxin influx can be facilitated by the *AUX1/LAX* family of influx transporters [Bennett *et al.*, 1996; Yang *et al.*, 2006], *AUX1* being specifically expressed in the L1. However, in the cytoplasm whose pH is close to 7, auxin is in its basic form (deprotonated - negatively charged) and can not diffuse through the membrane any more. Auxin efflux is then only possible thanks to active transport by *PIN-FORMED1* (*PIN1*) efflux carriers [Gälweiler *et al.*, 1998] which are more specifically expressed in the L1. *pin-1* mutants possess a SAM capable of initiating leaves, but an SAM incapable of initiating flowers [Okada *et al.*, 1991]. Finally, polar auxin transport can go against the gradient since it is an active process taking place in the vascular cambium.

2.2 Polar auxin transport regulation or self-organisation?

As demonstrated by Reinhardt *et al.* [2003], local accumulation of auxin is sufficient to trigger the various developmental roles of auxin, notably organ initiation at the periphery of the SAM apex (figure 2a). This auxin gradient is the result of complex interplays between auxin synthesis, degradation and active directed inter-cellular transport by influx and efflux carriers. However, if most of the major players have been identified, some aspects of polar auxin transport remains elusive.

It is indeed still unclear how the PIN1 efflux transporter polarisation is controlled at tissue scale. We know that PIN transporters are under the control of a rapid and complex mechanism of transcytosis at cellular scale. Moreover, recent studies suggest that auxin itself modify both expression levels and the polarity of PIN1 by inhibiting the latter mechanism [Heisler *et al.*, 2005; Paciorek *et al.*, 2005]. This has led to several theoretical cell-based models explaining self-organisation of the polar auxin transport (for review, see van Berkel *et al.* [2013]). The main drawback of the two main families of computational models, con-

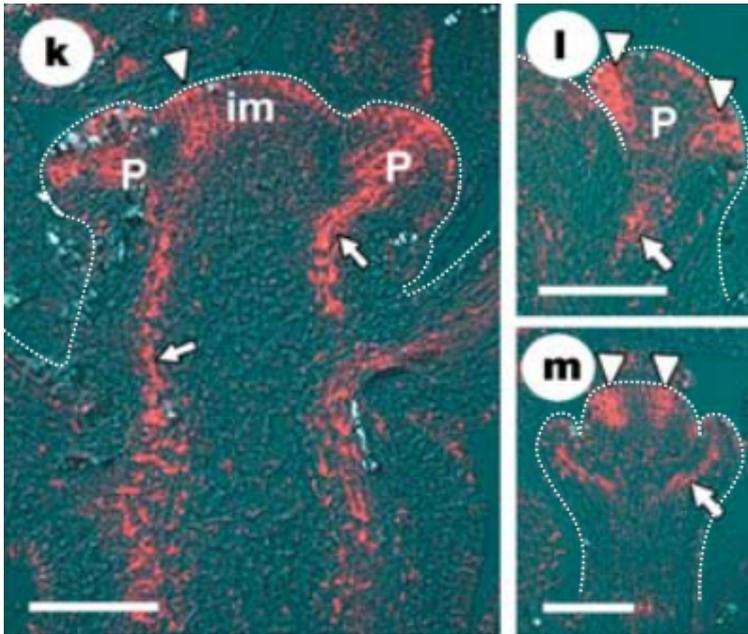


Figure 3: Localisation of PIN1 protein (red) in *A. thaliana* inflorescence and floral meristems. **k**, PIN1 localisation in inflorescence meristem (im) cross-section; arrowhead indicate initiation site. **l**, PIN1 localisation in young (stage 2) FM cross-section; arrowhead indicate sepal initiation site. **m**, PIN1 localisation in older FM (stage 3) cross-section; arrowhead indicate anthers initiation site. (P) indicate flower primordia, and white arrows indicate vasculature. *Source:* adapted (white outlines) from Reinhardt *et al.* [2003]

centration and flux-canalisation based models, cannot account for PIN1 dynamic pattern both in the L1 (organ initiation) and the vasculature (figure 3). Hybrid models have been developed and can reproduce a realistic pattern, but the existence of dual mechanisms seems unlikely. However they offers the merit to explain PIN1 polarisation mechanisms by using a simple feedback between auxin and its polar transport.

Anyhow, these models does not take into account the molecular mechanisms of auxin signalling network. Futhermore, since auxin modifies PIN polarisation through its signalling mechanism, it might play a role in regulating polar auxin transport. By incorporating auxin transport (influx and efflux) together with extracellular auxin receptor inhibiting PIN trafficking upon binding and PIN transcriptional regulation through nuclear signalling, Wabnik *et al.* [2010] proposed a model where patterns of PIN1 polarities are generated from synergy of intracellular and extracellular auxin signalling. This model successfully reproduces PIN distribution patterns both during formation of vascular systems and organogenesis in the L1. Finally, it is important to note that some of these models need sources of auxin, and if we understand “how” auxin is synthesised, the “where” is still unclear below organ scale. Knowing the latter would definitely help working on an even more realistic *in silico* dynamic model for auxin spatio-temporal patterning and polar transport.

Discussion.

Many models attempting to explain (self-)organisation of polar auxin transport have been proposed and the latests seems to reproduce quite faithfully the observed patterns of transporters as well as auxin distribution patterns. One could regret that these models were developed and tested for the SAM and not the FM. We can indeed clearly see similar expression patterns of PIN1 in the L1, the vasculature and at organ initiation sites in the FM as in the SAM (figures 3). Also, most attention was given to auxin mediated response after local accumulation by polar auxin transport, and most of the genes undergoing a modification of their expression levels in response to auxin (downstream) are identified.

However, as we will see in the next section, auxin response is mediated by a co-receptors and a network of transcription factors. This lead to a complex and tissue specific response to auxin signal, likely through the specific spatio-temporal expression patterns of the proteins implicated in the perception pathway leading to organogenesis. For example it has been shown that the apex of the SAM is insensitive to auxin local accumulation, since even though polar auxin transport accumulate auxin in the apex, no organ is initiated there. Auxin signalling pathway should therefore either be actively repressed there or absent. Finally, the signalling pathways between auxin and its response genes might modify the dynamic between auxin input and output since it implicate protein degradation and others non-instantaneous mechanisms as we will see in the next two sections.

We therefore worked at creating an expression map of proteins implicated in the auxin signalling pathway -receptors and transcription factors- to understand further the spatial regulation of auxin signalling pathway. This later phenomena will be detailed in the next section. Together with the analysis of the auxin transcription network structure I conducted, our work has enabled the modelling of a simplified version of the signalling pathway leading to a new dynamic model of auxin perception. We propose to explore the details of auxin transduction pathway in the following sections.

3 Auxin signal transduction is ensured by a TF network.

We call early auxin-related genes or auxin response elements (AuxRE) the genes contributing to the perception and transduction of the signal and late auxin-related genes those activated downstream the transduction pathway. Micro-array experiments demonstrate that hundreds of genes change their expression levels in response to auxin [Pufky *et al.*, 2003] but how such a small phytohormone can modify early and late auxin-related gene expression levels?

Another experiment by Reinhardt *et al.* [2003] demonstrated that *MONOPTEROS* (*MP*) / *AUXIN RESPONSE FACTOR 5* (*ARF5*) is essential for the SAM to enable the auxin response capacity since *mp/arf5* mutant meristems are insensitive to ectopic application of auxin (figure 2b). This indicates that, at least *MP/ARF5*, and maybe other transcription factors are recruited by auxin to regulate late auxin-related genes.

Made of 52 proteins in total, the Auxin Response Factor (ARFs) and Aux/IAA transcription factor family compose the perception network and are responsible for activating or repressing auxin-related genes. We will first discuss about their molecular and functional structures, then present the currently accepted descriptive model of auxin transduction pathway.

3.1 ARFs - Aux/IAA molecular and functional structures.

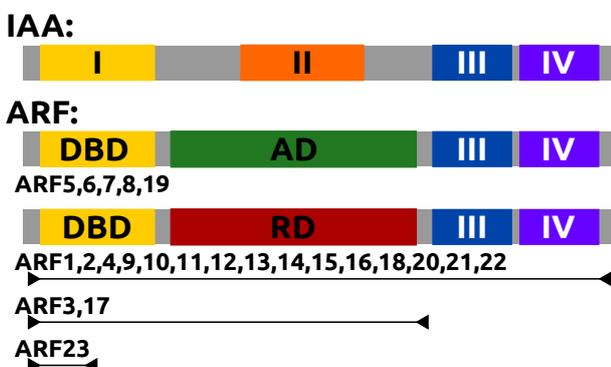


Figure 4: Schematic representation of the ARF and Aux/IAA structures as found in *Arabidopsis thaliana*. DBD: DNA binding domain. I: Aux/IAA specific putative homo-dimerisation domain. AD: Activation domain. RD: Repression domain. II: Aux/IAA specific degradation domain. III & IV: protein dimerisation domains. Arrowed lines indicate the extent of each inhibiting ARF structure. *Source:* adapted from Hagen and Guilfoyle [2002]; Guilfoyle and Hagen [2007] with the notable difference that we found domains III and IV when aligning full length protein sequences for ARF13.

Using a limited Yeast-2-Hybrid study, Kim *et al.* [1997] first proposed that protein-protein

interactions among the tested Aux/IAA proteins could regulate late auxin-related genes. The Aux/IAA possesses two domains unknown to the ARFs, domains I and II. While domain I function is unclear (putative homo-dimerisation function), domain II has been found to be responsible for the Aux/IAA instability [Hagen and Guilfoyle, 2002], targeting these proteins to the proteasome in presence of auxin.

Later, the ARFs dimerisation capacity was also proposed as a potential regulator of late auxin-related genes by Ulmasov *et al.* [1999b], their regulatory activity depending on their conserved middle domain amino-acid composition [Ulmasov *et al.*, 1999a]. In addition, the authors also uncovered the capacity of ARFs to bind to promoting sequences of these AuxRE by the intermediate of a DNA binding domain (DBD) found in N-terminal position of the proteins. The domains conserved across the two sub-families are protein-protein dimerisation domains, called domains III and IV, found in C-terminal position in both ARFs and Aux/IAA (see figure 4). These domains are conserved for nearly all the transcription factors except ARFs 3, 17 and 23.

The knowledge of the whole *Arabidopsis thaliana* genome finally extended the complex ensemble constituting the auxin transduction pathway to 23 ARF and 29 Aux/IAAs. A phylogenetic study by Remington *et al.* [2004] highlighted the strong conservation of function-specific domains in each sub-family, and two other conserved domains in both families (dimerisation domains III and IV). They showed that these two sub-families are long related through genes duplication and mutation-selection.

3.2 Auxin transduction pathway model.

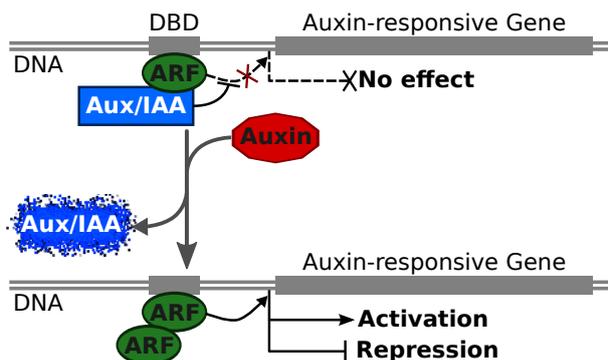


Figure 5: Model of auxin transduction pathway. Activation and repression activities depend on ARF middle domain amino-acid composition. DBD indicate the DNA binding domain found upstream of auxin-inducible genes. In absence of auxin Aux/IAA are dimerise with the ARFs, preventing them to exert their activating or repressing activity. When auxin is present, it target the Aux/IAA to the proteasome leaving the ARF free to dimerise and exert their regulating activity. Source: adapted from Chapman and Estelle [2009].

Starting from an initial situation without auxin, expressed ARF bind to the TGTCTC sequence of AuxRE found in their promoter [Ulmasov *et al.*, 1999b; Hagen and Guilfoyle, 2002; Pufky *et al.*, 2003]. In this configuration, expressed Aux/IAAs form hetero-dimers with ARF, thus preventing the ARF to regulate AuxRE expression levels. However, when auxin accumulate locally, the Aux/IAAs are degraded, because targeted to the proteasome by ubiquitination of the domain II by an SCF/TIR1 complex (for review see Paciorek and Friml [2006]; Leyser [2006]; Guilfoyle and Hagen [2007]; Chapman and Estelle [2009]). In turn, this enable locally expressed ARFs to exert their activator or repressor activity, depending on the amino-acid composition of their middle domain. The proposed model is summarised in figure 5.

3.3 Auxin perception is spatially regulated.

In addition to the control of auxin spatial distribution at the SAM, the spatial regulation of the auxin perception pathway has also been observed. The phytohormone is first perceived by a co-receptor complex made by the *TIR1/AFB* family of F-box proteins acting together with Aux/IAAs. Parry *et al.* [2009] demonstrated that *TIR1* and *AFB2-3* are post-transcriptionnally regulated by a microRNA *miR363*, thus identifying an element involved in the spatial regulation of the auxin perception. However, we should stress that the precise role of each member of the *TIR1/AFB* family is still unclear. Nevertheless, they could observe a complementary expression pattern between the microRNA and the auxin co-receptors in the root, further suggesting the possibility of a differential sensitivity to auxin depending on cell identity. The AFB family is not limited to *AFB2-3*, and working on seed, Greenham *et al.* [2011] have shown that *AFB4* and *5* interact in an auxin-dependent manner with *IAA3*, thus enlarging the family of auxin co-receptors. Also, by generating *AFB4:AFB4-GUS* reporter line, they observed a spatial restriction of *AFB4* expression pattern. This supports again the idea of auxin competent zones and of a cell-specific regulation of auxin co-receptors.

Discussion.

Only a limited amount of protein-protein interactions have been tested, and the models proposed by Leyser [2006] and [Chapman and Estelle, 2009] does not take into account that Aux/IAAs can dimerise with proteins from the same subfamily. The presence of five activators (ARF 5,6,7 and 19) among the 23 ARF, while the others are repressors, has not been taken into account in the model of auxin transduction pathways. The descriptive model is thus an interesting framework to understand the auxin signalling pathway but it does not fully integrate our knowledge on the subject. Finally a recent study linked *MP/ARF5* as a direct activator of genes initiating FM fate [Yamaguchi *et al.*, 2013], making the characterisation of the network even more interesting.

Y2H is an adequate technique to test protein-protein interaction at large scale and we used it to determine and characterise the structure of the “potential” auxin perception network. Potential because, as said earlier, the whole network is spatio-temporally regulated, thus not present as a whole in every cells. We could therefore hypothesise two to three levels of control, one depending on the dimerisation capacity, another depending on activation-repression by the transcription factors and finally one depending on the spatio-temporal control of TF expression patterns. The determination of auxin perception network structure would thus be an important contribution to the field, as well as a primary step toward its structure characterisation by statistical analysis. To further understand the perception network structure, we used proteic dimerisation sequences similarities to explain the origin of preferential mode of connexions. This led us to hypothesise a possible preferential use of one dimerisation domain out of two depending on the functional group of each interacting protein.

4 Growth patterns during floral morphogenesis lead to a stereotypic shape.

I now focus on a more macroscopic scale and examine the morphological events taking place during early flower development subsequent to its initiation by local auxin accumulation

and auxin response pathway. One striking aspect of flowers is that they are all made of four distinct organ types organised in concentric whorls: the sepals, the petals, the stamens and the carpels. Also, flowers from the same species display a great robustness in their overall shape, and organ numbers. Being responsible for seed production and pollen dispersion, the flowers execute a very important function of gene transmission.

Overall, flower development can be separated into three main steps. First the floral meristem identity genes determine the floral meristem fate. Second the floral organs identities are determined, later appearing as concentric whorls. Finally floral organ genes activate downstream effectors that specify the various tissues and cell types that constitute the different floral-organ types [Krizek and Fletcher, 2005].

4.1 Morphological description of early flower development.

The first description of flower development as a series of morphological events was made by Smyth *et al.* [1990]. This starts when the floral meristem outgrowth begins on the flanks of the SAM (stage 1), and goes up to the finale mature flower (stage 12) characterising each morphological change as a new stage. Stage 2 starts when the flower bud is separated from the SAM by the appearance of a small crease between the two meristems. Stage 3 is characterized by the emergence of the sepals from the sides of the floral meristem. These are therefore the three first morphological events displayed by the young floral meristem (FM) during its development (figure 6)

4.2 Induction of floral meristem fate.

According to Bossinger and Smyth [1996] the floral initium is made up of a group of four cells, which divide and expand radially, producing the group of cells from which all floral tissues are derived. In *A. thaliana*, the repression of *STM* and induction of *AS1* result in the expression of two key genes, *LEAFY (LFY)* [Weigel *et al.*, 1992] and *APETALA 1 (AP1)* in the flower primordium. Mutating one of these genes indeed results in a partial conversion of flowers into shoot-like organs; the *lfy ap1* double mutant shows a stronger phenotype. On the contrary, SAM can be converted to FM by the ectopic expression of *LFY* and *AP1*. In

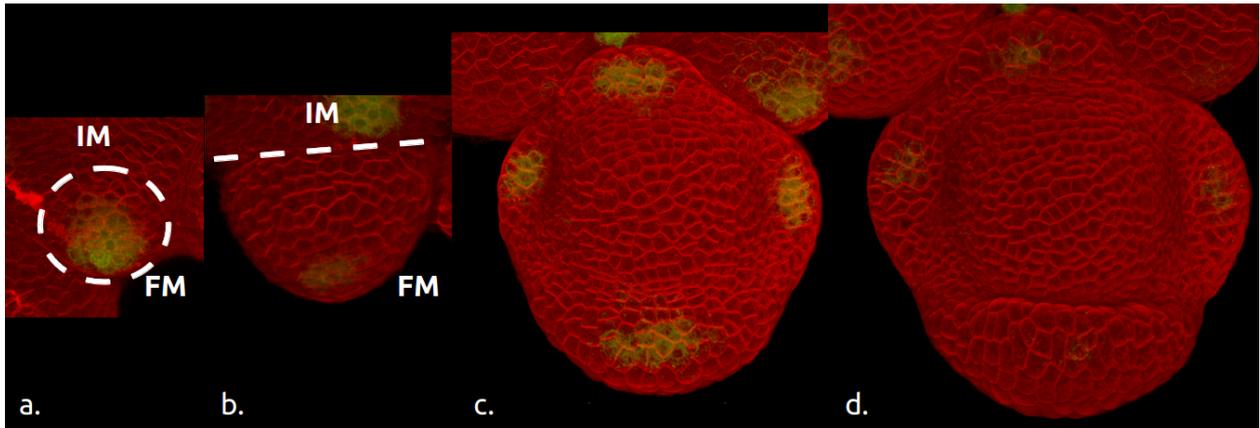


Figure 6: The three first stages of *A. th.* FM morphogenesis. **a.** Stage 1, no clear distinction between the FM and the SAM; **b.** Stage 2, a clear boundary is visible between the SAM and the FM (+26h); **c.** Early stage 3, sepals start emerging at the sides of the FM (+69h); **d.** Stage 3, the sepals are clearly visible (+80h). Green fluorescent marker: *AHP6::erGFP*. Source: Images acquired at the RDP by Pradeep DAS.

opposition to *LFY* and *AP1* the SAM fate is specified by *TERMINAL FLOWER 1* (*TFL1*).

Promoted by Gibberellin [Blazquez *et al.*, 1998], a phytohormone, *LFY* is one the first identity gene indicating flower primordium formation [Weigel *et al.*, 1992] and so one of the most upstream genes coding for a transcription factor. It targets downstream genes like *AP1*, *SEPALATA* (*SEP*) and other uncharacterised genes and putative transcription factors and signalling molecules detected by whole genome approach such as DNA microarray and Chromatin immunoprecipitation (Chip-chip).

Lastly, the boundary between the Inflorescence Meristem and the Floral Meristem is also controlled by a specific set of genes. The three known members of *CUP-SHAPED COTYLEDON* (*CUC*) family as well as *LATERAL ORGAN BOUNDARIES DOMAIN* (*LBD*) are known to play a role in establishing this boundary. However, no model linking gene or hormonal patterns to shape change is yet available to explain the SAM/FM boundary formation.

4.3 Floral meristem patterning by homeotic genes: the ABC model.

As shown in figure 7, three classes of genes were first identified by Coen and Meyerowitz [1991] to determine floral organ identity by analysing mutations affecting flower structure. In the wild-type background (fig.7a), all three classes are present and organise the four whorls. Mutating *APETALA 2* (*AP2*), the flowers produce carpels where sepals should be, and

stamens in position of the petals (fig.7b). Flowers of *apetala 3* and *pistillata (pi)* mutant plants does not display petals and stamens any more but -respectively- sepals and carpels instead. Finally the *AGAMOUS* flowers show an homeotic transformation of stamens into petals and carpels into sepals. *AG* also act as a termination signal for the flower, meaning when mutated *ag* flower from new ones inside the previous one, adding extra whorls of sepals and petals. Another key aspect is that class A and C genes function in an antagonistic manner, such that when class C genes are absent or non-functional, class A genes activity expand into class C and *vice versa*. This simple model provides a conceptual framework for explaining how the individual and combined activities of the ABC genes produce the four whorls of organ in most Angiosperm.

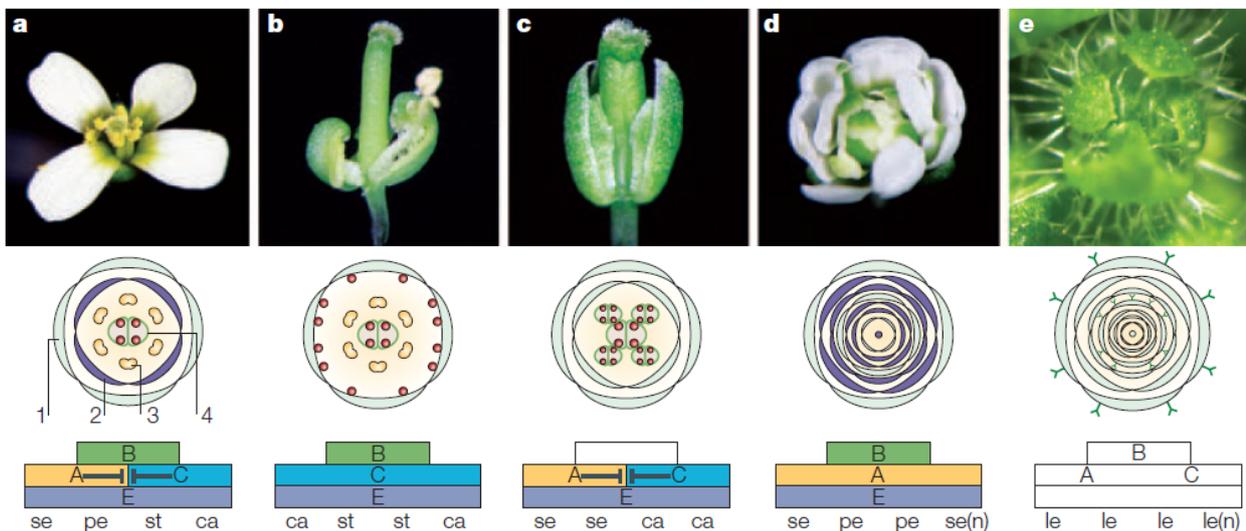


Figure 7: The ABCE model for floral-organ patterning illustrated with *A. thaliana* flower. **a.** the wild-type display four whorls of organs: the sepals, the petals, the stamens and the carpels. **b.** An *ap2* mutant flower, thus lacking class A genes, show the loss sepals and petals, replaced by carpels and stamens, respectively, subsequent to the expansion of class C genes throughout the flower. **c.** A *pi* mutant flower posses sepals and carpels instead of petals and stamens, respectively, subsequently to the absence of class B genes. **d.** An *ag* mutant flower, lacks class C genes activity, thus displaying sepals, two whorls of petals and further repeating the patterns in interior whorls. **e.** A *SEPALLATA* mutant flower for four *sep* genes (*sep1-4*) present reiterating whorls of leaf-like (le) organs subsequently to the lack of class E genes impairing the activity of ABC class genes, ultimately leading to loss of floral organ determinacy. Source: Krizek and Fletcher [2005].

Pelaz *et al.* [2000] first extended the ABC model by demonstrating the necessity of 3 additional MADS-box genes, the *SEPALATA1/2/3 (SEP1/2/3)* genes for the activity of B and C class genes. Moreover, *SEP3* as a flower-specific expression that restricts the action of the ABC genes to the flower [Honma and Goto, 2001]. Flowers of *sep1 sep2 sep3* triple

mutants indeed displays only sepals and *SEP4* is required redundantly with the other three *SEP* genes to confer sepal identity and contribute to the development of the petals, stamens and carpels [Ditta *et al.*, 2004].

Even though the details of the molecular mode of interaction between the homeotic genes remained a mystery for quite a while, the finding that MADS domain proteins can associate into higher-order complexes by Theissen [2001] shed light on the molecular aspect of floral organ identity. It is now accepted that MADS-domain proteins form ternary and quaternary complexes [Honma and Goto, 2001] known as the quartet model [Theissen, 2001; Krizek and Fletcher, 2005]. Thus the quartet model directly links floral organ identity to the action of four different tetrameric transcription factor complexes comprised of MADS-domain proteins [Theissen, 2001].

4.4 Structural and functional zonation of the early FM.

Subsequently to the change of identity from SAM to inflorescence meristem (IM), both IM and FM maintain the same layered structure as found in the SAM [Vaughan, 1955; Tilney-Bassett, 1991]. Studies of sectioned buds by Hill and Lord [1989] and Crone and Lord [1994] indicate that the first cell divisions generating sepals, petals and stamens usually occur in the L2 layer, whereas carpels might come from either L2 or L3 cells (figure 8). Overall, the L1 gives rise to the epidermis of all floral organs and to other tissues in the sexual organs (stamens and gynoecium), while the L2 and L3 gives rise to the mesophyll and other internal tissues [Jenik and Irish, 2000]. Since these three layers participate to tissue formation, they are called histogenetic. As reported by Jenik and Irish [2000], it is the control of the orientation of cell division that prevent layer mixing during morphogenesis. In addition, when dividing periclinally, the daughter cells are integrated into another layer and the development still proceed normally, leading to the idea that the final fate of a cell depends on its position rather than its lineage. Using a clonal analysis on *ap3-3* and *ag-1* mutants, Jenik and Irish [2000] found that cell division are regulated differently at early and late stages of FM development. Bossinger and Smyth [1996] suggested that the initiation and establishment of floral organ identity are separable processes; it is therefore of great

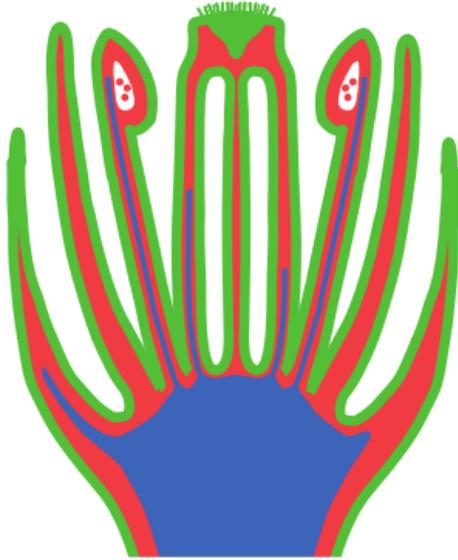


Figure 8: Contribution of the SAM histogenetic layers to floral organ tissues. L1 is in green, L2 in red and L3 in blue. The L1 gives rise to the epidermis of all floral organs and to some other tissues. L2 and L3 give rise to the mesophyll and other internal tissues. Source: Jenik and Irish [2000]

importance to quantify, in time and space, the patterns of lineage to more closely examine the cell fate determinacy.

4.5 Regulation of flower size.

The size of a mature flower can vary greatly between closely related species, but is usually invariant in any single species. Surprisingly, experimentally altering cell number or size is often not sufficient to change the size of the mature flower [Krizek and Fletcher, 2005].

First discovered by Elliott *et al.* [1996], *AINTEGUMENTA* (*ANT*) was originally associated to ovule development control, since *ant* mutants in *A. thaliana* displays partially developed ovules. Moreover, *ant ap2* double mutant plants display a total absence of floral organ [Elliott *et al.*, 1996]. Krizek [1999] linked *ANT* to cell growth regulation by observing cell features in a 35S::*ANT* background. The authors observed that the increased size of 35S::*ANT* sepals was the result of increased cell division, whereas the increased sizes of 35S::*ANT* petals, stamens, and carpels where in fact due to increased cell expansion [Krizek, 1999]. Another study revealed that *ANT* cell cycle regulation was, at least partially, the result of a prolonged expression of *CYCD3;1*, a D-type cyclin implicated in cell cycle regulation [Oakenfull *et al.*, 2002]. Following *ANT* identification, other actors that participate in the large network of cell proliferation regulation were also discovered. Acting downstream of auxin signalling, *ARF2* appears to have a role in control of organ size and cell proliferation,

by promoting [Hu *et al.*, 2003] and prolonging [Schruff *et al.*, 2006] *ANT* and *CYCD3;1* expression. Schruff *et al.* [2006] observed that *arf2* mutant plants display thicker stems and larger seeds, embryos, sepals and leaves. The putative role of *ARF2* is transcriptional repression of auxin response genes, and *ANT* expression is negatively regulated by *ARF2*. *AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE* (*ARGOS*), is highly induced by auxin and is also involved in organ size control, since transgenic plants expressing sense or anti-sense *ARGOS* cDNA display enlarged or reduced aerial organs, respectively. More recently, new genetic evidence by Krizek [2009]; Yamaguchi *et al.* [2013] have demonstrated that *ANT* acts redundantly with *AIL6/PLT3*, another member of the *AINTEGUMENTA-LIKE/PLETHORA* family.

Discussion.

Over the years, most of the effort to understand flower morphogenesis is related to the genetic determinacy of cell identity leading to organ initiation and organ development. However, a direct link between genes and shapes, as well as a precise quantification of the morphological events taking place during flower development, is still missing. Also, according to Vaughan [1955] floral organ growth in *Arabidopsis thaliana* is actually initiated from internal cells, making the sole observation of surface cells less informational regarding the potential sources of variation leading to shape change. Together with the previous details on structural zonation of the FM also lead to conclude that observing only the surface should not be enough.

In-depth live-imaging of growing tissues, together with the proper segmentation and lineage tools, allow us to explore and quantify the morphological events taking place during morphogenesis at cellular scale. We will develop in the next chapter the huge advances and tremendous potential that in-depth live-imaging of growing tissues bring to developmental biology.

5 Fluorescence live-imaging enable genes, shape and growth quantification at cellular scale.

I will now illustrate the great potential of fluorescence in-depth live-imaging, whose notable difference with nowadays -omic approaches, is to provide quantitative and dynamic data at single cell resolution. In this section, we will focus on methods using fluorophore-based biosensors which are widely used. The main reason for this success, especially for fluorescent protein, reside in its capacity to be synthesised by the cell, and therefore to label specific proteins by genetic manipulation. Hence it is possible to qualitatively and quantitatively observe these proteic biosensors inside the living tissue. I do not propose extensive reviews of every fluorescent biosensor (for review see Okumoto *et al.* [2012]), neither do I aim to describe all possibilities offered by fluorescence-based live-imaging techniques, but rather my aim is to simply their potential.

5.1 From staining to labelling.

To further grasp the full potential of these proteic biosensors, let's compare them to more traditional markers (not necessarily fluorescent) used in optical microscopy. Optical microscopy does not permit to see the inner parts of the tissue and thus require it to be sectioned. Unfortunately, after sectioning the biological object, light absorption/re-emission properties of the material are close to null, making the distinction of the tissue structures difficult. We call staining the external application of a dye, thus used to create a contrast by colouring parts of the tissue. They work by reacting to, or concentrating in, specific parts of a tissue, and using several dyes at once is possible to highlight different structures within a tissue.

Some of these dyes also possesses a fluorescent activity(emitting photons after excitation), making them useful for in-depth imaging, like the FM4-64 which selectively stain membranes. Although they present some limitations like being toxic at high doses and the fact they cannot label specific proteins.

The alternative to staining and the solution for labelling specific proteins is to create stable transgenic lines using proteic fluorophores. The first fluorescent protein that was isolated

is the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* by Shimomura *et al.* [1962]. Being a protein the GFP is producible by any cell possessing its DNA sequence, hence using a construct made of a gene of interest fused to the GFP gives the possibility to track (after maturation of the GFP) the resulting expressed protein inside the tissue. However, as we will see later, protein-GFP fusions also present some limitations.

Other techniques have been developed to label biological material, notably RNA, and they are reviewed in Dean and Palmer [2014].

5.2 What to label ?

Choosing what to label obviously depends on the goal of the experiment, on the biological object to work with, as well as the employed imaging technique. For instance it might not be possible to easily proceed to genetical modifications of the organism, in such cases the options are limited to staining as introduced earlier.

With the possibility to label proteins with a fluorescent marker, comes the possibility to follow crucial aspects of its dynamic such as its degradation, localisation or diffusion. Obviously, knowing the whole expression pattern of some proteins or of an hormone (indirectly) at cellular resolution may be crucial in understanding complex regulatory systems with multiple feedbacks and levels of control.

For example, using a novel Aux/IAA-based auxin signalling sensor termed DII-VENUS, Brunoud *et al.* [2012] managed to obtain a (negative) map of relative auxin distribution at cellular resolution in different tissues (SAM and RAM). Indeed the constitutively expressed, and rapidly maturing, DII-VENUS fluorescent protein is degraded in the presence of auxin, hence locally reducing the signal. This is due to the action of the DII domain (from the Aux/IAs, see sections 3.1 and 3.2) that cause the protein to be addressed to the proteasome in the presence of auxin. Together with another auxin activity reporter (DR5-VENUS), the analysis of their respective signals have shed light on the buffering capacity of the auxin perception network, which translates a fluctuating input (observed with DII-VENUS) into a stable output (observed with DR5-VENUS) [Vernoux *et al.*, 2011].

Another example of the unique potential of in-depth fluorescence live-imaging techniques

is the existence of biochemical sensitive fluorophore. For instance, some fluorescent biosensors are pH sensitive and acquiring their emitted signals, it is possible to quantify the pH levels of the cytoplasm. By targeting the fluorescent proteins toward specific sub-cellular compartments it is possible to obtain their pH [Kneen *et al.*, 1998] (for review see Bizzarri *et al.* [2009]). Although these pH-nanosensors are not perfect yet [Zhang *et al.*, 2013], they have been used to visualise fundamental processes like synaptic transmission [Miesenböck *et al.*, 1998], exocytosis and endocytosis of insulin [Ohara-Imaizumi *et al.*, 2002] or to detect primary and metastatic breast tumors [Lee *et al.*, 2011].

Ultimately, the key aspect of fluorescent proteins is their ability to provide quantitative data. A relationship indeed exist between the amount of fluorescence and the concentration of proteins given quantum yield and the dye extinction rate. In the case of fluorescence, quantum yield Φ is defined as:

$$\Phi = \frac{\text{number of photons emitted}}{\text{number of photons absorbed}}.$$

However this information is not always available and there are other means to obtain quantitative measurements of proteins concentrations. We will not explain here the principles of quantitative imaging, but for an extensive review of the available quantitative biosensors and their use see Okumoto *et al.* [2012]. Nevertheless, the numerous insights that available biosensors provide can be summarised as follows:

- quantitative protein tracking: localization (if membrane-localised, enables access to cell geometry), turnover and redistribution;
- quantitative protein dynamics: interaction and translocation (sub-cellular redistribution);
- quantification of biochemical properties provided by pH-sensitive, voltage-sensitive [Akemann *et al.*, 2012] or calcium sensing fluorochrome [Tada *et al.*, 2014].

Regarding diffusion or transport processes, adding a fluorescent protein to the protein of interest, might alter these processes by making a too “big” (context dependant) protein. Some alterations of the protein functionality have also been reported.

In summary, it is clear that live-imaging is a powerful tool to track fluorescent marker in the depth of the sample at high resolution (spatial: sub-cellular; temporal: minutes). All this at single-cell resolution but with a low genomic resolution. For an extensive review of the available quantitative biosensors and their use see Okumoto *et al.* [2012].

5.3 A 5D universe to compare -omic tools and live-imaging.

The idea of a 5D molecular universe was introduced in their review by Megason and Fraser [2007]. It is an interesting way of representing the problems faced in developmental biology. Noting ‘xyz’ the first three spatial dimensions, ‘t’ the temporal dimension and ‘g’ the genomic dimension, we thus face a molecular universe made of five dimensions ‘xyztg’. Picturing strengths and weaknesses of each techniques, in imaging or others conventional -omic, is facilitated by comparing how they will slice this multidimensional pie.

In order to further illustrate the differences with numbers, Megason and Fraser [2007] use the example of acquiring zebrafish genes expression patterns at near cellular resolution during its embryonic development (see figure 9). It possesses roughly 25 000 genes in its genome, would require a theoretical 100 points across each spatial dimension (high spatial resolution) and 1000 time points representing a temporal resolution of a few minute. With *in toto* live imaging technique, 25 000 developmental time-series would be required to build an atlas of genes expression patterns (if we label gene per genes). With conventional -omics much more data would be required (number of cells x number of time-points $\sim 1,000,000,000$) and we would still face the challenge of synchronisation.

Spatial dimension. Quite obviously, the conventional -omic techniques perform “poorly” in the spatial dimension since they require a lot more biological material than available in a single cell. Therefore the results of these methods are averaged over various tissues in case of multicellular organisms, and unsynchronised individuals in case of unicellular organisms. Another solution would be to do cell-sorting to only select the cells we are interested into. However, if we are interested in analysing a developmental process, it requires to be able to synchronise either the biological object development or the sampling to investigate the process at the same stage. In practice, this is often very difficult to do, even for single cell

organisms. On the contrary, with imaging techniques, it is possible image the development of the organism at single cell resolution, even sub-cellular resolution can be achieved in certain cases.

Temporal dimension. Considering the temporal (also called longitudinal) dimension, for they are destructive techniques, conventional -omic strategy are difficult to implement in time. Indeed, destroying the sample, these techniques impose a synchronisation between the samples, making the acquisition of accurate temporal data almost impossible. Conversely, being non-invasive, live-imaging techniques are perfect to acquire temporal data since they permit to follow the development of the same object in

time. We will note that, there are a few limitations to this assertion, since fluorochrome labelling can sometimes limit the temporal resolution or coverage of the acquisition by interfering with the normal development of the organism under study, or even kill it. Nevertheless fluorescent imaging also suffer of time-dependant drawbacks, inherent to the acquisition technique, like photo-bleaching (fluorophores extinction rate superior to maturation rate) then demanding greater time-interval between successive acquisitions.

Genomic dimension. On the contrary, for the genetic dimension, conventional -omics techniques are far more powerful (in term of resolution) than imaging. It is the very strength of -omics techniques that can analyse thousands of gene expression levels at once. However, -omics techniques relying on mRNA like microarray also posses a drawback. In case of regulating networks for example, it is often the proteins that apply the function. Knowing that there are several steps between mRNA and proteins, implies that mRNA quantification does not necessarily mean protein quantification. Gygi *et al.* [1999]; Greenbaum *et al.* [2003]

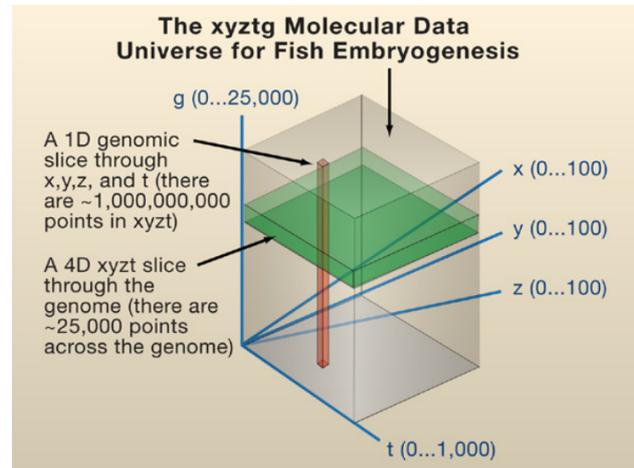


Figure 9: Theoretical example of a molecular universe for *in toto* imaging of zebrafish embryogenesis. The 'xyz' space correspond to spatial dimension, 't' to temporal dimension and 'g' to genomic dimension. Source: Megason and Fraser [2007].

worked at establishing this correlation between mRNA and protein levels, in yeast and other multicellular tissues. Their finding indicates that it is quite poor ($r \sim 0.4 - 0.6$) and can vary greatly depending on the cell type, developmental stage and category of proteins. On the other side, it is imperative for imaging techniques to label (with a fluorochrome) the desired object under study to acquire it. Protein quantification can be achieved by several means but remains a tricky task. Nevertheless obtaining its spatio-temporal pattern or relative expression level only require to attach a fluorochrome to the protein. This process can be limiting for organism where genetic transformation techniques are difficult to set-up. Moreover, tracking several proteins at once require spectrally non-overlapping fluorochromes, so their respective signals can be separated. The number of such fluorochromes currently available is still limited in number compared to the number of genes of most organism.

5.4 Examples of live-imaging in animals morphogenesis.

I propose to illustrate the excellent capacity of live-imaging to study morphogenesis by using examples from the animal community. Some model animals like the nematode *Caenorhabditis elegans*, ascidian *Ciona intestinalis*, zebrafish *Danio rerio* or fruit fly *Drosophila melanogaster* have proven adequate experimental subjects to advance in the field of developmental biology. In depth live-imaging of these biological models is indeed possible because most technical requirements are met: small size (fit as a whole under the microscope), short developmental time, transparency (good signal to noise ratio in the deeper parts of the tissues) and possibility to genetically modify the object. As we can see, it is the combination of the latest developments in every field (optic, biochemistry, image analysis...) that allow these acquisitions and analyses.

For example in the case of *C. elegans* the developmental plan of this worm is so precise that both lineage and cell fate are invariants from one individual to another [Bao *et al.*, 2006; Murray *et al.*, 2008]. Starting from the zygote, 13 hours later (larval stage 1) the organism is made of 558 cells. It will reach 959 cells at its adult size of about 1 millimetre long [Bao *et al.*, 2006]. Being easily genetically transformable and presenting a stereotypic development, this organism is perfect for making a 3D digital atlas (of the first larval stages)

at cellular resolution [Long *et al.*, 2009] and analysing genetic expression profiles of genes implicated in its embryogenesis [Murray *et al.*, 2008]. The 3D reconstruction of cell nuclei positions was made possible by segmentation algorithms after straightening the curved body. An interesting outcome of this 3D reconstruction is the possibility to estimate the degree of precision of cell nuclei positions by computing the average standard deviation of the cell nuclei position. Along the antero-posterior axis, the average standard deviation of the cell nuclei position is equal to $1.87\mu\text{m}$, demonstrating the extreme precision of the developmental plan.

Another illustration of image based approach to understand morphogenesis is given by the work of Bosveld *et al.* [2012] on *Drosophila melanogaster* thorax morphogenesis. They acquired fruit fly dorsal thorax using multi-scale imaging of ubiquitously expressed E-Cadherin (cell-cell adhesion proteins) fused with GFP. This time-lapse was made over 26 hours every five minutes, which after watershed segmentation yielded 2.8 millions cells overall. Using physical modelling, notably cells velocity map, they revealed how the Fat/Dachsous/Four-Jointed planar cell polarity pathway control morphogenesis of the thorax. They indeed first observed that the planar polarisation is first established by Dachsous which polarizes the myosin Dachs, then promoting anisotropy of junction tension. Finally using physical modelling on the basis of quantitative image analysis, they could demonstrate that this junction tension anisotropy is defining the pattern of local tissue contraction, in turn contributing to the epithelium shaping mainly via oriented cell rearrangements.

5.5 Imaging cell geometry to quantify shape and growth at cellular scale.

To access information on cells geometry, an obvious solution is to obtain a signal that marks cell outlines. Two different solutions to that end can be considered: either externally staining or using wall- or membrane-based fluorescent labelling.

For instance, external application of a fluorescent dye may be achieved using FM4-64 which selectively stain vacuolar membranes with red fluorescence (excitation/emission maxima 515/640 nm). However, the main drawback of most vital dyes is that they are lethal

at high doses and therefore not always well suited to study living tissues.

Labelling specific proteins with a fluorescent protein is possible if the organism is genetically transformable. Then using membrane addressing sequences, like LTI6b [Cutler *et al.*, 2000], fused to -say- a GFP it is possible to image the cells outlines.

Finally, on the basis of the outline, it is possible to manually measure 2D features but most of the time, due to the large number of cells, it is advisable to reconstruct an abstracted representation of the tissue by delimiting cells individually. In image analysis this process is called segmentation. Some have using Delaunay triangulations based on the cells nucleus to reconstruct the cells outline. We argue that quantifying geometrical features on such artificial reconstructions is erroneous, notably because the nucleus is not always at the center of its cell. However it can be use for illustration purpose, or 3D digital atlas construction as seen before with *C. elegans* [Long *et al.*, 2009].

5.6 Image processing: from intensity-based to object-based representations.

Going from a pixel (2D) or voxel (3D) based representation of intensity-based data to an abstracted object-based representation of the tissue is the core of *pattern recognition* methods. Implementing automated algorithms to carry out this task is a pre-requisite towards the large-scale quantification of spatio-temporal (or at least spatial) data. The main idea is to “explain” to the computer how to recognise a cell (outline) or a nucleus, a task the human brain achieve so easily (at least in 2D).

More and more trouble arises with missing information (e.g. absence or holes in the membrane signal) or noise (auto-fluorescence or random errors), which will introduce segmentation errors such as over segmentation (too many cells) or fused cells (e.g. if the cell contours are too thin). This is why it is of the greatest importance to obtain the best possible images, and also why intensity-based image filtering and corrections is also an active research area. Ultimately, segmentation algorithms will reconstruct cell shapes or recover cell nuclei positions in space, thus quantification of cell features, comparison between series and their statistical analysis will be possible.

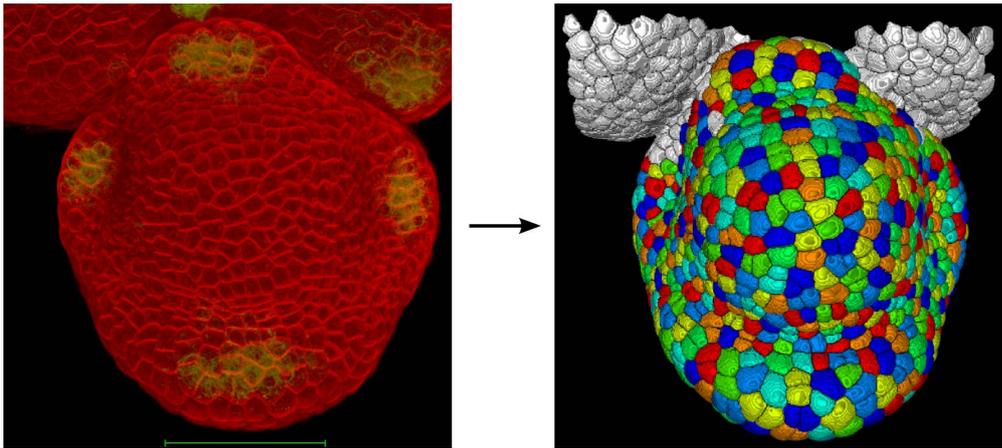


Figure 10: Illustration of intensity based image (left) transformed into a cell-based image (right). Source: Images and segmentation were generated at the RDP.

In the wake of a great increase in image acquisitions, and the relative youth of the image 3D reconstruction field, many dedicated segmentation tools have been built depending on their intended application and the nature of the object to recognise (e.g. cell shape, cell nucleus, neurons...). This resulted in a rather large choice of libraries, packages or software. I will focus on presenting 3D cell shape reconstruction algorithms since we want to achieve a statistical analysis of the spatio-temporal data. For a review about biological imaging software see Eliceiri *et al.* [2012].

Among the integrative platforms, we can cite ImageJ [Abràmoff *et al.*, 2004], Fiji [Schindelin *et al.*, 2012] and BioImageXD [Kankaanpää *et al.*, 2012]. ImageJ is the oldest of the free image analysis platforms and is widely used. Fiji is a bundled version of ImageJ containing plug-ins and features oriented toward microscopy images analysis. In both case, it is possible to readily use the integrated segmentation algorithms, and even to write or integrate your own.

Recently, dedicated 3D images analysis libraries have been developed, such as the Insight toolkit (ITK) [Yoo *et al.*, 2002]. However, except for its integration in BioImageXD, this library requires expert programmer to be used. Nevertheless, it offers many different 3D segmentation possibilities: Geodesic Active Contour, Shape Detection Level Set, Fast Marching Level Set, Region Growing and Watershed segmentation. In addition it integrate intensity-based image filters as well as segmentation correction tools and classification analysis methods.

Another software that is gaining in popularity is MorphographX [Kierzkowski *et al.*, 2012], which propose a graphical interface and the possibility to rapidly edit the segmentation using various correction tools. Initially proposing 3D surfacic segmentation, it will soon integrate the segmentation algorithms of the ITK library, thus offering full 3D segmentation.

Lastly, I cite the multi-angle reconstruction and segmentation - automatic lineage tracking (MARS-ALT) software [Fernandez *et al.*, 2010], which presents the unique feature to combine multiple views of the same object. It was developed to solve the problem of the lower resolution in z -direction (orthogonal to the focal plane), thus after deforming and fusing several images, even with a low tilt angle, it returns a 3D stack with the same resolution in all three directions (e.g. from $0.2 \times 0.2 \times 1\mu m$ to $0.2 \times 0.2 \times 0.2\mu m$) This reconstruction step is semi-automated since manual definition of landmarks is necessary to initiate the rigid and non-linear transformations. The segmentation algorithm used by MARS is based on a watershed algorithm. It has been fully automated using automatic seed detection, necessary to initialise the watershed algorithm, but it might be necessary to play with some of the parameters to achieve a reasonable segmentation.

Discussion.

Fluorescent live-imaging is still in its youth but it is rapidly growing and necessitates a lot of computational, storage and processing means. The possibilities offered are tremendous and there is no doubt that it will accelerate many research fields like biology and medicine. New or more powerful methods and hardware are invented every year, and we are already talking about the next generation of microscopy, the one that will get rid of optical limits: the “super-resolution” fluorescence microscopes Huang *et al.* [2009]. Conventional fluorescence microscopy is indeed limited in terms of spatial resolution because of the diffraction of light. This diffraction limit, permit to image up to 200-300nm in the lateral direction (xy) and 500-700nm in the axial direction (z). This is sufficient to properly image the cells, but it remains comparable or larger than most sub-cellular structures, thus it is not possible to observe them in details.

Also, in-depth imaging of the tissue requires energy (photons) to pass through the bio-

logical sample in order to activate a fluorescent molecule. This process sometime interfere with the normal development of the biological object under study, and can even kill it.

Another important challenge is obtaining signal with the best signal to noise ratio, causing the segmentation methods to fail and images to be blurry. In that view, Schleifenbaum *et al.* [2010] have proposed a method to enhance the contrast of the images, using differences in decay dynamic of both signals (true signal and auto-fluorescence). Other researches have focused on improving the signal quality by proposing monomeric fluorescent protein, with higher quantum efficiency, faster maturation and more photo-stability [Shaner *et al.*, 2004].

To conclude, many improvements can still be made at various levels (imaging hardware, biochemistry and image processing), but due to the widespread use of these still-emerging methods, many researchers are working toward this end.

Conclusions and questions

We have seen that plants present structures with zones of active organogenesis called meristems. The primary meristem of the plant aerial parts, the SAM, is responsible for the architecture of these aerial part by spatially and temporally regulating the site and rhythms of organs initiation. It is therefore an organising center with unique features like the presence of a stem cell niche located at the very tip, whose maintenance is essential for organogenesis to carry on. In the case of the angiosperms, the SAM is also responsible for the organogenesis of the reproductive organs: the flowers. FM organogenesis (initiation and development) then appears as a biologically crucial process.

The resulting distribution of organs along the stem axis is called phyllotaxis, and we have seen that it involves phytohormones that are distributed in spatio-temporal patterns and perceived by protein networks. These networks seems to act upstream developmental genes regulating the organ morphogenesis. However, if the regularity and stability of the phyllotaxy is attributable to spatio-temporal hormonal patterns, what could be the use and function of the auxin perception network?

Some preliminary results indicate that the auxin perception network could also be spatio-

temporally regulated, since *in-situ* hybridisation of some members of the network displays specific expression patterns. Moreover, three groups of distinct biological functions are believed to structure the network. Under the assumption that the auxin perception network structure is linked to its function, I first thought at determining its structure. It then proved crucial for ordinary differential equation modelling of the network, notably by simplifying the equation system, ultimately leading to propose putative function for the perception network. Ultimately, model validation came through the extensive characterisation of the TF expression patterns as well as the observation of the auxin distribution (input) and transcriptional (output) patterns.

In addition, it has been demonstrated that the TF composing the network are phylogenetically linked, since they originate from events of sequences duplication and mutation. In that wake, amidst some rare exceptions (ARFs 3, 17 and 23), they all present two highly conserved domains responsible for their dimerisation capacity. To further explain the TF network structure, I therefore thought it could be interesting to model the influence of phylogenetic distances between dimerisation sequences on the auxin signalling interactome structure.

After organ initiation comes morphogenesis, the organ shaping phase, controlled by the organ-specific developmental plan encoded in the homoetic genes. As demonstrated in all studied multi-cellular organisms, morphogenesis is a highly complex process involving genes regulatory networks, hormonal signals and mechanical feedbacks from the growing tissues. Moreover, these mechanisms act at different scales (molecular, cellular and tissular). This complexity has led some researchers to make use of modelling techniques to try to link gene functions and shape emergence into a growth framework. However, we still lack a precise 3D spatio-temporal cellular features quantification of a real growing tissue to test such predictive models. Nevertheless, fluorescent live-imaging has recently enabled this and we can thus ask if quantifying real spatio-temporal cellular features, together with dedicated statistical methods, can help the identification of morphogenetic patterns.

To do so, I will first define and organise the features extractable from a growing multi-

cellular tissue. Then using statistical methods, I will try to identify and characterise links between genes, local changes in cellular features, and shape emergence.

A striking aspect of both auxin perception network and morphogenesis is that they are both spatio-temporally regulated. Therefore, both the TF network and the growing tissue can be represented as graph. The TF network have been summarised as an undirected binary and valued graph as we tried to explain its function by analysing its structure. The growing tissue has been formally defined as an attributed spatio-temporal graph, summarising the spatial adjacency relations as undirected edges and the lineage as directed edges. The vertices and both types of edges can receive attribute describing the observed development of the FM. Finally, under the assumption that hidden functional patterns are in both cases structuring these graphs, we have used clustering methods to identify closely related vertices of the graphs that led to the characterisation of fundamental patterns.

To summarise, the main questions I will address in this thesis are:

- can we relate the structure and function of the auxin perception network?
- is the phylogenetic information relevant to the network structure?
- is it possible to identify and characterise morphological patterns of a growing floral meristem on the basis of its observed geometrical features at cellular scale?

Chapter 1

The auxin signalling network translates dynamic input into robust patterning at the shoot apex

Contents

Abstract	40
1 Introduction	41
2 Result and Discussion	43
2.1 Auxin signalling is spatially regulated in the shoot apical meristem	43
2.2 The Aux/IAA-ARF interaction network has a simple structure . .	46
2.3 The structure of the Aux/IAA-ARF network provides a plausible model for auxin signalling	48
2.4 The auxin signalling model allows predicting differential sensitivity to auxin and buffering capacities for the Aux/IAA-ARF pathway in the meristem	51
2.5 A novel signalling sensor, DII-VENUS, reports on the auxin signalling input in the meristem	53
2.6 Auxin signalling sensors distribution and dynamics confirm the model predictions	55
3 Discussion	56
4 Materials and methods	60
4.1 Plant material, growth conditions and plant treatments	60
4.2 Analysis of gene expression	60
4.3 Confocal microscopy and live imaging	61
4.4 Generation of plasmids and transgenic plants	61
4.5 <i>In vitro</i> transcription/translation of TIR1/AFB tagged proteins, immunoprecipitations and pull-down assays	62
4.6 Protein interaction analyses	63
4.7 Cluster analysis of Aux/IAA-ARF interaction network	65
4.8 Modelling of the Aux/IAA-ARF signalling pathway	68
Acknowledgements	70
Author contributions	70
Supplementary Materials	72

Abstract

The plant hormone auxin is thought to provide positional information for patterning during development. It is still unclear, however, precisely how auxin is distributed across tissues and how the hormone is sensed in space and time. The control of gene expression in response to auxin involves a complex network of over 50 potentially interacting transcriptional activators and repressors, the Auxin Response Factors (ARFs) and Aux/IAAs. Here we perform a large-scale analysis of the Aux/IAA-ARF pathway in the shoot apex of *Arabidopsis*, where dynamic auxin-based patterning controls organogenesis. A comprehensive expression map and full interactome uncovered an unexpectedly simple distribution and structure of this pathway in the shoot apex. A mathematical model of the Aux/IAA-ARF network predicted a strong buffering capacity along with spatial differences in auxin sensitivity. We then tested and confirmed these predictions using a novel auxin signalling sensor that reports input into the signalling pathway, in conjunction with the published DR5 transcriptional output reporter. Our results provide evidence that the auxin signalling network is essential to create robust patterns at the shoot apex.

1 Introduction

Auxin is a key morphogenetic signal involved in the control of cell identity throughout plant development. A striking example of auxin action is in the regulation of organogenesis at the shoot apical meristem (SAM). The SAM, a population of stem cells generating the aerial parts of the plant [Traas and Doonan, 2001], continuously produces new organs at precise positions at its periphery [Vernoux *et al.*, 2010]. The dynamics and robustness of organ positioning and patterning is thought to depend on local accumulations of auxin, generated by the PIN-FORMED1 (PIN1) efflux carrier controlling the direction of auxin polar fluxes [de Reuille *et al.*, 2006; Bayer *et al.*, 2009; Gälweiler *et al.*, 1998; Heisler *et al.*, 2005; Jönsson *et al.*, 2006; Reinhardt *et al.*, 2003; Smith *et al.*, 2006; Vernoux *et al.*, 2000] together with the AUX1/LAX influx carriers [Bainbridge *et al.*, 2008; Reinhardt *et al.*, 2003]. While the role of polar auxin transport has received extensive attention in the recent years, both auxin distribution and the contribution of its signal transduction pathway to patterning in the SAM are still largely uncharacterized.

A complex ensemble of 29 Aux/IAAs and 23 ARFs is central to the regulation of gene transcription in response to auxin (for review: Chapman and Estelle [2009]; Guilfoyle and Hagen [2007]; Leyser [2006]). Protein-protein interactions govern the properties of this transduction pathway [Del Bianco and Kepinski, 2011]. Limited interaction studies suggest that, in the absence of auxin, the Aux/IAA repressors form heterodimers with the ARF transcription factors (For review: Guilfoyle and Hagen [2007]) and recruit co-repressors of the TOPLESS (TPL) family, preventing the ARFs from regulating target genes [Szemenyei *et al.*, 2008]. In the presence of auxin the Aux/IAA proteins are targeted to the proteasome, by an SCF E3 ubiquitin ligase complex [Chapman and Estelle, 2009; Leyser, 2006]. In this process, auxin promotes the interaction between Aux/IAA proteins and the TIR1 F-box of the SCF complex (or its AFB homologs) that acts as an auxin co-receptor [Dharmasiri *et al.*, 2005a,b; Kepinski and Leyser, 2005; Tan *et al.*, 2007]. The auxin-induced degradation of Aux/IAAs would then release ARFs to regulate transcription of their target genes. This includes activation of most of the *Aux/IAA* genes themselves, thus establishing a negative feedback loop [Guilfoyle and Hagen, 2007]. Although this general scenario provides a framework for

understanding gene regulation by auxin, the underlying protein-protein network remains to be fully characterized. In addition, while auxin predominantly activates transcription (Paponov *et al.* [2008] and references therein), sequence analysis and transient assays suggest that most of the 23 ARFs act as transcriptional repressors while only 5 (ARF5, 6, 7, 8 and 19) of them are activators [Guilfoyle and Hagen, 2007], further highlighting the need for an integrated view of this pathway.

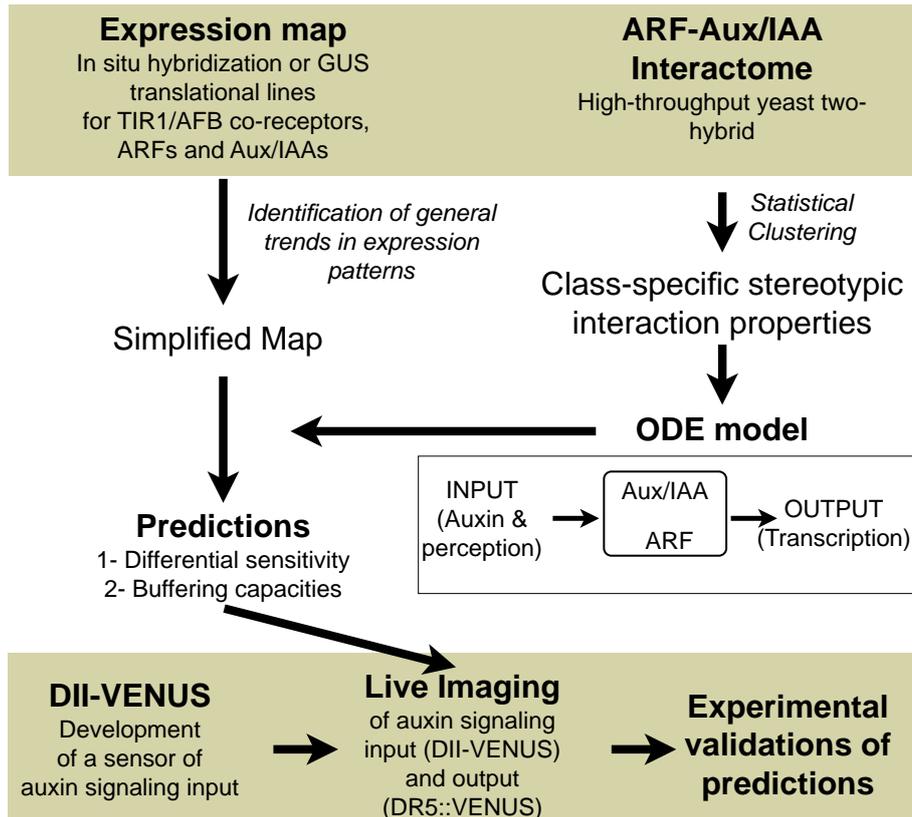


Figure 1.1: Flowchart representation of the strategy and main findings. The experimental parts of the work are shown in light green boxes. The connections between the different parts of the work are represented by directed arrows.

To understand how the Aux/IAA-ARF pathway contributes to sensing auxin in space and time, we have conducted a large-scale analysis of the Aux/IAA-ARF network in the inflorescence SAM. Our strategy and findings are summarized in the flowchart presented in Figure 1.1. A comprehensive expression map and full interactome uncovered a relatively simple distribution and structure of this pathway in the shoot apex. Using Ordinary Differential Equation (ODE) modelling, we then predicted spatial differences in auxin sensitivity and a strong buffering capacity of the Aux/IAA-ARF network in the SAM. Lastly, the development of a novel auxin signalling sensor allowed us to dynamically visualize the input

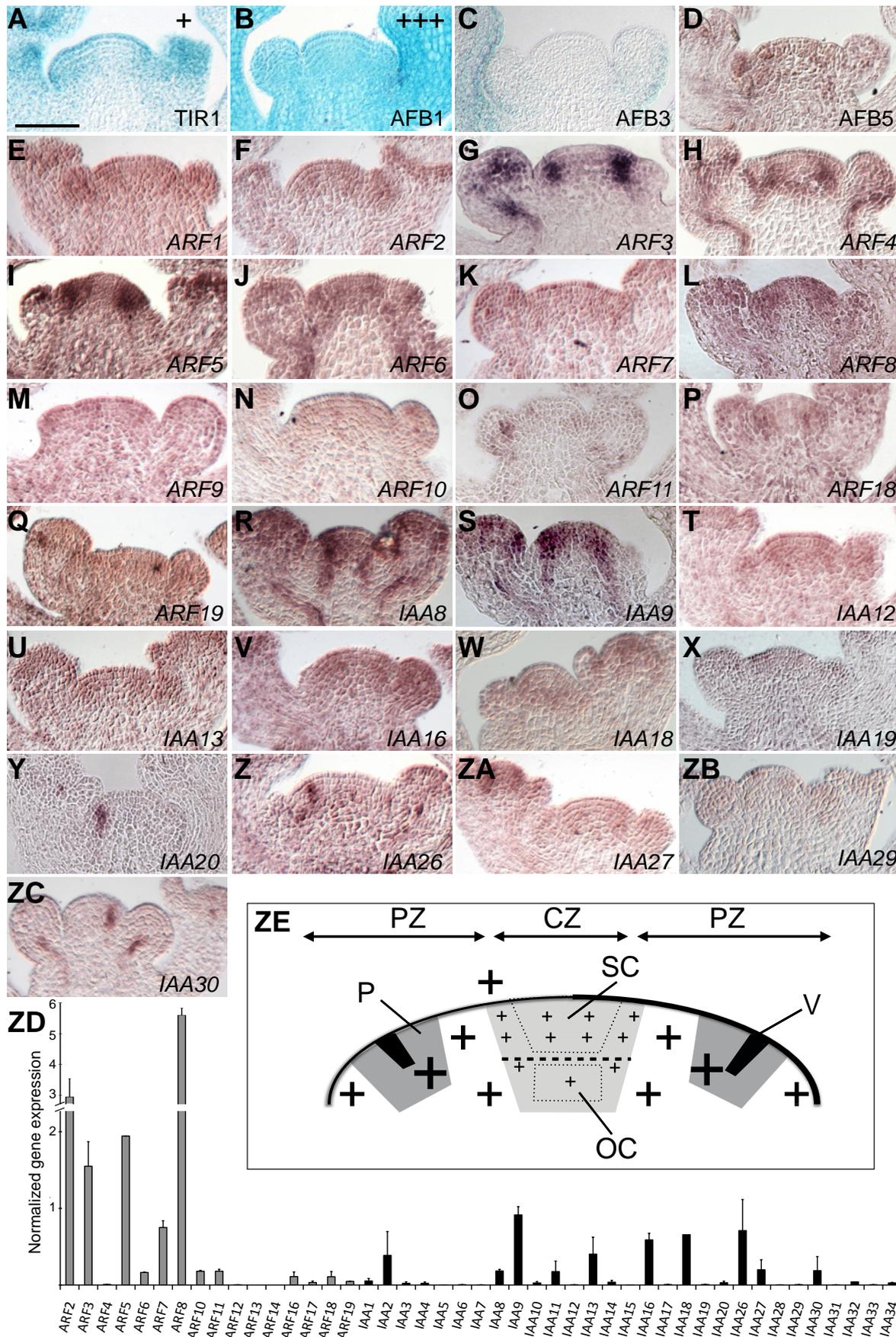
into the signalling pathway (a parameter that depends on both auxin concentration and perception) and to confirm the predictions of our model. Taken together, our data provide evidence that robust patterning at the shoot apical meristem depends not only on auxin distribution but also on the local properties of the Aux/IAA-ARF signalling network.

2 Result and Discussion

2.1 Auxin signalling is spatially regulated in the shoot apical meristem

To fully understand how auxin signalling is regulated in the SAM, we analysed the expression patterns of the F-box co-receptors as well as of the *Aux/IAA* and *ARF* genes. First, we used translational fusions to the GUS reporter gene to study the expression of the AFB1-3 and TIR1 auxin co-receptors in the inflorescence (figure 1.2A-C; Supplementary Figure S1.1A-D for serial sections) [Parry *et al.*, 2009]. We found that AFB1 is the most abundant auxin co-receptor (based on GUS activity), homogeneously expressed in the meristem, while AFB3 is expressed in tissues below the meristem. AFB2 is undetectable in the meristem proper. TIR1 is weakly expressed throughout the meristem and shows a reduced expression in the L2 and L3 layers at the centre. Auxin receptor activity of the AFB4 and AFB5 F-box proteins was also recently demonstrated [Greenham *et al.*, 2011]. We thus used *in situ* hybridisation to study the expression patterns of the corresponding genes in the SAM. While AFB4 could not be detected at significant levels, we did detect a low expression of AFB5 throughout the meristem that was slightly higher in the L2 and L3 layers of organ primordia (figure 1.2D, Supplementary Figure S1.1E,F). Hence, our results suggest that TIR1, AFB1 and AFB5 control auxin-induced degradation of Aux/IAAs in the inflorescence meristem.

We next used systematic RNA *in situ* hybridisation to obtain the patterns of all *ARFs* and *Aux/IAAs* (with the exception of *ARF15* and *21* for which we could not obtain a cDNA). We found 13 *ARFs* and 12 *Aux/IAAs* expressed in the meristem (figure 1.2E-ZC, Supplementary Figure S1.2-S1.3 for serial sections). Expression was confirmed using quantitative RT-PCR on total RNA from dissected meristems (Figure ZD), except for *ARF4* and *IAA12* that could



(Caption next page.)

Figure 1.2: Spatial regulation of Aux/IAA-ARF signalling in the inflorescence. **(A-D)**. Expression patterns of TIR1/AFB F-box co-receptors. Expression was analyzed using GUS translational fusions for TIR1, AFB1 and AFB3 and *in situ* hybridisation for *AFB5*. The relative levels of the protein are indicated for AFB1 and TIR1, as revealed by GUS activity detection time (TIR1: 48h; AFB1: 8h). **(E-ZC)** Expression patterns of *ARFs* and *Aux/IAAs* revealed by RNA *in situ* hybridization. **(ZD)** Detection of *Aux/IAA* and *ARF* expression by RT-qPCR in the inflorescence meristem. The analysis was done in duplicate on meristem mRNAs. Error bars represents the range of values. **(ZE)** Schematic representation of the *Aux/IAAs* and *ARFs* distribution in the meristem. The meristem is represented as a dome (PZ: peripheral zone; CZ: central zone, in grey; SC: stem cells; OC: organizing centre). Global tendency in expression levels are indicated by the size of the + sign. A dashed line was drawn between the upper and lower part of the centre to indicate differences in signalling capacities since *ARF4* and *IAA18* are expressed in the inner core. The primordia (P) have been delineated by a dashed line in the PZ to indicate that several *Aux/IAAs* and *ARFs* shows an even high accumulation in the organ primordia. Several *Aux/IAA* and *ARF* are also more specifically associated with vasculature (V; see main text). Median or near-median sections are shown. Scale bar: 50 μm .

not be amplified, as well as *IAA2* and *11* that were amplified but not detected by *in situ* hybridisation. The expression of 2 *ARFs* (*ARF11* and *19*: figure 1.2O,Q) and 4 *Aux/IAAs* (*IAA20*, *26*, *29* and *30*: figure 1.2Y,Z,ZB,ZC) appeared to be restricted to the presumptive vasculature in the organ primordia or just a few cells in the epidermis (see also Supplementary Figure S1.2 and S1.3) while *ARF9* and *10* showed a weak homogenous expression (figure 1.2M,N and Supplementary Figure S1.2). Most strikingly, all the other *ARF* and *Aux/IAA* genes (9 and 8 genes respectively) were expressed at higher levels at the periphery and at lower levels in the centre of the meristem (figure 1.2E-ZC and Supplementary Figure S1.2-S1.3) [Hardtke and Berleth, 1998; Pekker *et al.*, 2005; Sessions *et al.*, 1997; Wu *et al.*, 2006]. Amongst these, *ARF1*, *2*, *5*, *7*, *8*, *18* and *IAA12*, *13*, *18*, *27* were expressed homogeneously throughout the periphery of the meristem (although with different intensities), while *ARF3*, *4*, *6* and *IAA8*, *9* showed an even stronger expression in organ primordia. *ARF4* and *IAA18* were also expressed higher in the inner core of the meristem, below the stem cells at the centre of the meristem.

These *in situ* hybridisation results show that the expression of the *Aux/IAA* and *ARF* genes defines 5 different domains in the meristem (figure 1.2ZE). However the general tendency observed for most of the 25 *ARFs* and *Aux/IAAs* detected in the SAM is a differential expression with low levels at the centre of the meristem (where the stem cells are located) and high levels at the periphery of the meristem (where organ initiation takes place). For some of the genes an even higher expression is observed in the organ primordia. This differential

expression of auxin signalling regulators is also paralleled by a lower expression of TIR1 in the inner core of the meristem and the higher expression of AFB5 in the internal part of primordia. Furthermore our data show that ARF activators and repressors are largely co-expressed, suggesting a role for this co-expression in the regulation of transcription in response to auxin.

2.2 The Aux/IAA-ARF interaction network has a simple structure

To understand the functional significance of the distribution of *ARFs* and *Aux/IAAs* in the SAM, we next investigated the global structure of the Aux/IAA-ARF network using a high-throughput yeast two-hybrid approach. This analysis revealed 433 positive interactions amongst 1225 tested, indicating a highly connected network (figure 1.3A and Supplementary Table S1.1). 78% (51 in 65) of the interactions tested in the literature were confirmed in our data, indicating a very good coverage (Supplementary Table S1.2 and references therein). We also confirmed 28 interactions out of 31 (90%) tested *in planta* using bimolecular fluorescence complementation, further supporting the biological significance of our data (Supplementary Figure S1.4).

To explore the organization of this network, we applied a graph clustering method [Picard *et al.*, 2009] that groups proteins based on their connectivity profile (i.e. proteins with similar interactors). Three well separated clusters, characterised by contrasting probabilities for within- and between-cluster connectivity, were uncovered (figure 1.2B, Supplementary Figure S1.5): proteins in cluster I were strongly and similarly connected to each other and to the proteins of cluster II; proteins of the cluster II were strongly connected to those of cluster I, but sparsely to themselves; finally, proteins from cluster III showed a low connectivity to the rest of the network and to each other. Similar results were obtained when we restricted the analysis to the subset of ARFs and Aux/IAAs present in the SAM, indicating a similar interactome topology (Supplementary Figure S1.6). Further examination showed that cluster I contained mainly Aux/IAAs, cluster II mainly ARF activators and cluster III ARF repressors (figure 1.2C,D, Supplementary Figure S1.5-S1.6). For the SAM-specific

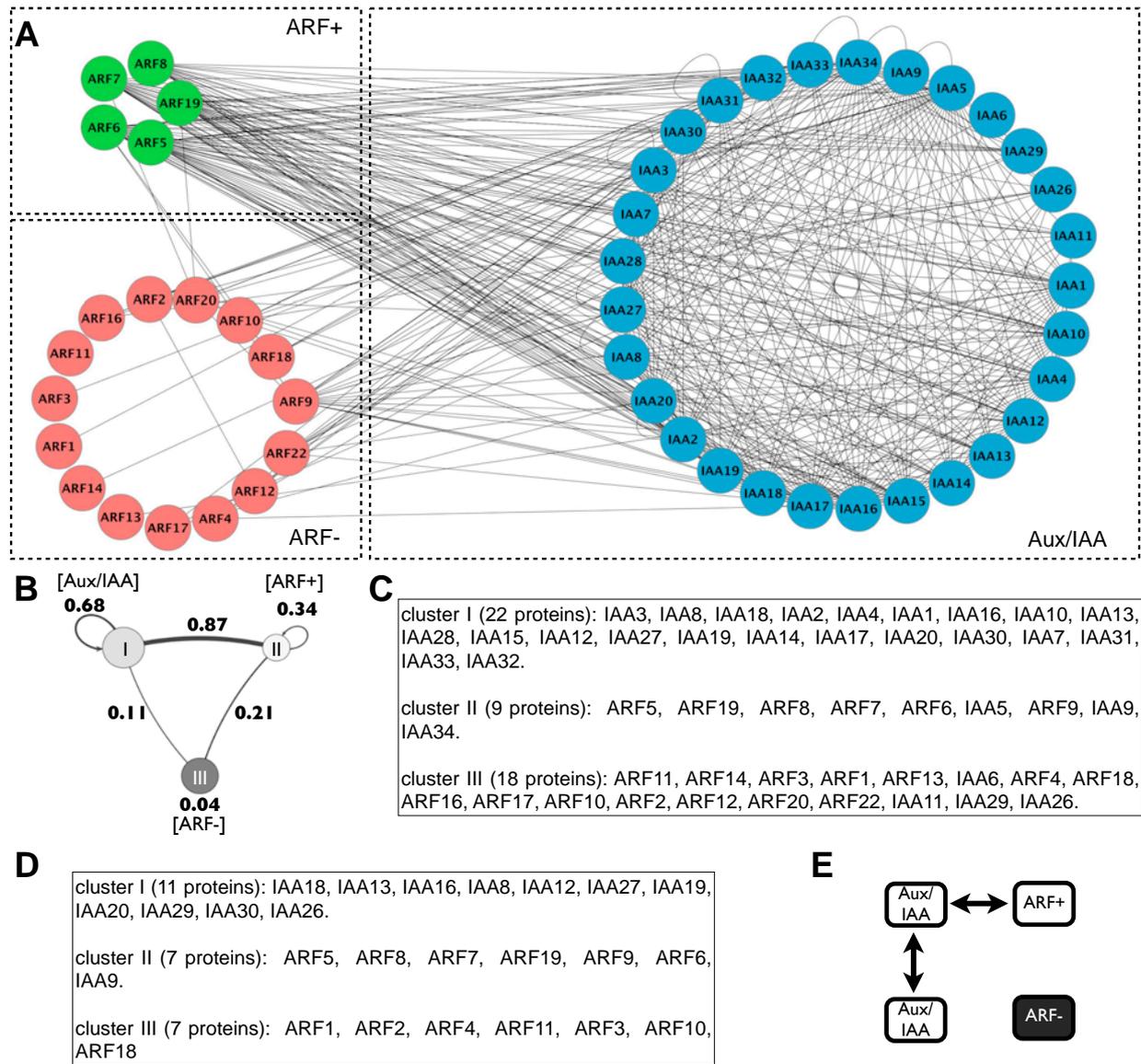


Figure 1.3: Structure the auxin signalling network. **(A)** Visual representation of the Aux/IAA-ARF interactome using Cytoscape (www.cytoscape.org). The proteins are grouped according to their biological identity as indicated. Note the global differences in connectivity of the three biological groups **(B-D)** Connectivity graph and clusters identified by the MixNet algorithm. The probabilities associated with the connectivity structure for the global network are indicated in **(B)**. The three clusters are mainly composed of Aux/IAA (I), ARF activators (II) and ARF repressors (III) as indicated in brackets in **(B)**. The identity of the proteins in these clusters for both the global network **(C)** and the SAM-specific network **(D)** is shown. The proteins are ordered from the most to the least central in each cluster based on the distance of the protein to the cluster. **(E)** The topology of the network relies on stereotypic interaction capacities for the different classes of proteins as represented. ARF+: ARF activators; ARF-: ARF repressors.

network (figure 1.2D) the only exceptions were the ARF9 repressor and IAA9, which showed a connectivity profile closer to the ARF activators and were assigned to cluster II.

Our results thus indicate that the topology of the whole network, as well as of the SAM specific network, relies on three principal features (figure 1.2E): (i) Aux/IAA proteins

interact with themselves, (ii) Aux/IAA proteins interact with ARF activators and (iii) ARF repressors have no or very limited interactions with other proteins in the network.

2.3 The structure of the Aux/IAA-ARF network provides a plausible model for auxin signalling

Having obtained the structure of the Aux/IAA-ARF network, we next investigated the impact of this structure on transcriptional regulation in response to auxin. The results of our interaction analyses suggest a model for the Aux/IAA-ARF signalling pathway in the SAM where transcriptional activation by ARF activators would be negatively regulated by two independent systems, one involving the ARF repressors and the other, the Aux/IAs. The presence of auxin would remove the inhibitory action of Aux/IAs but leave the ARF repressors to compete with ARF activators for promoter binding sites. To explore the regulatory properties of this signalling network involving multiple elementary reactions and a feedback loop, we developed a mathematical model using ODEs.

In addition to data on the Aux/IAA-ARF interactome, the model was based on a set of four other general considerations. First, that ARF activators and ARF repressors may regulate the same target genes, since they can bind the same AuxRE element found in promoters of auxin-induced genes [Goda *et al.*, 2004; Nemhauser *et al.*, 2004; Pufky *et al.*, 2003; Ulmasov *et al.*, 1999b]. Second, that the promoters of auxin-induced genes have one or two AuxREs, though a few may contain three or more [Goda *et al.*, 2004]. Third, that all the *Aux/IAA* genes expressed in the SAM except IAA16 and 27 are induced by auxin (Paponov *et al.* [2008] and references therein). Fourth, that amongst the 23 ARF genes, only ARF4 and ARF19 expression can be induced by auxin [Paponov *et al.*, 2008], thus indicating that expression of the majority of *ARFs* is independent of auxin.

These experimental observations lead to the reaction scheme shown in figure 1.4A (a more detailed version of this scheme is given in Figure 1 of Note S1) where:

1. the interactions between proteins occur according to figure 1.3E;
2. target genes promoter contains two AuxREs;

3. target genes are regulated by both ARF activators and ARF repressors;
4. the target genes include the *Aux/IAA* genes but no *ARF* genes;

In this scheme we also considered three additional assumptions:

5. A cooperative effect occurs when two ARF activators are bound on target gene promoters as suggested by Ulmasov *et al.* [1999b]
6. binding of auxin to TIR1/AFB co-receptors and Aux/IAs is a faster process than ubiquitination, implying that the effect of auxin on the system can be directly described as an increase of Aux/IAA degradation rate (see Note S1 for further details).
7. the auxin-mediated degradation affects the stability of both Aux/IAA monomers and dimers. In the latter case, the partner of the Aux/IAA (ARF activator or Aux/IAA) is released.

Concentrations of populations of Aux/IAs, ARF activators, ARF repressors and of mRNA of auxin-induced genes were then described mathematically using a system of 5 ODEs (figure 1.4A; see Material and Methods for full description of the model). For the analysis of gene transcription in response to auxin, the most important aspect of this mathematical description is that it allows analysing the transcriptional output R as a function of the auxin level x . It is important to stress that, upon assumption (6), the auxin level x directly modulates the Aux/IAA degradation rate as a function of both auxin concentration and perception of auxin (see Material and Methods). The parameter x can thus be viewed more generally as the input into the signalling pathway and changes in this parameter can reflect either changes in auxin concentrations or changes in the TIR/AFB co-receptors levels.

We performed a mathematical and numerical study of the model (detailed in Note S1). The main conclusions of this analysis are as follows:

1. We proved that in a range of plausible parameters, the system always reaches a unique steady state. Simulations performed by varying the different parameters further indicate that this steady state is stable. The effect of modifying the auxin concentration x was to shift the system to a new value of the steady state, in other words to change the level of all variables and notably the transcriptional output R .

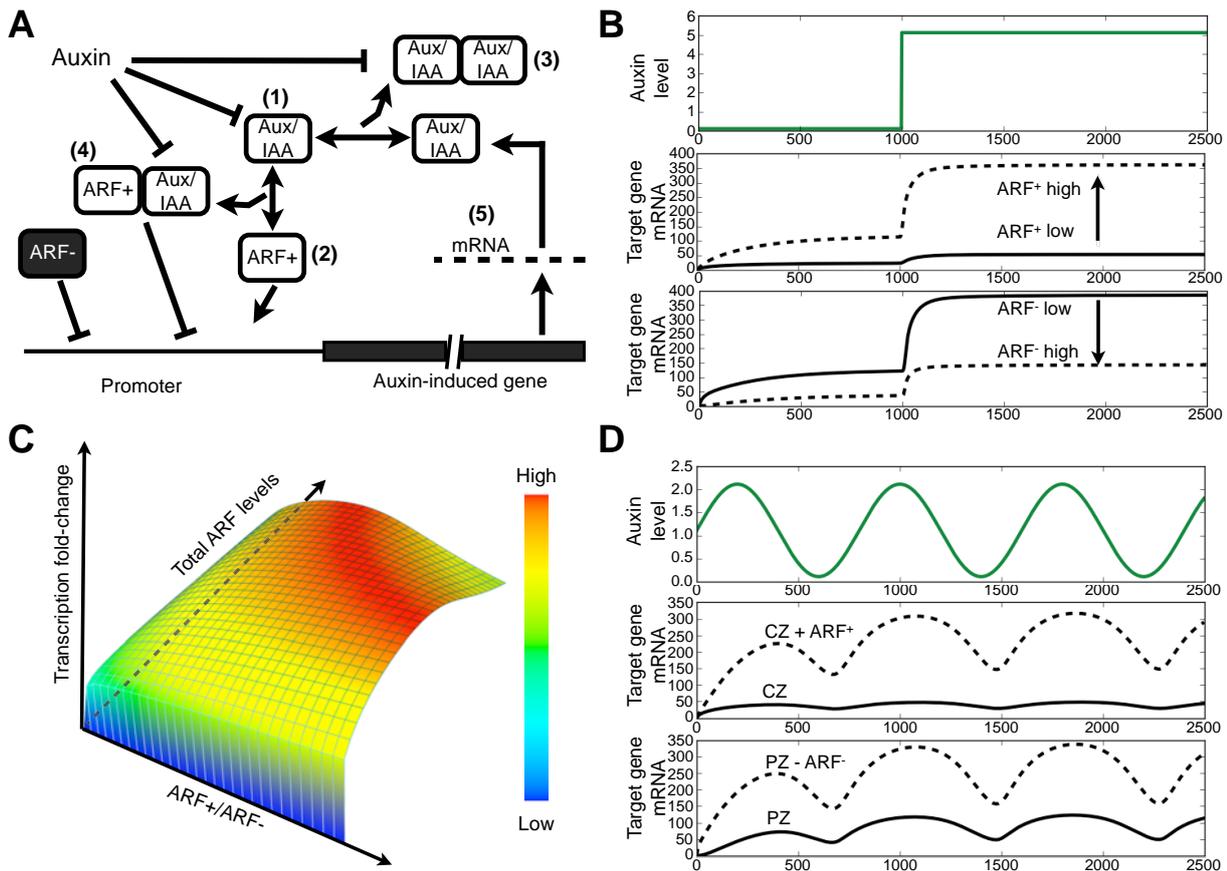


Figure 1.4: Mathematical model of the auxin signalling network. **(A)** Reaction scheme considered for the model. The numbers in brackets indicate the 5 populations of molecule described by ODEs in the model. **(B)** Effect of the level of ARF activators (middle panel) and ARF repressors (bottom panel) on target gene induction capacity upon an increase in auxin (upper panel). **(C)** 3D representation of the induction capacity as a function of ARF levels and ARF activator to ARF repressor ratio. The surface has been obtained by calculating the transcription fold-change, i.e. the ratio of the mRNA levels at equilibrium before and after a step increase in auxin. A colour map representing the parameter values is shown. **(D)** Effect of the level of ARF activators (middle panel) and ARF repressors (lower panel) on the stability of target gene induction upon varying auxin level. Transcription in response to sinusoidal changes in auxin levels (upper panel) has been studied. Two situations, corresponding to the centre (CZ) or the periphery (PZ), were considered. The effect of increasing ARF activators was tested for the first situation (CZ + ARF+) and of decreasing ARF repressors for the second (PZ - ARF-). For simplicity mRNA levels are shown in **(B,D)** for only the higher and the lower concentration of the variable parameter used in the simulation. See figures 4 and 5 of Note S1 for the full range of values.

2. The model reproduced gene activation in response to an increase in the auxin concentration x for all tested parameters. More generally, all variables of the system displayed stereotypical response curves upon variations in auxin concentration showing the robustness of our model.

Both the robustness and the ability to reproduce a biological observation support the plausibility of this model and prompted us to use it for investigating the role of auxin

signalling in SAM function.

2.4 The auxin signalling model allows predicting differential sensitivity to auxin and buffering capacities for the Aux/IAA-ARF pathway in the meristem

Our expression study suggested a simplified view of the meristem where the same population of Aux/IAs, ARF activators and ARF repressors exhibit a low expression at the centre and a high expression in the peripheral zone (figure 1.2). In order to study the significance of such a spatial distribution for auxin-regulated patterning, we used the auxin signalling model described above.

As could be intuitively expected, the ARF repressors limit the intensity of target gene induction by ARF activators in our model (figure 1.4B; see also Figure 4,5 in Note S1), which implies that the activation level of transcription in response to auxin (or more generally to an increase in the auxin signalling input) depends on both the absolute levels and the balance between ARF activators and repressors. This can be visualised by a 3D plot representing the gene induction levels in response to auxin as a function of both the absolute levels of ARF proteins (both ARF activators and repressors) and the ratio between ARF activators and repressors (figure 1.4C). For a given ARF activator-to-repressor ratio, the gene induction capacity increases with the absolute levels of ARF proteins, although with a slight decrease at the highest concentrations of proteins when this ratio is elevated. Based on these simulations we propose that, at the periphery of the meristem and notably in the organ primordia, the higher expression of *ARF* activators might allow for a high capacity of induction of gene transcription in response to auxin, despite the high expression of *ARF* repressors. On the other hand, the low expression of *ARF* activators at the centre of the SAM would create a low sensitivity to auxin. However, the fact that the expression of *ARF* repressors is also reduced at the centre of the SAM might allow these cells to not be completely insensitive to auxin. In addition, the distribution of the TIR1/AFB co-receptors could also contribute to creating a difference in sensitivity between the centre and the periphery of the SAM by reducing it in the L2 and L3 layers at the centre (lower TIR1 expression) and increasing it

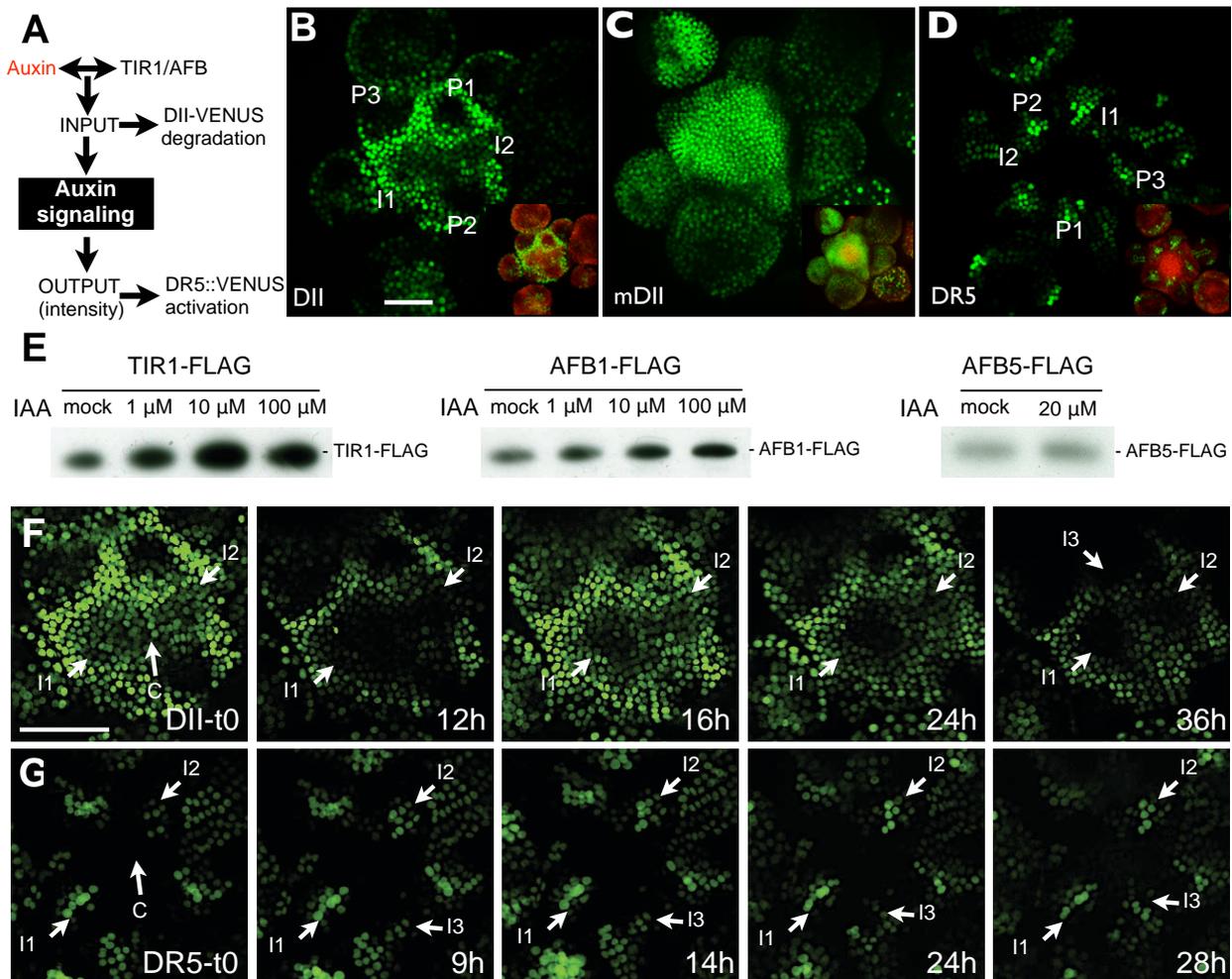


Figure 1.5: Spatial control and dynamics of auxin signalling at the inflorescence meristem. (A) Schematic representation of signalling parameters monitored by DII-VENUS as compared to DR5::VENUS. (B-D) Expression of DII-VENUS (B), mDII-VENUS (C) and DR5::VENUS (D) visualized by confocal microscopy. Insets: overlay of the VENUS signal (Green) with the autofluorescence signal (red). In (B) and (D) the 3 first primordia (P) are indicated and numbered from the youngest to the oldest. Two *initia* (I) are indicated and numbered from the oldest to the youngest following standard nomenclature. (E) Auxin-dependent binding of IAA28 domain II to TIR1/AFB auxin co-receptors. Anti-FLAG immunoblots of IAA28 domain II peptide pull-down assay with TIR1-FLAG, AFB1-FLAG or AFB5-FLAG. IAA treatments are as indicated. (F,G) Time-course of DII-VENUS (F) and DR5::VENUS (G) expression followed by confocal microscopy (VENUS fluorescence in green). Images were taken at indicated time after t0. The *initia* (I1 to 3) and the localization of the centre of the meristem (C) are indicated. Scale is the same in all images. Scale bar: 50 μ m.

in the internal tissues of the organ primordia (higher *AFB5* expression). In conclusion, we predict that the differential expression of the ARF activators and repressors in the meristem (figure 1.2) generates differences in auxin sensitivities between the centre (low sensitivity) and the periphery (high sensitivity), and thus leads to a differential expression of the *Aux/IAAs* (prediction 1).

Next, we sought to use our model to investigate the effects of altered auxin levels on the

auxin signalling, since it is known that auxin levels can be modified in response to various environmental factors (e.g. Gray *et al.* [1998]; Stepanova *et al.* [2008]; Tao *et al.* [2008]). We analysed the transcriptional response to changing auxin concentrations by simulating sinusoidal variations. These variations in auxin trigger oscillations in the transcriptional level of target genes for all the parameter tested (see Note S1). However, we observed that increasing the level of ARF repressors for a fixed level of ARF activators leads to an attenuation of the amplitude of the fluctuations. Decreasing the level of ARF activators for a fixed level of ARF repressors gave a similar result (see Note S1). This suggests that the ARF activator-to-repressor ratio can affect the stability of gene expression in response to auxin, or more generally to fluctuations in the auxin signalling input. This result can be related directly to our simplified representation of the SAM by simulating a situation equivalent to the centre of the meristem, where both ARF repressors and activators are present at similarly low levels. This simulation showed a low, but stable induction of target genes when the auxin input varies ("CZ" in figure 1.4D middle panel). Likewise a situation equivalent to the periphery of the meristem with high levels of both ARF activators and ARF repressors ("PZ" in figure 1.4D lower panel) also resulted in a relatively stable output. In both cases, stability was perturbed when the balance between positive and negative ARFs was altered (figure 1.4D). Increasing ARF activator levels when ARF repressor levels are low ("CZ + ARF+" in figure 1.4D middle panel), or decreasing ARF repressor levels when ARF activator levels are high ("PZ - ARF-" in figure 1.4D lower panel), caused strong fluctuations in target gene transcription levels. Thus, we predict that the signalling pathway buffers its response to the auxin input via the balance between ARF activators and repressors, in turn generated by their differential spatial distributions (prediction 2).

2.5 A novel signalling sensor, DII-VENUS, reports on the auxin signalling input in the meristem

To test the predictions from our model experimentally, we needed to assess both the input (auxin level and/or perception) and the output (target gene induction) of the signalling cascade. For measuring the transcriptional output, the widely used DR5 reporter which

consists of several ARF binding sites driving the expression of a reporter gene, is perfectly adapted (figure 1.5A) [Benková *et al.*, 2003; Heisler *et al.*, 2005; Sabatini *et al.*, 1999; Ulmasov *et al.*, 1997]. For assaying pathway input, given the absence of tools to visualize auxin *in situ*, we designed a novel auxin signalling sensor that relates more directly to auxin concentrations. This sensor comprises a constitutively expressed fusion of the auxin-binding domain (termed domain II or DII) [Dreher *et al.*, 2006; Tan *et al.*, 2007] of several Aux/IAA proteins (IAA8, 9 and 28: see material and methods; Vernoux *et al.*, in preparation) to a fast-maturing variant of YFP, VENUS [Nagai *et al.*, 2002]. By design, these fusion proteins should monitor local degradation of Aux/IAs, and thus the input in the signalling pathway depending on auxin levels and/or on local differences in auxin perception (figure 1.5A). Confocal imaging revealed that the strongest fluorescence signal was obtained using a nuclear-targeted VENUS sequence fused to the IAA28 domain II (Vernoux *et al.*, in preparation), which we henceforth refer to DII-VENUS.

DII-VENUS fluorescence could be detected in both shoot and root apical tissues where it showed clear differential distributions (figure 1.5B, Supplementary Figure S1.7A; Vernoux *et al.*, in preparation). Using root tissues, we showed that DII-VENUS abundance is dependent on auxin and the TIR1/AFB1-3 receptors (Vernoux *et al.*, in preparation). We also observed that the disruption of ubiquitin-dependent breakdown of Aux/IAA proteins using a proteasome inhibitor blocks the auxin-induced degradation of DII-VENUS in the SAM (Supplementary Figure S1.8). In addition, introducing a mutation in domain II of DII-VENUS (mDII-VENUS), which disrupts the interaction between Aux/IAs and the auxin co-receptors [Tan *et al.*, 2007] largely abolished the differential distribution of fluorescence in the SAM (figure 1.5C, Supplementary Figure S1.7B). We conclude that DII-VENUS abundance is regulated by auxin via TIR1/AFB activities, consistent with the model for Aux/IAA degradation [Chapman and Estelle, 2009].

We next tested auxin-dependent binding of the domain II of IAA28 to several TIR/AFB co-receptors, to estimate their relative contributions to DII-VENUS degradation. Pull-down experiments revealed that TIR1, AFB1 and AFB5 exhibited auxin-enhanced binding to domain II of IAA28 (figure 1.5E). Similarly to what had been observed for other Aux/IAs

[Parry *et al.*, 2009], the strongest binding was with TIR1 while the binding and the magnitude in the auxin-induced increase in binding with AFB1 appeared lower. Binding to IAA28 domain II was also low for AFB5. Hence, in the SAM, the homogenous expression of AFB1 (figure 1.2B) might not be sufficient to ensure a homogenous degradation capacity of DII-VENUS throughout the structure. The lower expression of TIR1 in the L2 and L3 layers at the centre of the meristem (figure 1.2A) might significantly diminish DII-VENUS degradation capacity in this part of the meristem, while *AFB5* might lead to a more limited increase of this capacity in the internal layers of primordia (figure 1.2D). Taken together, our results suggest that DII-VENUS reports auxin signalling input in the SAM. DII-VENUS thus likely provides information on auxin distribution but the differential of TIR1 and AFB5 needs to be taken into account.

2.6 Auxin signalling sensors distribution and dynamics confirm the model predictions

The degradation patterns displayed in the SAM from the DII-VENUS line indicated a high auxin signalling input in flower primordia and a low input in the cells immediately surrounding the primordia, in agreement with the organ-specific expression pattern of DR5::VENUS (figure 1.5B,D) [Heisler *et al.*, 2005]. A low DII-VENUS signal was also observed at the centre of both inflorescence and vegetative SAM (figure 1.5B, Supplementary Figure S1.7A). DII-VENUS is thus efficiently degraded at the centre of the meristem, despite the lower expression of TIR1 in most cells of this domain. This demonstrates a high auxin signalling input at the centre of the meristem, in contrast to the complete exclusion of DR5::VENUS expression from these cells (figure 1.5D), indicating that the signalling pathway limits gene activation in response to auxin at the meristem centre. These results confirm the prediction that ARF distribution in the meristem creates a differential sensitivity to auxin between the centre and the periphery (prediction 1), thus contributing to patterning of the meristem and higher expression of *Aux/IAAs* at the periphery of the SAM.

To test the buffering capacities of the signalling pathway (prediction 2), we next took advantage of the dynamic properties of DII-VENUS by carrying out live imaging experiments

to monitor DII-VENUS levels in real time (figure 1.5F) [Fernandez *et al.*, 2010]. Our initial intention was to perform auxin treatments to test the prediction. However, this turned out to be unnecessary as we observed notable fluctuations in DII-VENUS signal intensities that occurred naturally and without any treatments in the SAM. These variations were most evident in the epidermal (L1) layer in a zone connecting the centre to the sites of *initia* I1 and I2. This demonstrates the presence of important temporal variations in the overall auxin signalling input, especially in the centre and during the earlier steps of organ initiation. In contrast to DII-VENUS, the auxin signalling output, visualized with DR5::VENUS, did not show such global variations: the fluorescence remained stable in the organs after initiation and no signal was ever detected outside the organs (figure 1.5G; see notably I2 and I3). Since the VENUS protein does not appear to be unduly stable (Supplementary Figure S1.9), we conclude that the stable expression of DR5::VENUS over time reflects a stable output of the signalling pathway. Taken together, our results thus indicate that the Aux/IAA-ARF network in the SAM buffers fluctuations in the auxin signalling input and stabilises the activation of genes in response to auxin, thus providing biological evidence in support of the second prediction made by our model.

3 Discussion

It has been proposed that the plant could modulate auxin signalling during development by combining subsets of regulators in specific domains of tissues [De Rybel *et al.*, 2010; De Smet *et al.*, 2010; Muto *et al.*, 2007; Weijers *et al.*, 2005]. Our results, showing that ARF and Aux/IAAs are generally expressed at a lower level at the meristem centre and at a higher level at the periphery (figure 1.2) lead to an alternative scenario for the SAM, where the network is at least partially regulated by alterations in global expression levels. By combining expression data with the interactome, mathematical modelling and two auxin related markers (Figure 1), we provide evidence that this simple mode of regulation helps to establish a differential sensitivity between the centre and the periphery of the SAM that would be instrumental in translating dynamic auxin distributions into robust patterns.

A notable exception to this general regulatory principle at the SAM seems to be the

provascular tissue where specific expression of most notably *IAA20* and *30* could provide different signalling properties to the cells (figure 1.2). It has been demonstrated that *IAA20* and *30* lack a conserved domain II and are likely insensitive to auxin [Dreher *et al.*, 2006; Sato and Yamamoto, 2008]. They could thus locally diminish auxin sensitivity during the patterning of the provascular tissue. Further analyses of the corresponding mutants will likely help understand how these two non-canonical Aux/IAAs contribute to the development of the vasculature in the shoot apex.

In contrast to most Aux/IAAs, the expression of 10 of the 12 ARF active in the SAM is independent of auxin [Paponov *et al.*, 2008]. This indicates that the distribution of ARFs is most likely the primary factor controlling auxin sensitivity in the SAM and could be regulated by auxin-independent patterning factors. Recent evidence indicating that modifying the activities of two stem cell regulators, *CLAVATA3* and *WUSCHEL*, can affect the expression level of the auxin-inducible *DR5* reporter supports this hypothesis [Yadav *et al.*, 2010], although the precise link between these patterning genes and auxin signal transduction remains to be established.

We have also identified a differential regulation of the TIR1/AFB co-receptors in the SAM (figure 1.2). Genetic evidence indicates that the *TIR1/AFB* genes act redundantly in the regulation of auxin responses although with different contributions [Dharmasiri *et al.*, 2005b; Parry *et al.*, 2009]. Previous work has shown that the Aux/IAA proteins *IAA3* and *IAA7* exhibit a stronger interaction with TIR1 than with AFB1 [Parry *et al.*, 2009]. Our analysis of the interactions between TIR1/AFBs and the domain II of *IAA28* (figure 1.5) are consistent with these observations [Dharmasiri *et al.*, 2005b; Parry *et al.*, 2009] and supports the conclusion that at least one important component of the differences in TIR1/AFB co-receptor function relates to a broad variation in the affinity of their interactions with Aux/IAAs. Taken together, this suggests that each of the F-box co-receptors present in the SAM might be able to interact with the Aux/IAAs with a different affinity: high affinity for TIR1 and lower affinity for AFB1 and AFB5. Although further analysis will be needed to confirm these ideas, spatial changes in one of the TIR1/AFBs would therefore affect globally the capacity to degrade Aux/IAAs. The distribution of TIR1/AFB co-receptors in the SAM

would thus be expected to participate in the control of auxin sensitivity in parallel with the ARFs by lowering the Aux/IAA degradation capacity at the centre (lower expression of TIR1) while increasing it in the organs (higher expression of AFB5).

In our theoretical analysis of auxin signalling in the SAM (figure 1.4), we use a model that appears to be the simplest possible interpretation of our current understanding of auxin signal transduction. It is largely based on the rather simple topology of the network revealed by the Aux/IAA-ARF interactome (figure 1.3). In this network, the vast majority of the Aux/IAs interact with all the ARF activators, suggesting that most Aux/IAs have the potential to act as repressors of the transcriptional activity of ARF activators. In addition, the limited connectivity of ARF repressors suggests that the role of these transcription factors is essentially auxin-independent and that they might simply compete with the ARF activators for binding to the promoter of auxin-inducible genes. Whereas this general scenario most likely applies to the SAM, it is important to point out that more specific interactions might affect the dynamics of the ARF-Aux/IAA signalling pathway elsewhere in the plant. For instance, ARF9 is the only repressor that interacts with a high number of Aux/IAs thus suggesting a different mode of action for this ARF compared to the other repressors. Its weak homogenous expression in the SAM suggests that it likely does not play a dominant role in the SAM. However it might be dominant in other tissues and identification of the local network together with simulations using a modified version of our model would help unravelling a putative function in these tissues. A limited number of interactions between the other ARF repressors and some of the Aux/IAs have also been detected, even in the SAM-specific network. It is thus possible that such interactions participate in regulating the activity of the auxin signalling network but further analysis will be necessary to explore this possibility and understand its functional significance.

The development of the DII-VENUS sensor was instrumental in confirming the predictions made by our model (figure 1.5). Although the level of DII-VENUS depends on both absolute auxin concentrations and the levels of the TIR/AFB co-receptors, given the TIR1/AFB expression patterns DII-VENUS levels can be used to estimate relative auxin levels in the SAM. Both TIR1 and AFB1 appear to be homogenous throughout the periph-

ery of the SAM, while *AFB5* is expressed specifically in the internal tissue of the primordia (figure 1.2). Thus, given the low affinity of AFB5 for IAA28, the degradation of the DII-VENUS sensor in the organ primordia and its accumulation in the organ frontiers likely indicate a high concentration of auxin in the organ primordia and a low concentration in the surrounding cells. In the centre of the SAM, the lower TIR1 expression in the internal tissues should lead to less degradation of the DII-VENUS sensor and thus to an under-estimation of the auxin parameter. The low level of DII-VENUS in these cells is thus most likely due to a high auxin concentration at the centre of the SAM. Whereas high concentrations of auxin at young organ primordia have been predicted, for instance based on the high activity of the DR5-promoter [de Reuille *et al.*, 2006; Benková *et al.*, 2003; Heisler *et al.*, 2005; Reinhardt *et al.*, 2003; Smith *et al.*, 2006], the presence of auxin at the meristem summit was still a matter of debate. Low DR5 activity in the central zone seemed to indicate low auxin levels [Smith *et al.*, 2006], but this was contradicted by the patterns of PIN transporters at the SAM suggesting that auxin was preferentially transported to the meristem centre [de Reuille *et al.*, 2006]. Here we provide a simple explanation for this apparent discrepancy, showing that significant amounts of auxin do accumulate at the meristem centre, but without resulting in a high transcriptional output due to the low sensitivity there.

In conclusion, our work supports a key role for local auxin signalling in the regulation of patterning at the SAM. It also provides a plausible auxin distribution map in the structure that will need to be taken into account when further exploring the role of auxin in the SAM. In this context, it will be also important to further analyze the mechanisms controlling DII-VENUS level fluctuations in the SAM that might be linked to changes in auxin concentration and to explore a potential physiological role for these fluctuations.

4 Materials and methods

4.1 Plant material, growth conditions and plant treatments

All transgenic plants (see below) were generated in the Columbia (Col-0) ecotype of *Arabidopsis thaliana*. Plants expressing translational fusion of TIR1, AFB1, AFB2 and AFB3 to GUS have been described [Dharmasiri *et al.*, 2005b; Parry *et al.*, 2009]. For the live imaging, plants were grown on soil at 20°C in short-day conditions (8h light/16h darkness) for 4 weeks before being transferred in long-day conditions (16h light/8h darkness). The chemical treatments were done by immersing dissected inflorescence apices into liquid MS supplemented with the chemicals. Indole-acetic acid (IAA; Sigma) was dissolved in ethanol and used at the indicated final concentration. MG132 was dissolved in DMSO and used at the final concentration of 50 μ M for 2h30. For MG132/IAA co-treatments plants were pre-treated with MG132 for 1h30 before adding IAA and waiting an extra hour.

4.2 Analysis of gene expression

RNA in situ hybridization was performed as described [Vernoux *et al.*, 2000] with at least 3 independent experiments for each probe tested. Analysis of GUS expression in the inflorescence meristem was done according to Vernoux *et al.* [2000]. The RT-qPCR was performed on a Step one plus cycler (Applied Biosystems) using the SYBR green reagent kit (Roche). mRNA were extracted from around 100 dissected inflorescence meristems (with all the flowers older than the P1 stage removed) from soil-grown plants. The primers used are from Czechowski *et al.* [2004] except for ARF8, IAA29 and IAA32 (Table S3). Expression of several ARF genes could not be assessed due to non-reliable amplification (ARF1, 9, 15, 20, 21, 22 and 23). Expression of the *TCTP* gene (Table S3) was used as a standard and calculations were as described (Pfaffl, 2001). Expressions were analyzed on 2 independent mRNA extractions.

4.3 Confocal microscopy and live imaging

Live imaging of living shoot apical meristems was performed as described previously [Fernandez *et al.*, 2010]. Observations were done either on a LSM-510 laser-scanning confocal microscope (Zeiss, Jena, Germany) or a SP5 spectral detection confocal Microscope (Leica, Germany).

4.4 Generation of plasmids and transgenic plants

The DR5::VENUS plasmid has been described in Heisler *et al.* [2005]. The DII-VENUS and mDII-VENUS binary vectors were generated using Gateway technology and following the Multisite Gateway three-fragment vector construction kit instructions (Invitrogen). Full details of the cloning are given elsewhere (Vernoux *et al.*, in preparation). Briefly we used the degron region of IAA8, 9 and 28 starting from the conserved lysine to the end of domain II (IAA8: 107-178, IAA9: 120-195, IAA28: 28-61) [Dreher *et al.*, 2006]. To generate mDII-VENUS, we introduced the previously described P53L mutation [Rogg *et al.*, 2001] in the wild-type IAA28 degron sequence. These sequences were then fused in frame to VENUS tagged with the N7 nuclear localization signal [Heisler *et al.*, 2005] and put under the control of the strong constitutive 35S promoter. The different plasmids were then introduced in plants by floral dipping [Clough and Bent, 1998].

For the plasmids used for generating the pull-down data (see below), a plant expression vector containing a 3xFLAG tag was created by annealing the 3xFLAG-F and the 3xFLAG-R oligonucleotides (Table S3) and cloned into a *pFP101* binary vector (digested with *HindIII* and *BamHI* to remove the 2x35SPro). A Gateway[®] Cassette B (Invitrogen) was cloned in the Klenow filled-in *HindIII* site flanking the 3xFLAG fragment to form the *pFPGW-3xFLAG* vector. The coding sequences for TIR1, AFB1, and AFB5 (obtained from ABRC) were PCR-amplified to add Gateway attB sites (table S3) and recombined into pDONR P5-P2 Gateway entry vector (Invitrogen). The TIR1/AFBs were then recombined in a multi-gene Gateway reaction with pDONR P1P5r (Invitrogen) containing the cauliflower mosaic virus 35S promoter (CaMV35S) into the *pFPGW-3xFLAG* vector, in frame with the 3xFLAG tag, to create the plasmids *pFPGW-35S-TIR1/AFB-3xFLAG*. The TIR1/AFB-3xFLAG

sequence (without the promoter) was PCR-amplified from the pFPGW-35S-TIR1/AFB-3xFLAG plasmid by using oligonucleotides AsiSIAFB1, AsiSITIR1 and FLAGPmeI (Table S3) and cloned as an *AsiSI-PmeI* fragment into the pF3A WG (BYDV) Flexi[®] vector (Promega).

A 6xHIS tag was fused at the N-terminus of the ASK1 coding sequence by PCR using oligonucleotide *ASK1PmeI* successively with the nested primers HISASK1 and AsiSIHIS, and cloned as an *AsiSI-PmeI* fragment into the pF3A WG vector (generating the pF3A WG ASK1 vector).

The *35S:AFB5:3xFLAG* (AFB5-FLAG) transgenic line used for pull-down assays from plant extract (see below) was created by first transforming pFPGW-35S-AFB5-3xFLAG into *Agrobacterium tumefaciens* strain GV3101 [Koncz and Schell, 1986] by electroporation. Transgenic plants were generated using the floral dip method [Clough and Bent, 1998] and transformants were selected by seed coat fluorescence. In the T₂ generation, lines showing a 3:1 ratio of seed coat fluorescent:non-fluorescent plants, were selected for further study. Homozygous lines were selected from the T₃ generation.

4.5 *In vitro* transcription/translation of TIR1/AFB tagged proteins, immunoprecipitations and pull-down assays

For pull-down assays with TIR1-FLAG and AFB1-FLAG expressed in wheat germ extract, the pFPGW-35S-TIR1/AFB-3xFLAG and pF3A WG ASK1 plasmids were used with the TnT SP6 High-yield Wheat Germ Protein Expression System (Promega) to synthesize TIR1/AFB-3xFLAG co-expressed with HIS-ASK1 by *in vitro* transcription/translation (IVTT) in accordance with the manufacturer's instructions.

Pull-down assays with wheat germ expressed TIR1/AFB-3xFLAG were performed by combining 22 μ l of IVVT reaction extract with 6.5 μ g of biotinylated IAA28 domain II peptide (biotinyl-NH-EVAPVVGWPPVRSSRRN-COOH, synthesized by Thermo Scientific), 70 μ l 50% streptavidin-agarose suspension and 440 μ l of extraction buffer (EB; 0.15 M NaCl, 0.5% Nonidet P40, 0.1 M Tris-HCl pH 7.5, containing 1 mM phenylmethylsulphonyl fluoride, 1 μ M dithiothreitol, 10 μ M MG132 and 1 mg.ml⁻¹ BSA (Sigma) with auxin treatments as

indicated. The assays were incubated for 1 hour at 4°C with mixing then washed three times for 5 minutes in EB containing the auxin treatment.

For pull-down assays with plant-expressed AFB5-FLAG, extracts of 10 day old AFB5-FLAG seedling were made as described previously [Kepinski and Leyser, 2005] and used in pull-down assays by combining 2.5 mg of crude extract with 6.5 µg of biotinylated IAA28 domain II peptide, 70 µl 50% streptavidin-agarose suspension. The assays were incubated for 1 hour at 4°C with mixing then washed three times for 5 minutes in EB containing the auxin treatment.

The final processing of all pull-down assays including electrophoresis and western transfer were performed as described previously [Kepinski and Leyser, 2005]. The immunodetection of TIR1/AFB-FLAG was performed with a 1:10000 dilution of anti-FLAG 2-Peroxidase (HRP) antibody (Sigma) followed by chemiluminescent detection with ECL plus reagents (Amersham).

4.6 Protein interaction analyses

For generating the different constructs for the interaction analyses, full-length cDNA for all the *Aux/IAAs* and *ARFs* (except *ARF15* and *ARF21*) were either obtained from ABRC for the majority of them or RIKEN (for *ARF4*) or cloned by RT-PCR from cDNA libraries obtained from various tissues (for *ARF6*, *ARF14*, *IAA8*, *IAA9* and *IAA28*). Gateway technology was then used to generate the different constructs following the Multisite Gateway three-fragment vector construction kit instructions (Invitrogen). For the Y2H analysis, full length cDNAs were used for all the *Aux/IAA* and *ARF3*, 13 and 17 (ATG to Stop). Partial cDNAs encoding domain III and IV were used for the other *ARFs* and corresponded to the following region of the proteins: *ARF1*: 538-Stop; *ARF2*: 727-Stop; *ARF4*: 660-Stop; *ARF5*: 788-Stop; *ARF6*: 790-Stop; *ARF7*: 1030-Stop; *ARF8*: 699-Stop; *ARF9*: 519-Stop; *ARF10*: 575-Stop; *ARF11*: 500-Stop; *ARF12*: 500-Stop; *ARF14*: 497-Stop; *ARF16*: 570-Stop; *ARF18*: 482-Stop; *ARF19*: 948-Stop; *ARF20*: 484-Stop; *ARF22*: 498-Stop. The cDNAs or parts of cDNA were cloned directionally in pENTR/D-Topo (Invitrogen). They were then systematically transferred by recombination to a Gateway-compatible pACT2-

based vector (downstream of Gal4-AD) and a pGBKT7-based vector for the ARFs (downstream of Gal4-BD). The Y2H screen was performed using the mating method in a microtiter plate format as described [Boruc *et al.*, 2010]. Briefly, each interaction was tested in the two directions (each protein was used alternatively as a bait or as a prey) and using two independent reporters (*LacZ* and *HIS3*). *LacZ* activity was detected visually in the presence of 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal; Blue staining). *HIS3* activity was detected by the restoration of growth in the absence of histidine (See below for details on the analysis of the Y2H data). We also tested expression in yeast for all the constructs using standard Western blot analysis with an anti-HA antibody (clone 3F10, Roche) for the fusions to Gal4-AD and a Gal4-BD monoclonal antibody (Clontech) for the fusions to Gal4-BD. We confirmed expression for all the Gal4-BD fusions but a few Gal4-AD fusion were very weak or could not be detected indicating that the number of interactors might be underestimated for ARF9, ARF17, IAA11, IAA20 and IAA30. For BiFC analysis, we tested various interactions (see Supplementary Figure S1.4) between ARF5, ARF6, ARF9, ARF18 and/or the following stabilized Aux/IAAs: IAA8 P172S, IAA12 P74S, IAA17 P88L, IAA18 P101S and IAA28 P53L. These mutations have been described for IAA12, 17 and 28 [Hamann *et al.*, 2002; Rogg *et al.*, 2001; Rouse *et al.*, 1998] and similar mutations were introduced in IAA8 and 18. The full-length cDNAs were fused in frame to YFP molecule halves and interactions were tested by transient expression in *Nicotiana Benthamina* leaves as described [Desprez *et al.*, 2007]. Empty vectors were used as a negative control. Each interaction was tested independently 2 to 4 times and similar leaves areas were scanned for the different tests. The protein interactions from this publication have been submitted to the IMEx (<http://imex.sf.net>) consortium through IntAct (pmid 19850723) and assigned the identifier IM-15409.

4.7 Cluster analysis of Aux/IAA-ARF interaction network

4.7.1 Construction of the interaction network on the basis of the yeast-2-hybrid data

The output of the X-gal test took the form of a mark chosen from among the following ordered marks: -, +; ++; ++++. The output of the HIS3 test took the form of an optical density (OD) ratio reflecting the efficiency of recovery (ratio of the OD in the absence of histidine to the OD in the presence of histidine). The Spearman's rank correlation coefficient between the outputs of the X-gal and HIS3 tests was approximately 0.7, suggesting a partial but good agreement between these two tests. We thus chose to build a decision rule that exploits the redundancy between the X-gal and HIS3 tests and the two possible configurations for a given pair of proteins. This decision rule involved two thresholds, one for the X-gal and another for the HIS3 test. In order to determine these thresholds, we analyzed the empirical distributions of the optical density ratio (HIS3 test output) for each possible X-gal test output using Gaussian mixtures estimated with the `mclust` R package (<http://www.stat.washington.edu/fraley/mclust>). In particular, the empirical optical density ratio distribution for the '+' X-gal test mark was well fitted by a two-component Gaussian mixture where the component of lowest mean was interpreted as false positives. This interpretation was supported by the fact that the Gaussian component of lowest mean almost disappeared in the empirical optical density ratio distribution for the '++' and '+++' X-gal test marks (i.e. for higher stringency of the X-gal test). On the contrary it was over-represented for the '-' mark. Based on this statistical analysis, the thresholds were fixed between '-' and '+' for the X-gal test and at 0.45 for the HIS3 test. We finally defined the following decision rule in order to minimize false positives: there is interaction between proteins if at least an X-gal test and an HIS3 test are positive.

4.7.2 Cluster analysis

Structuring connectivity patterns were uncovered using a probabilistic clustering method implemented in the MixNet software [Daudin *et al.*, 2007; Picard *et al.*, 2009]. The key feature of the MixNet model is to give a probabilistic summary of the connectivity structure

by uncovering clusters of proteins that share the same connectivity profiles. Briefly, instead of directly describing the clustered structure of vertices, the MixNet model describes the topology of the network using connectivity probabilities $\pi_{q,\ell}$ i.e. the probability for a vertex from class q to be connected with a vertex from class ℓ . The protein interaction network was modelled as a random graph with $(X_{ij}; i, j = 1, \dots, N)$ representing its adjacency matrix, such that $X_{ij} = 1$ if vertices i and j are connected and 0 otherwise. The idea of the MixNet model is to consider that vertices can be spread into Q connectivity clusters that are hidden, with Q being unknown as well. The parameters of this model are the proportions of each cluster $\alpha = (\alpha_q; q = 1, \dots, Q)$, and the connectivity probability matrix $\Pi = (\pi_{q,\ell}; q, \ell = 1, \dots, Q)$. To this extent, Π is a summary of the connectivity of the protein interaction network at the cluster level. For a given number of clusters Q , the outputs of the MixNet algorithm are the estimated model parameters $\hat{\alpha}$ and $\hat{\Pi}$ and the probabilities of assignment of vertices to clusters $(\tau_{iq}; i = 1, K, N; q = 1, K, Q)$. The posterior distribution $(\tau_{iq}; q = 1, \dots, Q)$ represents the probabilistic assignment of vertex i to the clusters.

4.7.3 Analysis of the adequacy of the clustering

We assessed the adequacy of the clustering obtained by the MixNet algorithm by evaluating the separability of the clusters and the dispersion of the proteins within the clusters. Since, in our case, the assignment of proteins to clusters is almost deterministic (i.e. $\tau_{iq} \approx 1$ for a unique cluster q and $\tau_{i\ell} \approx 0$ for $\ell \neq q$), this assignment can be viewed as a partition. The model parameters $(\pi_{q\ell})$, which parameterized the edges of the graph, cannot be used directly to define dispersion measures of the vertices assigned to a given cluster. We thus used the edges incident to the vertices to derive dissimilarity measures for the vertices using the adjacency information. The Sokal-Michener distance between vertices i and j defined as $D_{i,j} = \sum_k \mathbf{I}(x_{ik} \neq x_{jk})/N$, where \mathbf{I} denotes the indicator function, is the proportion of mismatches or disagreement between the i^{th} and j^{th} rows of the adjacency matrix. This distance naturally expresses the difference in connectivity profiles between vertices. The

distance between vertex i and cluster q is then given by:

$$D(i, q) = \frac{\sum_{j \neq i} \tau_{jq} \sum_k \mathbf{I}(x_{ik} \neq x_{jk})}{\left(\sum_{j \neq i} \tau_{jq}\right) N}$$

If the vertices are deterministically assigned to a given cluster, this distance simplifies to

$$D(i, q) = \frac{\sum_{j \in q; j \neq i} \sum_k \mathbf{I}(x_{ik} \neq x_{jk})}{(n_q - 1) N}, \text{ if } i \in q$$

$$D(i, q) = \frac{\sum_{j \in q} \sum_k \mathbf{I}(x_{ik} \neq x_{jk})}{n_q N}, \text{ if } i \notin q$$

where n_q is the number of elements of cluster q .

The distance between cluster q and cluster ℓ is given by:

$$D(q, \ell) = \frac{\sum_{i, j; i \neq j} \tau_{iq} \tau_{jl} \sum_k \mathbf{I}(x_{ik} \neq x_{jk})}{\left(\sum_{i, j; i \neq j} \tau_{iq} \tau_{jl}\right) N}$$

If the vertices are deterministically assigned to a given cluster, this distance simplifies to

$$D(q, q) = \frac{\sum_{i, j \in q; i \neq j} \sum_k \mathbf{I}(x_{ik} \neq x_{jk})}{n_q(n_q - 1) N}$$

$$D(q, \ell) = \frac{\sum_{i \in q} \sum_{j \in \ell} \sum_k \mathbf{I}(x_{ik} \neq x_{jk})}{n_q n_\ell N}, \text{ if } q \neq \ell$$

4.7.4 Graphical representation of proteins preserving pairwise distances between them

We applied a multidimensional scaling (MDS) method [Hastie *et al.*, 2009] to visualize the proteins. Briefly, the MDS approach allowed us to find a low-dimensional projection of the data such as to preserve, as closely as possible, the pairwise distances between data points $\{D(i, j); i, j = 1, \dots, N\}$.

4.8 Modelling of the Aux/IAA-ARF signalling pathway

4.8.1 Model description

Except for gene transcription, we assumed that all reactions involved in this network are well described using mass action law kinetics. Variables are denoted using upper case characters, and parameters with lower case. Let I and A denote the concentrations of the IAA and ARF proteins respectively, and D_{II} and D_{IA} the concentration of the dimers formed by these proteins. R denotes the total concentration of mRNA that is transcribed from a pool of auxin responsive genes. This set of genes includes Aux/IAs, so that I has a production rate proportional to R . Auxin is represented by the variable x appearing in decay rates involving Aux/IAA proteins. The system can be written as follows:

$$\begin{aligned}\frac{dI}{dt} &= \pi_I R + 2k'_{II} D_{II} - 2k_{II} I^2 + k'_{IA} D_{IA} - k_{IA} I A + \delta_{II}(x) D_{II} - \delta_I(x) I \\ \frac{dA}{dt} &= \pi_A + k'_{IA} D_{IA} - k_{IA} I A + \delta_{IA}(x) D_{IA} - \delta_A A \\ \frac{dD_{II}}{dt} &= k_{II} I^2 - (k'_{II} + \delta_{II}^* + \delta_{II}(x)) D_{II} \\ \frac{dD_{IA}}{dt} &= k_{IA} I A - (k'_{IA} + \delta_{IA}^* + \delta_{IA}(x)) D_{IA} \\ \frac{dR}{dt} &= h(I, A, D_{IA}) - \delta_R R\end{aligned}$$

where h is the transcription rate of the target genes, which depends on the levels of the different transcription factors. This function was determined based on thermodynamic assumptions, as described e.g. in Bintu *et al.* [2005]:

$$h(I, A, D_{IA}) = \frac{1 + \frac{f}{B_d} + A \left(1 + \frac{f_A \omega_A}{B_d} A\right)}{1 + \frac{A}{B_d} \left(1 + \frac{\omega_A}{B_d} A\right) + \frac{\omega_I}{K_d B_d} A I + \frac{\omega_D}{B_d} D_{IA} + \kappa_A^-}$$

These equations include the following parameters:

- π_I is the rate of translation of mRNA (R) into Aux/IAA (I).
- π_A is the production rate of ARF activators, supposed constant.
- The parameters k'_{IX} and k_{IX} for $X \in I, A$ respectively denote dissociation and associ-

ation rates of the dimer forming reactions.

- κ_A^- represents a background level of transcriptional repression, due to ARF repressors.
- f and f_A represent the strength of transcriptional enhancement due to a single ARF activator and two ARF activators being bound to the promoter, respectively.
- K_d (B_d respectively) is the dissociation constant of the ARF:Aux/IAA dimerization reaction (resp. ARF to promoter binding reaction). Hence, $K_d = k'_{IA}/k_{IA}$.
- The coefficients ω_A , ω_I and ω_{DA} represent cooperativity effects induced by the binding of two ARF activators (ω_A) on the promoter or by ARF:Aux/IAA dimer formation (ω_I and ω_{DA}). For the latter, the occurrence of two terms, proportional to IA and D_{IA} , is due to the fact that ARF and Aux/IAA may dimerise either prior to DNA binding, or on the promoter after ARF binds to DNA.
- The parameters of the form δ_X denote degradation rates. In case of Aux/IAA proteins, the latter is auxin dependent. Then, $\delta_I(x)$ takes the form of saturating functions of the auxin level : $\delta_I(x) = \gamma_I \delta_I \frac{Kx}{1 + Kx}$, which was obtained by applying a quasi-steady state assumption to the equations governing the kinetics of auxin mediated degradation (see Note S1). In this function, the parameters δ_I , γ_I and K respectively represent the basal decay rate of Aux/IAs, the maximum fold increase in decay rate that can be induced by auxin and the affinity of auxin to TIR1/AFB co-receptors.
- As mentioned above, we have also assumed that even bound in a dimer, Aux/IAA proteins are accessible to auxin and degraded, freeing the other protein involved in the dimer. Hence the terms $\delta_{II}(x)$ and $\delta_{IA}(x)$ in the equations above, which take the same form as $\delta_I(x)$. An additional decay rate of the form δ_{II}^* has been included as well, to account for the natural decay of the dimers.

4.8.2 Numerical simulations

All simulations were performed using tools from the scientific library Scipy (www.scipy.org) of the programming language Python (www.python.org). More specifically, after preliminary

tests of different solvers, we have chosen to use the `odeint` routine, which relies on `lsoda`, an adaptive step-size solver for stiff and non-stiff systems from the FORTRAN library `odepack`. All figures were produced using the `Matplotlib` (2D) and `Mayavi2` (3D) libraries. A python script containing the main routines can be provided upon request. The model is also available with this paper as an SBML file and has been deposited in the `Biomodels` database (model accession number: MODEL1105290000).

Acknowledgements

We would like to thank M. Heisler for the DR5::VENUS plasmid; A. Miyawaki for VENUS; Y. Jacob for the yeast 2-hybrid destination vectors; A. Chaboud and B. Blanquier (IFR128) for advice on the yeast 2-hybrid and qPCR respectively; F. Rozier for help with *in situ* hybridization; M. Kieffer for assistance with cloning; C. Chamot and C. Lionnet (IFR128 PLATIM) for confocal microscopy; A. Boudaoud, O. Hamant, F. Parcy, I. Bohn-Courseau, A.P. Mahönen, D. Weijers, F. Besnard, V. Mirabet, F. Monéger and all the members of the Meristem team for discussions on the manuscript. T.V was supported by the Human Frontier Science Program Organization (CDA 0047/2007 HFSP0) and the Agence National de la Recherche (AuxFate ANR-07-JCJC-0115); F.P. by the Agence Nationale de la Recherche (NeMo ANR-08-BLAN-0304-01); L.D.V., J.T. and T.V. by an E.U. training Grant (MRTN-CT2004-005336 - SY-STEM); L.A and S.K by the Biotechnology and Biological Research Council (BB/F007418/1 and BB/F013981/1); L.D.V. is a postdoctoral fellow of the Research Foundation-Flanders.

Author contributions

T.V., G.B. and E.F. contributed equally to this work. T.V., G.B., V.M., S.G., R.K. and J.H.D. did the expression analysis. G.P. and M.E. constructed and characterized the GUS lines for the auxin receptors. H.V.D., G.B, V.M, C.C. and L.D.V. performed the protein interaction analysis. J.L., F.P. and Y.G. analyzed the data and the structure of the interactome. E.F. did the ODE modelling and interpretations were done by E.F. and T.V.

T.V, G.B., M.O., A.L. and D.W. characterized the DII-VENUS sensor. G.B, M.O. and P.D. performed the live imaging in the shoot meristem. L.A. and S.K performed the pull-down analyses. T.V. designed the project and most of the experiments together with E.F, Y.G, J.H.D, S.K., M.E., C.G., M.B., L.D.V and J.T. T.V., E.F. and J.T. wrote the manuscript with input from the other authors. T.V. is corresponding author.

Supplementary Materials

Supplementary text: Model of the ARF-Aux/IAA signalling network.

See separate file NoteS1.pdf, available at <http://onlinelibrary.wiley.com/doi/10.1038/msb.2011.39/suppinfo>.

Supplementary material: SBML version of the ARF-Aux/IAA signalling model.

See separate file AuxinSignalingODEmodel.xml, available at <http://onlinelibrary.wiley.com/doi/10.1038/msb.2011.39/suppinfo>.

Table S1.1: Results of the yeast two-hybrid analysis.

See separate file SuppTable1.xls, available at <http://onlinelibrary.wiley.com/doi/10.1038/msb.2011.39/suppinfo>.

Table S1.2: Summary of Aux/IAA-ARF interactions tested in the literature.

Interaction	Results	Method used	References	Full data
IAA1/IAA1	positive	Y2H/His3 - LacZ	(1)	positive
IAA1/IAA1	positive	Y2H/LacZ	(2)	
IAA1/IAA2	positive	Y2H/His3 - LacZ	(1)	positive
IAA1/IAA3	positive	Y2H/His3 - LacZ	(1)	positive
IAA1/IAA4	positive	Y2H/His3 - LacZ	(1)	positive
IAA1/IAA8	positive	Y2H/His3 - LacZ	(1)	positive
IAA1/IAA9	positive	Y2H/His3 - LacZ	(1)	positive
IAA1/IAA16	positive	Y2H/His3 - LacZ	(1)	positive
IAA1/IAA17	positive	Y2H/His3 - LacZ	(1)	positive
IAA1/IAA18	positive	Y2H/His3 - LacZ	(1)	positive
IAA1/IAA19	positive	Y2H/His3 - LacZ	(1)	positive
IAA1/IAA20	positive	Y2H/His3 - LacZ	(1)	negative
IAA2/IAA2	positive	Y2H/His3 - LacZ	(1)	positive
IAA3/IAA17	positive	Y2H/LacZ	(3)	positive
IAA6/IAA6	positive	Y2H/LacZ	(2)	negative
IAA12/IAA12	positive	Y2H/LacZ	(4)	positive
IAA12/IAA12	positive	BiFC	(5)	
IAA13/IAA13	positive	Y2H/LacZ	(2)	positive

IAA17/IAA17	positive	Y2H/LacZ	(3)	positive
IAA19/IAA19	positive	Y2H/LacZ	(2)	positive
IAA19/IAA19	20% dimerization ¹	FCCS	(6)	
ARF1/ARF1	positive ²	Y2H/LacZ	(3)	negative
ARF5/ARF1	positive ²	Y2H/LacZ	(3)	negative*
ARF5/ARF1	negative	Y2H/LacZ	(7)	
ARF5/ARF2	negative	Y2H/LacZ	(7)	negative
ARF5/ARF4	negative	Y2H/LacZ	(7)	negative
ARF5/ARF5	positive ²	Y2H/LacZ	(7)	
ARF5/ARF5	positive ²	Y2H/LacZ	(2)	negative*
ARF5/ARF5	No association	FCCS	(6)	
ARF5/ARF6	negative	Y2H/LacZ	(7)	negative
ARF5/ARF7	positive	Y2H/LacZ	(7)	negative*
ARF5/ARF7	<20% dimerization ³	FCCS	(6)	
ARF5/ARF8	negative	Y2H/LacZ	(7)	negative
ARF5/ARF9	negative	Y2H/LacZ	(7)	negative
ARF5/ARF11	negative	Y2H/LacZ	(7)	negative
ARF7/ARF1	negative	Y2H/LacZ	(7)	negative
ARF7/ARF2	negative	Y2H/LacZ	(7)	negative
ARF7/ARF4	negative	Y2H/LacZ	(7)	negative
ARF7/ARF6	positive	Y2H/LacZ	(7)	negative
ARF7/ARF7	positive	Y2H/LacZ	(7)	
ARF7/ARF7	<20% dimerization ³	FCCS	(6)	negative*
ARF7/ARF7	positive ²	Y2H/LacZ	(2)	
ARF7/ARF8	positive ²	Y2H/LacZ	(7)	negative
ARF7/ARF9	negative	Y2H/LacZ	(7)	negative
ARF7/ARF11	negative	Y2H/LacZ	(7)	negative
ARF8/ARF8	negative	Y2H/LacZ	(2)	negative

ARF1/IAA12	positive	Y2H/LacZ	(8)	negative
ARF1/IAA13	positive	Y2H/LacZ	(8)	negative
ARF1/IAA17	positive	Y2H/LacZ	(3)	negative
ARF5/IAA1	positive	Y2H/LacZ	(2)	positive
ARF5/IAA3	positive	Y2H/LacZ	(9)	positive
ARF5/IAA6	positive	Y2H/LacZ	(2)	negative
ARF5/IAA12	positive	Y2H/LacZ	(4)	
ARF5/IAA12	positive	Y2H/LacZ	(7)	positive
ARF5/IAA12	positive	Y2H/LacZ	(9)	
ARF5/IAA12	positive	BiFC	(5)	
ARF5/IAA13	positive	Y2H/LacZ	(2)	positive
ARF5/IAA14	positive	Y2H/LacZ	(10)	positive
ARF5/IAA17	positive	Y2H/LacZ	(3)	positive
ARF5/IAA19	positive	Y2H/LacZ	(2)	positive
ARF5/IAA19	100% dimerization	FCCS	(6)	
ARF6/IAA1	positive	Y2H/His3 - LacZ	(1)	positive
ARF7/IAA1	positive	Y2H/His3 - LacZ	(1)	positive
ARF7/IAA1	positive	Y2H/LacZ	(2)	
ARF7/IAA6	positive	Y2H/LacZ	(2)	negative
ARF7/IAA12	positive	Y2H/LacZ	(7)	positive
ARF7/IAA13	positive	Y2H/LacZ	(2)	positive
ARF7/IAA14	positive	Y2H/LacZ	(10)	positive
ARF7/IAA18	positive	Y2H/LacZ	(11)	positive
ARF7/IAA19	positive	Y2H/LacZ	(2)	
ARF7/IAA19	positive	Pull-down	(2)	positive
ARF7/IAA19	100% dimerization	FCCS	(6)	
ARF8/IAA1	positive	Y2H/LacZ	(2)	positive
ARF8/IAA6	positive	Y2H/LacZ	(2)	negative

ARF8/IAA13	positive	Y2H/LacZ	(2)	positive
ARF8/IAA19	positive	Y2H/LacZ	(2)	positive
ARF11/IAA1	positive	Y2H/His3 - LacZ	(1)	negative
ARF19/IAA3	positive	Y2H/LacZ	(9)	positive
ARF19/IAA12	positive	Y2H/LacZ	(9)	positive
ARF19/IAA14	positive	Y2H/LacZ	(10)	positive
ARF19/IAA18	positive	Y2H/LacZ	(11)	positive

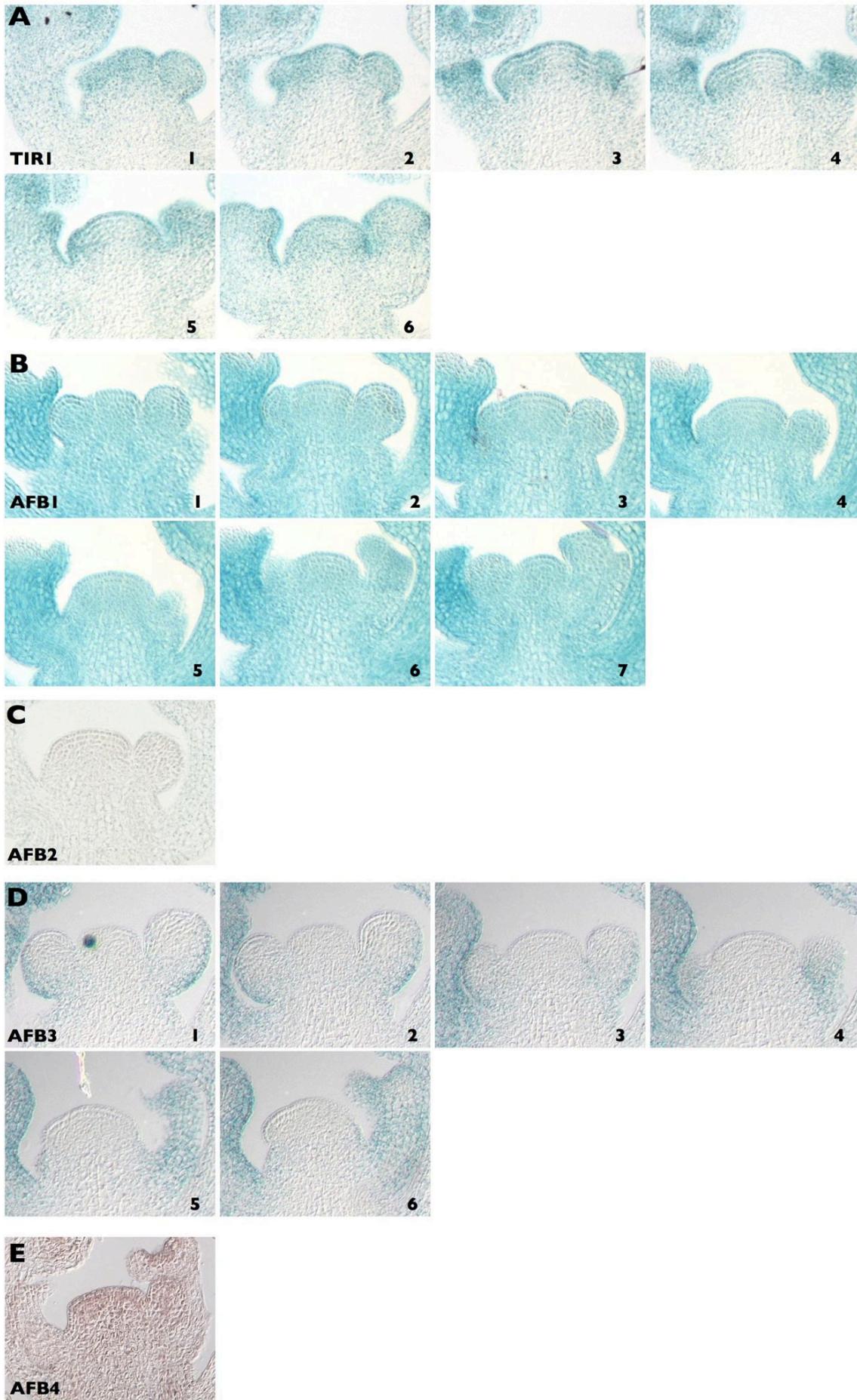
The method used to test the interactions is indicated along with the reporter systems for yeast two-hybrid (Y2H). Divergent interaction results in the full interactome analysis are indicated in bold and * indicates divergent results in the literature. ¹ Statistically different from all controls; ² Low LacZ activity; ³ Not statistically different from all controls.

1. Kim, J., Harter, K. and Theologis, A. (1997). Protein-protein interactions among the Aux/IAA proteins 136. *PNAS*, **94**, 11786–11791.
2. Tatematsu, K., Kumagai, S., Muto, H., Sato, A., Watahiki, M. K., Harper, R. M., Liscum, E. and Yamamoto, K. T. (2004). MASSUGU2 encodes Aux/IAA19, an auxin-regulated protein that functions together with the transcriptional activator NPH4/ARF7 to regulate differential growth responses of hypocotyl and formation of lateral roots in *Arabidopsis thaliana*. *The Plant Cell*, **16**(2), 379–93.
3. Ouellet, F., Overvoorde, P. J. and Theologis, A. (2001). IAA17/AXR3: biochemical insight into an auxin mutant phenotype. *The Plant Cell*, **13**, 829–841.
4. Hamann, T., Benkova, E., Bäurle, I., Kientz, M. and Jürgens, G. (2002). The *Arabidopsis* BODENLOS gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning. *Genes and Development*, **16**, 1610–1615.
5. Szemenyei, H., Hannon, M. and Long, J. A. (2008). TOPLESS mediates auxin-dependent transcriptional repression during *Arabidopsis* embryogenesis. *Science*, **319**, 1384–1386.
6. Muto, H., Nagao, I., Demura, T., Fukuda, H., Kinjo, M. and Yamamoto, K. T. (2006). Fluorescence cross-correlation analyses of the molecular interaction between an Aux/IAA

- protein, MSG2/IAA19, and protein-protein interaction domains of auxin response factors of arabidopsis expressed in HeLa cells. *Plant & Cell Physiology*, **47**(8), 1095–101.
7. Hardtke, C. S., Ckurshumova, W., Vidaurre, D. P., Singh, S. A., Stamatiou, G., Tiwari, S. B., Hagen, G., Guilfoyle, T. J. and Berleth, T. (2004). Overlapping and non-redundant functions of the Arabidopsis auxin response factors MONOPTEROS and NONPHOTOTROPIC HYPOCOTYL 4. *Development*, **131**(5), 1089–1100.
 8. Ulmasov, T., Murfett, J., Hagen, G. and Guilfoyle, T. J. (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *The Plant Cell*, **9**, 1963–1971.
 9. Weijers, D., Benkova, E., Jäger, K. E., Schlereth, A., Hamann, T., Kientz, M., Wilmoth, J. C., Reed, J. W. and Jürgens, G. (2005). Developmental specificity of auxin response by pairs of ARF and Aux/IAA transcriptional regulators. *The EMBO Journal*, **24**, 1874–1885.
 10. Fukaki, H., Nakao, Y., Okushima, Y., Theologis, A. and Tasaka, M. (2005). Tissue-specific expression of stabilized SOLITARY-ROOT/IAA14 alters lateral root development in Arabidopsis. *Plant Journal*, **44**, 382–395.
 11. Uehara, T., Okushima, Y., Mimura, T., Tasaka, M. and Fukaki, H. (2008). Domain II mutations in CRANE/IAA18 suppress lateral root formation and affect shoot development in Arabidopsis thaliana. *Plant & Cell Physiology*, **49**, 1025–1038.

Table S1.3: Oligonucleotides used in this study.

Targets names	Forward	Reverse
RT-qPCR		
ARF8	5' TGGCAGCTTGTATTCGTTGA3'	5' CCAAACGTTATTCACAAATGACTC3'
IAA29	5' CCGAGTCTTCAATAGTTTACGATG3'	5' CGAATATGACGATGATGATAACTACC3'
IAA32	5' AGGGTGGTGGGGTAATCG3'	5' CATACCCGTCGAGACCTATCAT3'
TCTP	5' ACACCCAAGCTCAGCGAAGAA3'	5' CATGCATACCCTCCCCAACAA3'
<i>In vitro</i> transcription/translation of TIR1/AFB tagged proteins		
3xFLAG	5' AGCTTAGACTACAAAGACCATGACGGTG ATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGTGATATCG3'	5' GATCCGATATCACTTGTGCATCGTCATC CTTGTAATCGATGTCATGATCTTTATAATC ACCGTCATGGTCTTTGTAGTCTA3'
TIR1; F: attB5; R: attB2	5' GGGGACAACCTTTGTATACAAAAGTTGAA GCATGCAGAAGCGAATAGC3'	5' GGGGACCACTTTGTACAAGAAAGCTGGG TATAATCCGTTAGTAGTAA3'
AFB1; F: attB5; R: attB2	5' GGGGACAACCTTTGTATACAAAAGTTGAA GCATGGGTCTCCGATCCCACC3'	5' GGGGACCACTTTGTACAAGAAAGCTGGG TACTTTATGGCTAGATGTG3'
AFB5; F: attB5; R: attB2	5' GGGGACAACCTTTGTATACAAAAGTTGAA GCATGACACAAGATCGCTC3'	5' GGGGACCACTTTGTACAAGAAA GCTGGGTATAAAATCGTGACGAACCTTTGGT GC3'
TIR1-3xFLAG	5' ATTGCGATCGCATGCAGAAGCGAATAGC C3' (AsiSITIR1)	5' ACCGTTTAAACTCACTTGTGCATCGTCAT CC3' (FLAGPmeI)
AFB1-3xFLAG	5' ATTGCGATCGCATGGGTCTCCGATTCCC 3' (AsiSIAFB1)	(FLAGPmeI)
AFB2-3xFLAG	5' ACCGCGATCGCATGAATTATTTCCAG3' (AsiSIAFB2)	(FLAGPmeI)
6xHISASK1	5' GCATCATCACCATCACCATATGTCTGCG AAGAAGATTGT3' (HISASK1)	5' AGCTTTGTTTAAACTCATTCAAAGCCC ATTGGTTCTC3' (ASK1PmeI)
6xHISASK1	5' GTGTGCGATCGCCATGCATCATCACCAT CACCAT3' (AsiSIHIS)	(ASK1PmeI)



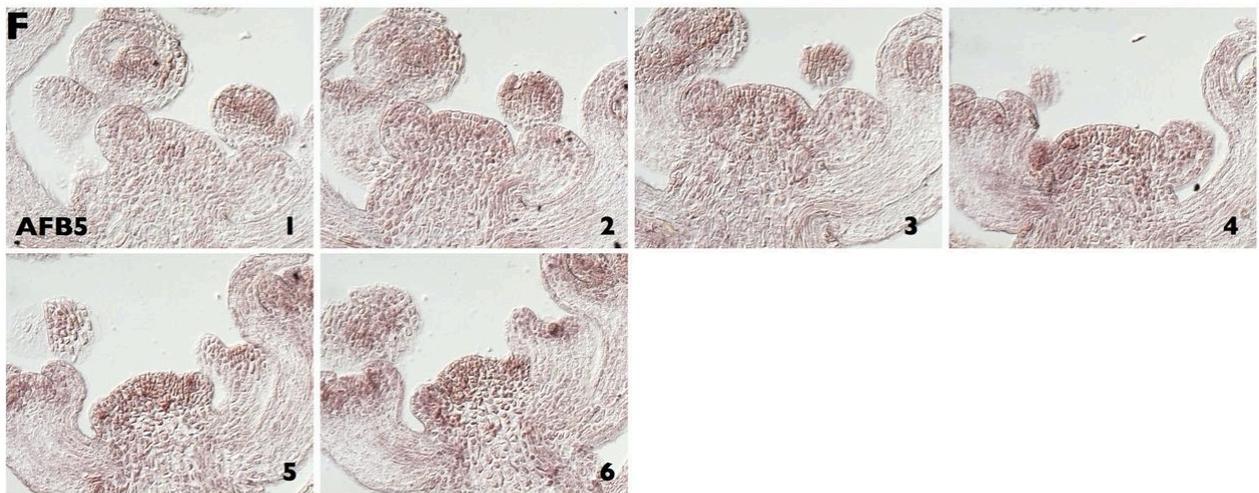
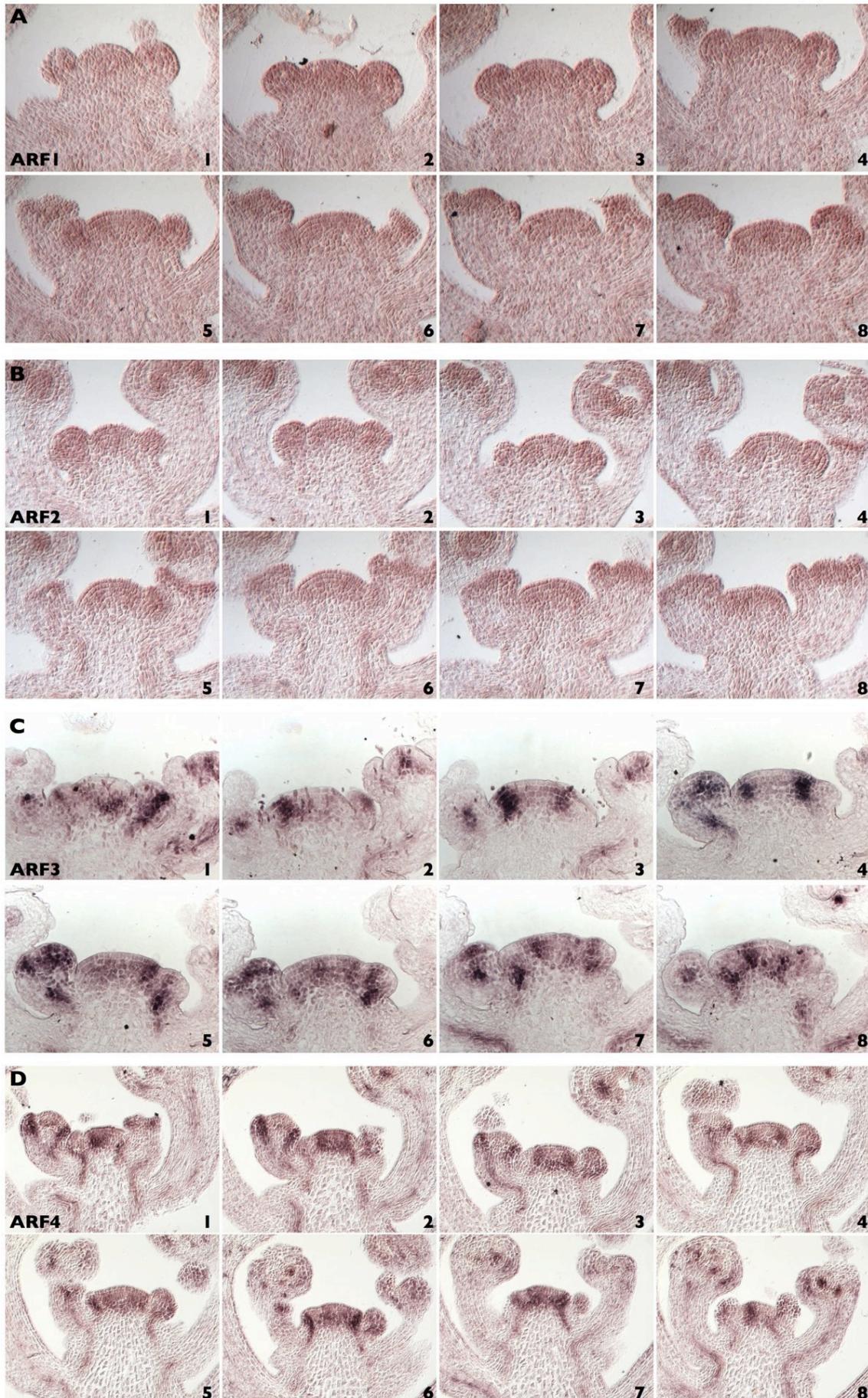
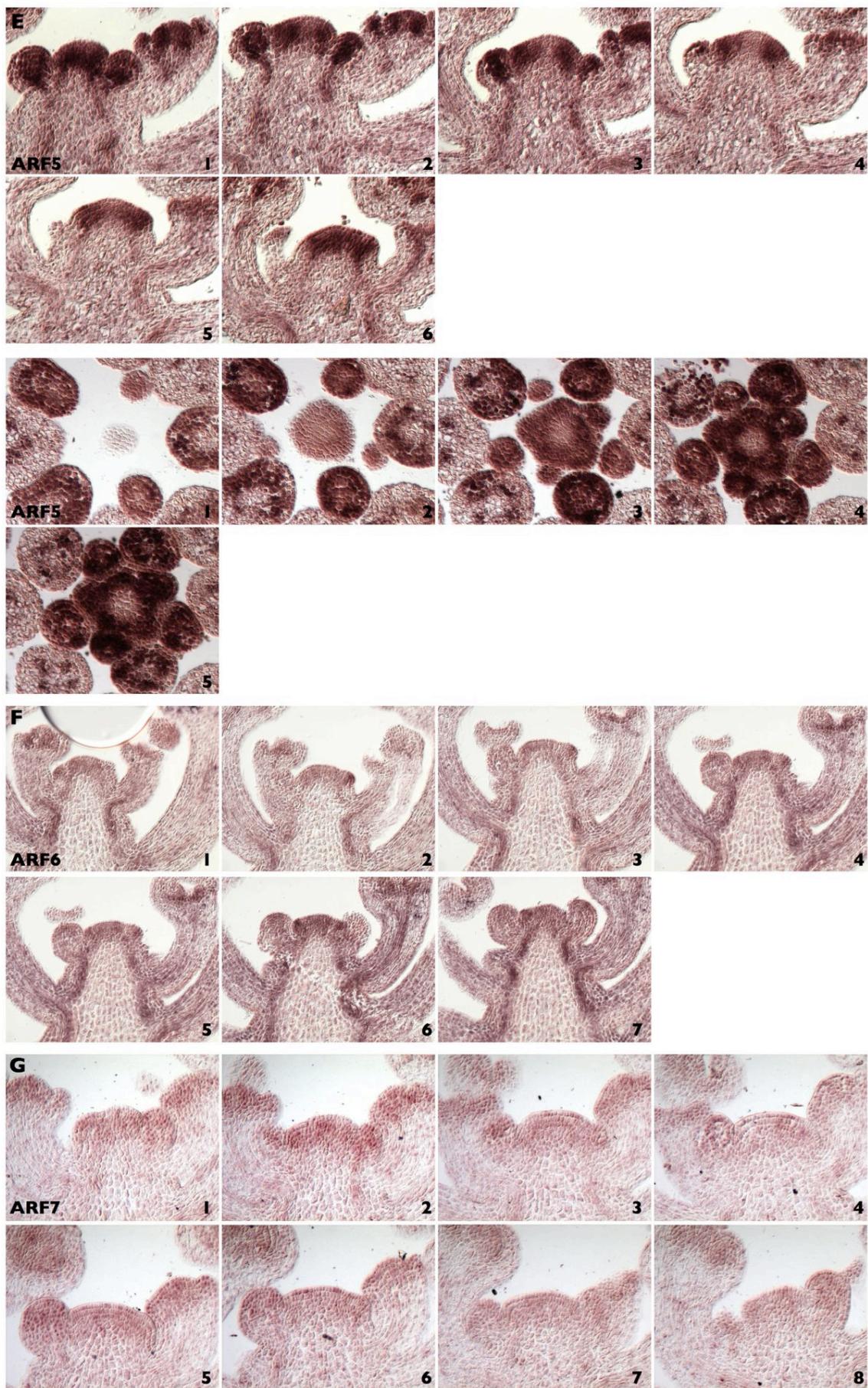
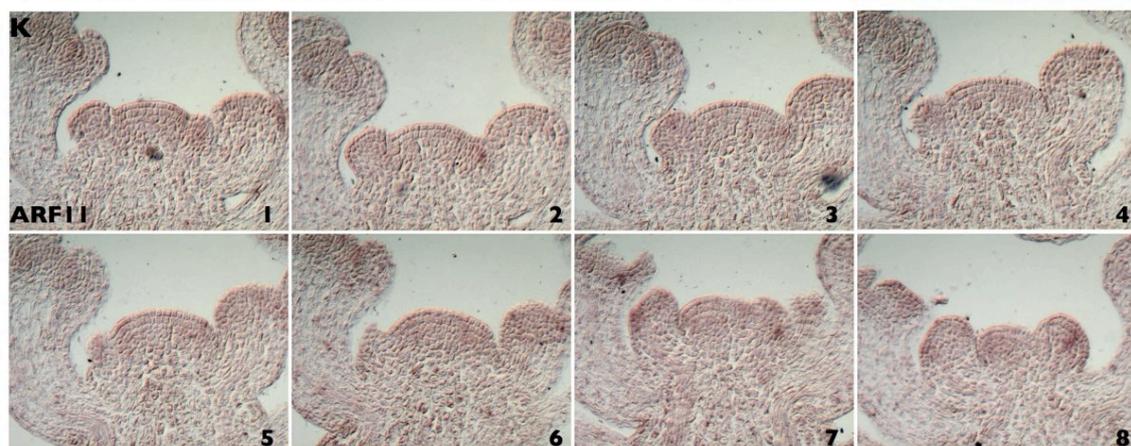
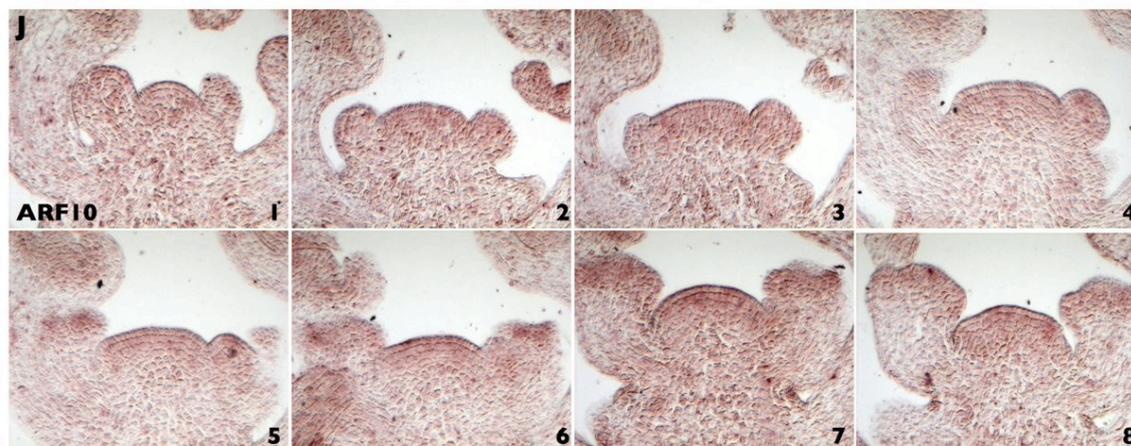
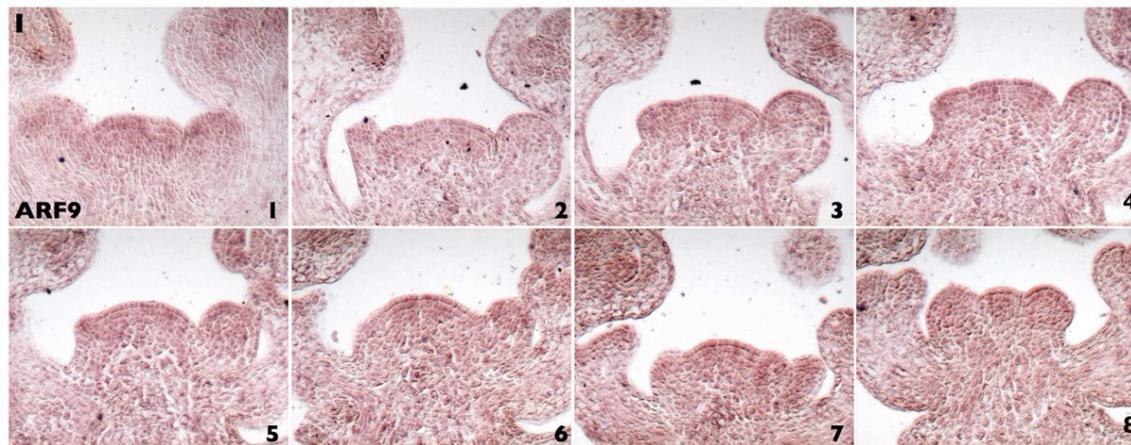
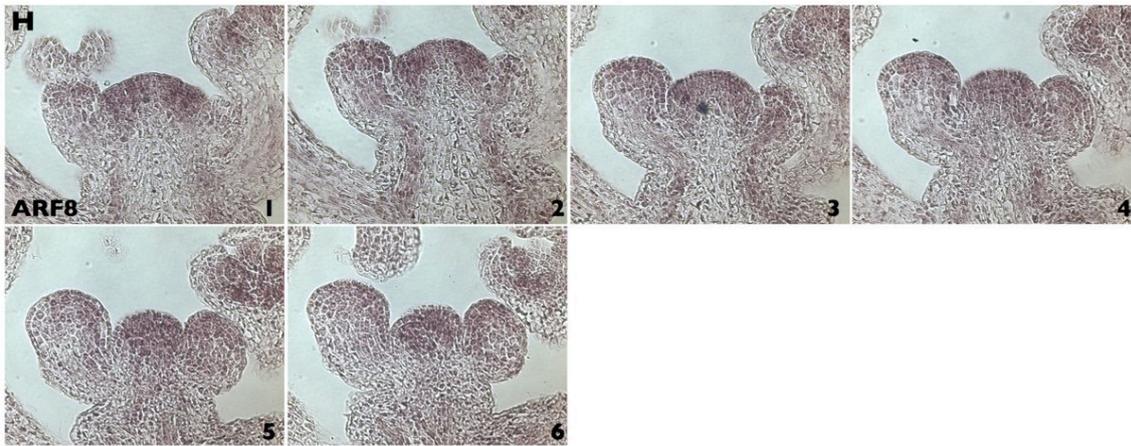


Figure S1.1: TIR1/AFB expression patterns in the inflorescence meristem. Serial sections showing expression of TIR1/AFB proteins fused to GUS (**A-D**) or *AFB4/5* mRNA as visualized by RNA *in situ* hybridization (**E,F**). The name of the gene is indicated on the figure. Scale is identical in all images for each serial section.







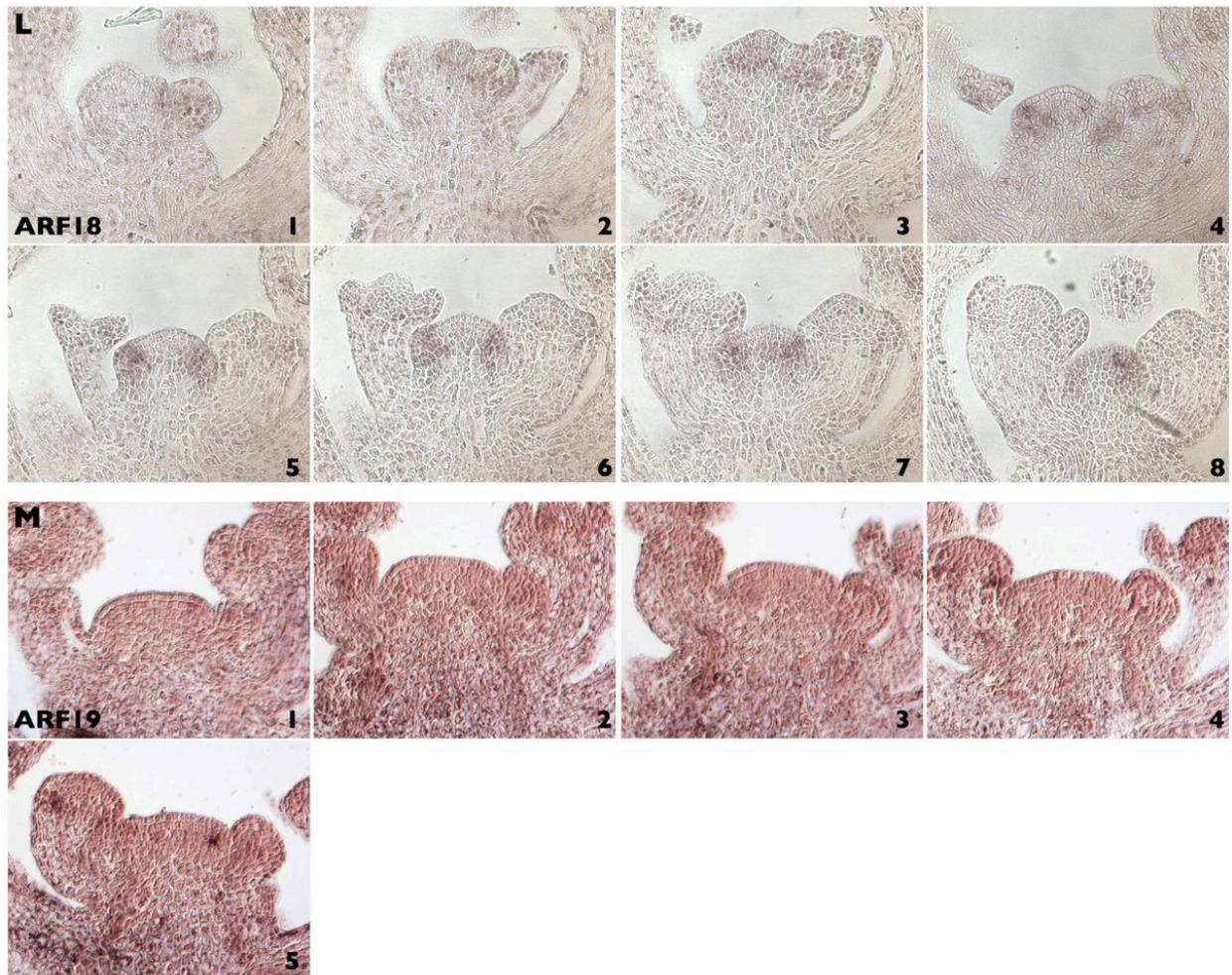
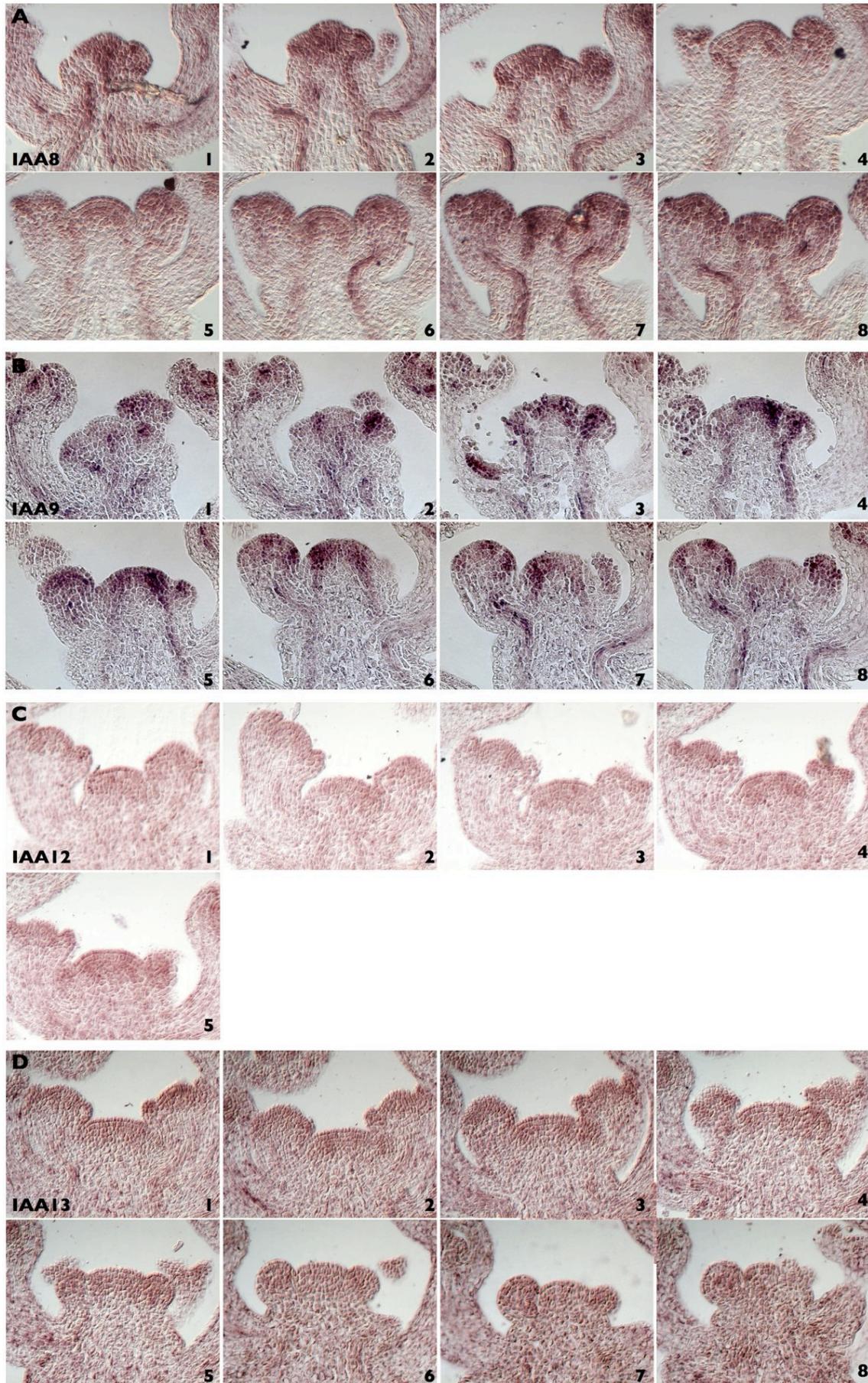
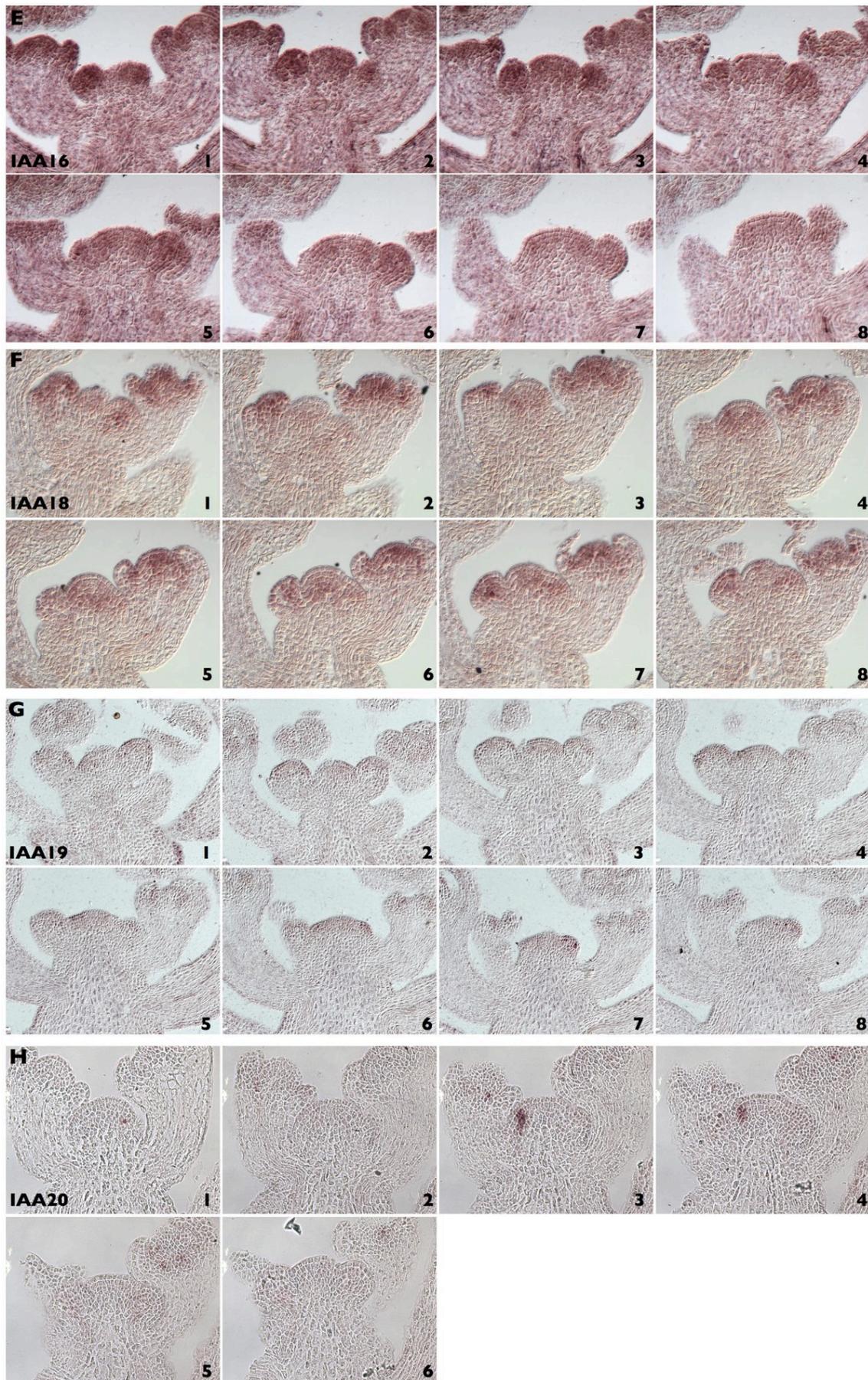


Figure S1.2: ARF expression patterns in the inflorescence meristem. (**A-M**) Serial sections showing expression of *ARF* genes as visualized by RNA *in situ* hybridization. The name of the gene is indicated on the figure. Scale is identical in all images for each serial section.





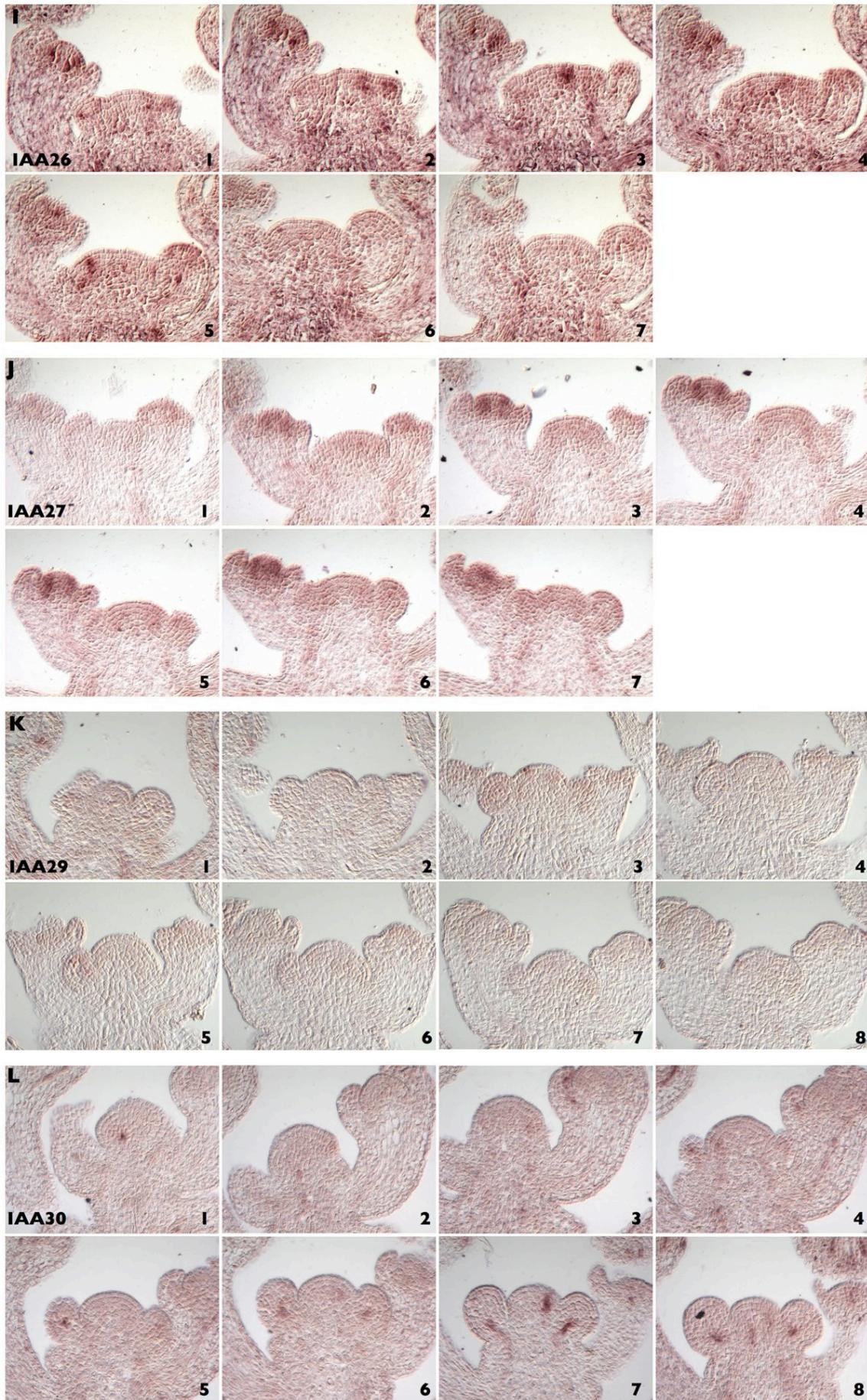


Figure S1.3: Aux/IAA expression patterns in the inflorescence meristem. (A-L) Serial sections showing expression of *Aux/IAA* genes as visualized by RNA *in situ* hybridization. The name of the gene is indicated in the figure. For each serial section, scale is identical in all images.

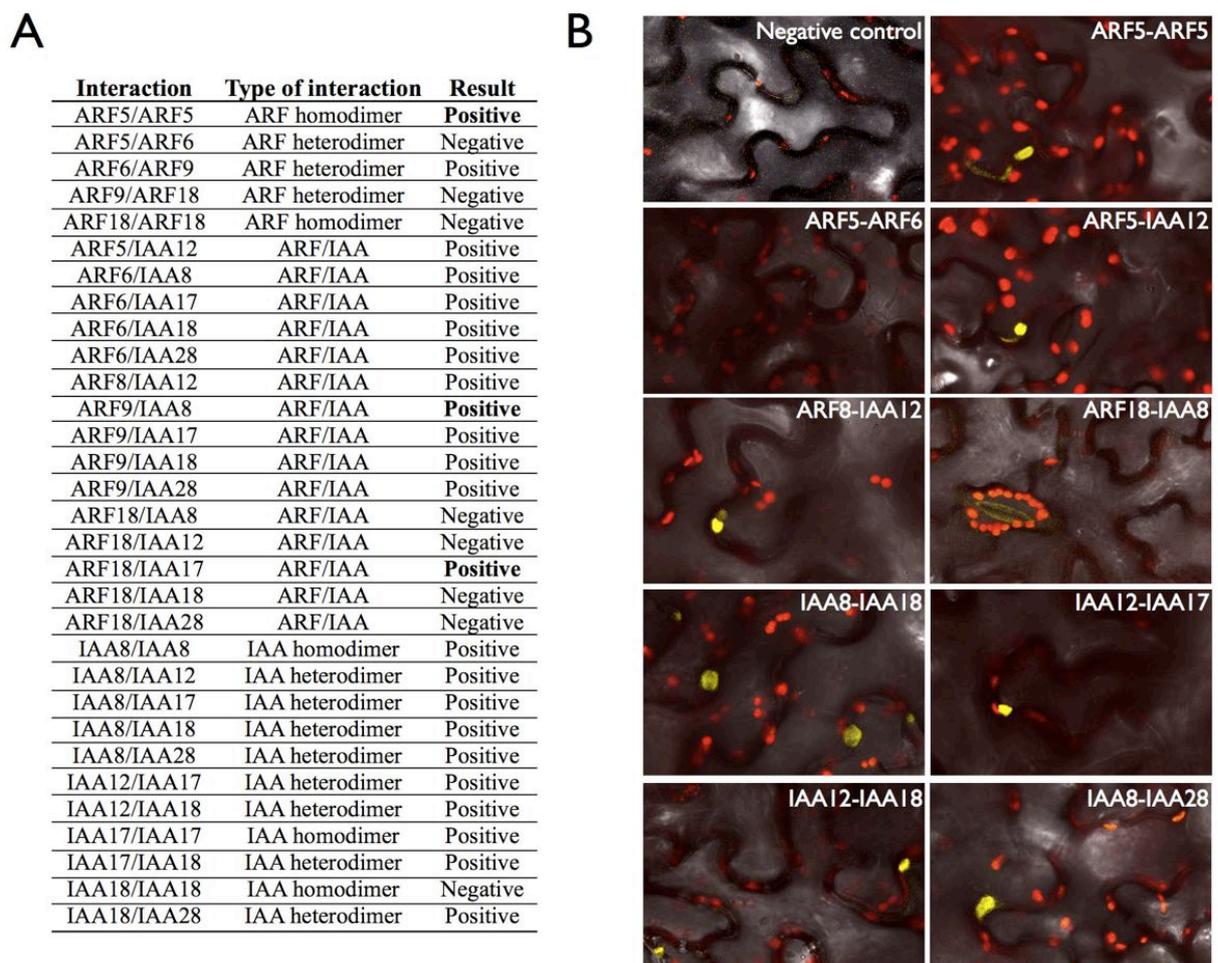


Figure S1.4: *In planta* analysis of Aux/IAA-ARF interactions. Bimolecular Fluorescence complementation (BiFC) was used to test 31 Aux/IAA-ARF interactions *in planta*. The set of interactions included both positive and negative interactions as predicted from the Y2H interactome analysis. (A) Summary of the results of the BiFC analysis. All the results were coherent with the results from the Y2H analysis except for 3 interactions (result indicated in bold). (B) Confocal images for a selection of the tested interactions showing results of BiFC in *Nicotiana Benthamina* leaves. Yellow: YFP; Red: autofluorescence; Grey levels: transmission channel. Scale is identical in all images.

A cluster I (22 proteins): IAA3, IAA8, IAA18, IAA2, IAA4, IAA1, IAA16, IAA10, IAA13, IAA28, IAA15, IAA12, IAA27, IAA19, IAA14, IAA17, IAA20, IAA30, IAA7, *IAA31*, IAA33, IAA32.

cluster II (9 proteins): ARF5, ARF19, ARF8, ARF7, ARF6, IAA5, ARF9, IAA9, IAA34.

cluster III (18 proteins): ARF11, ARF14, ARF3, ARF1, ARF13, IAA6, ARF4, ARF18, ARF16, ARF17, ARF10, ARF2, ARF12, ARF20, ARF22, IAA11, IAA29, IAA26.

$$\hat{\Pi} = \begin{bmatrix} 0.68 & 0.87 & 0.11 \\ 0.87 & 0.34 & 0.21 \\ 0.11 & 0.21 & 0.04 \end{bmatrix} \quad D = \begin{bmatrix} 0.29 & 0.36 & 0.47 \\ 0.36 & 0.25 & 0.50 \\ 0.47 & 0.50 & 0.16 \end{bmatrix}$$

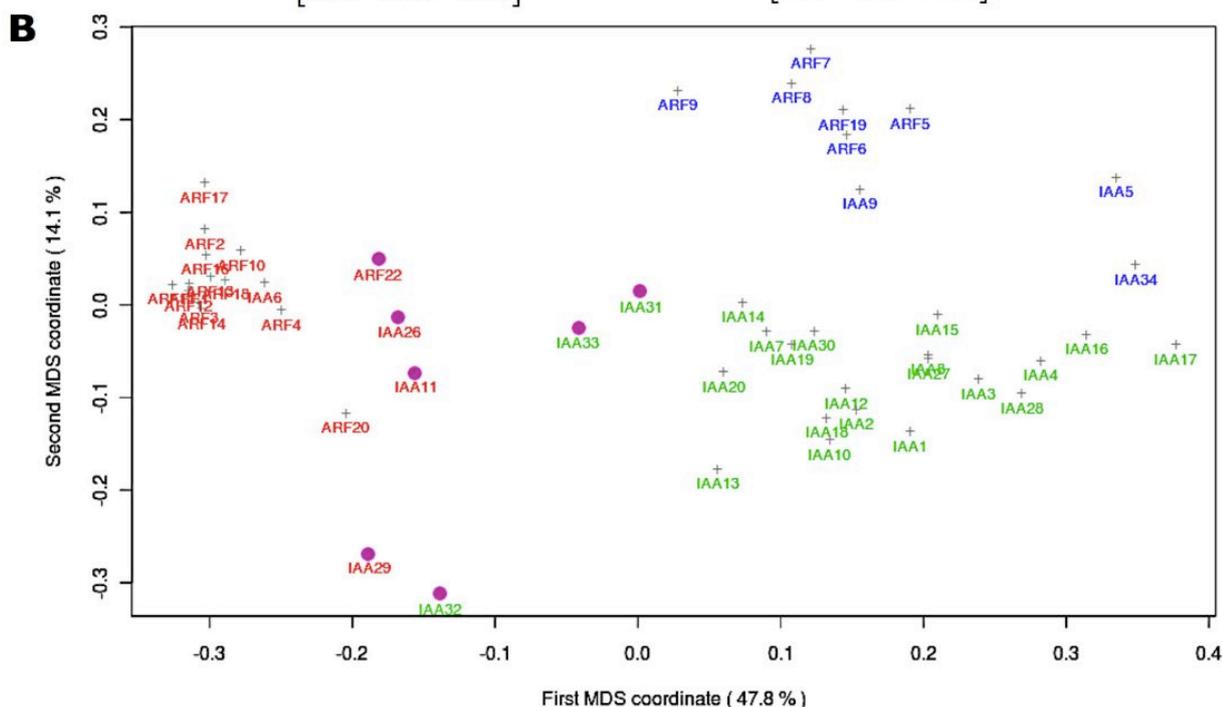


Figure S1.5: Cluster analysis of the global Aux/IAA-ARF interactome using MixNet. **(A)** Results of the MixNet algorithm for the global interactome. We applied the MixNet algorithm to the ARF-Aux/IAA protein network for $Q = 2$ to 10 clusters. The model selection criterion favours 4 clusters. Since this model selection criterion is only asymptotically valid (i.e. for large N), this number of clusters should only be considered as indicative in our case and we thus explored neighbouring solutions. The results are presented for $Q = 3$ clusters. The proteins are ordered from the most to the least central in each cluster based on the distance of the protein to the cluster ($D(i, q)$; data not shown). The connectivity probability matrix $\hat{\Pi}$ and the cluster distance matrix $\{D(q, \ell); q, \ell = 1, \dots, Q\}$ are given. The three clusters differ strongly in terms of connectivity profiles; compare the rows of matrix $\hat{\Pi}$. The three clusters are also well separated; compare the diagonal elements of the cluster distance matrix $\{D(q, \ell)\}$ corresponding to within-cluster distances to the off-diagonal elements corresponding to between-cluster distances. The proteins in italics form the fourth cluster when the MixNet algorithm is applied for $Q = 4$ clusters. In this case, the 3 most peripheral proteins of cluster I are grouped with the 4 most peripheral proteins of cluster II to form a second Aux/IAA cluster. This fourth cluster is not well defined (the within-cluster distance $D(IV, IV) = 0.34$ is larger than the between-cluster distance $D(II, IV) = 0.3$ (data not shown), indicating that the 3-cluster solution is the most adequate. **(B)** Visualization of the clusters using MDS. The first two MDS coordinates were deduced from the pairwise distances between proteins $\{D(i, j); i, j = 1, \dots, N\}$. These first two coordinates account for 61.9% of the total variation. The proteins from cluster I, II and III are figured in green, blue and red respectively. Note that the clusters are compact (except for a few outliers) and well separated. The proteins that are in the 4th cluster when the MixNet algorithm is applied for $Q = 4$ are found amongst the outliers (labelled in purple).

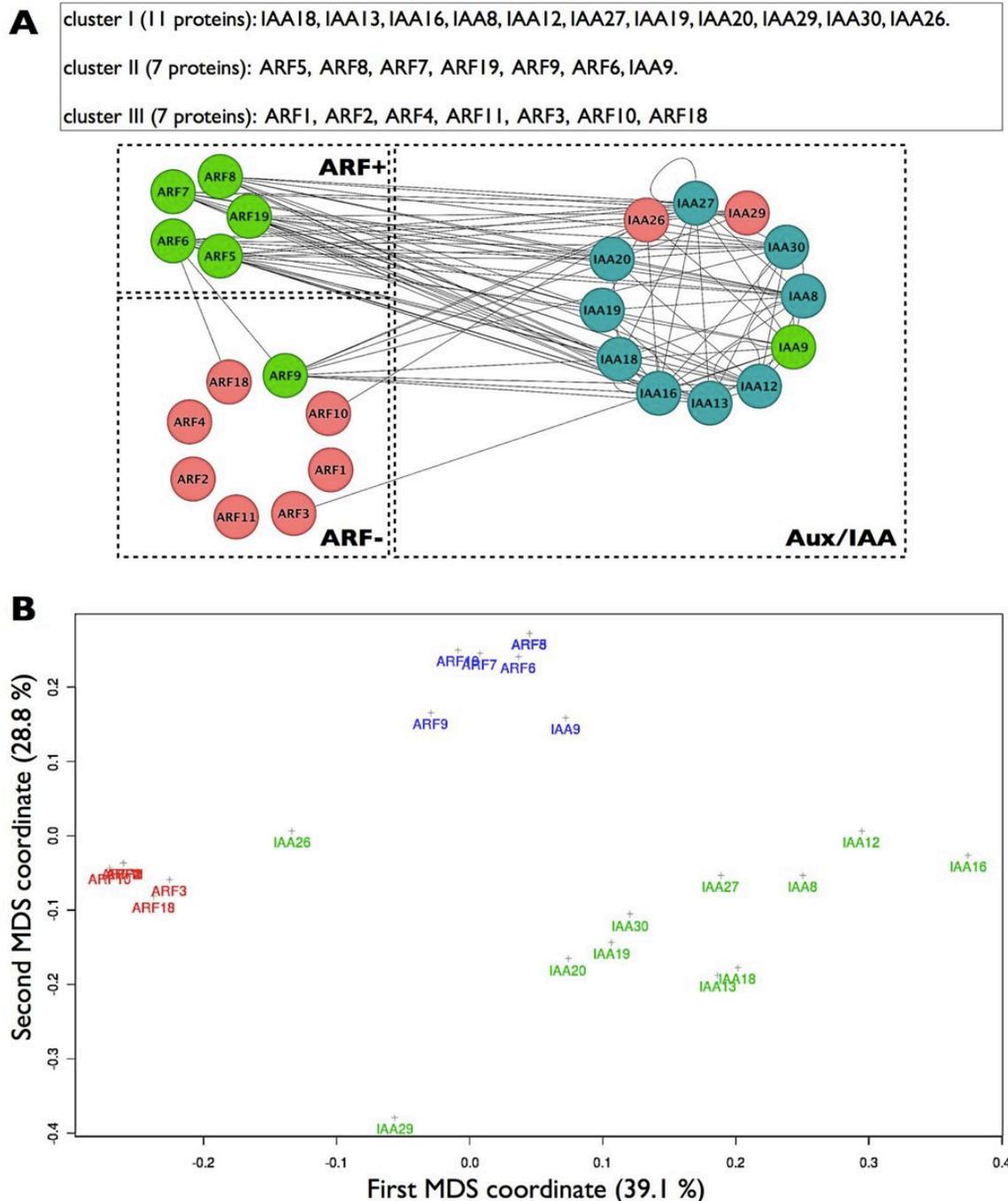


Figure S1.6: Cluster analysis of the meristem Aux/IAA-ARF interactome using MixNet. **(A)** Results of the MixNet analysis for the SAM interactome. We applied the MixNet algorithm to the SAM subnetwork for $Q = 3$ clusters. The three clusters obtained are almost nested in the three clusters obtained for the complete network, the only exceptions being IAA29 and IAA26, peripheral in cluster III, which are then assigned to cluster I with almost all the other Aux/IAs. As in the case of the complete network, the proteins are ordered from the most to the least central in each cluster based on the distance of the protein to the cluster ($D(i, q)$; data not shown) **(B)**. A graphic representation of the structure of the network using Cytoscape (www.cytoscape.org) is given. The Aux/IAA, ARF activator (ARF+) and repressors (ARF-) have been grouped. The vertices are colored according to their MixNet cluster: blue for cluster I, green for cluster II and pink for cluster III. **(C)** Visualization of the clusters using MDS. The first two MDS coordinates were deduced from the pairwise distances between proteins $\{D(i, j); i, j = 1, \dots, N\}$ and explain 67.9% of the total variation. The proteins assigned to cluster I, II and III are figured in green, blue and red respectively. Note again that the clusters are compact (except for a few outliers) and well separated.

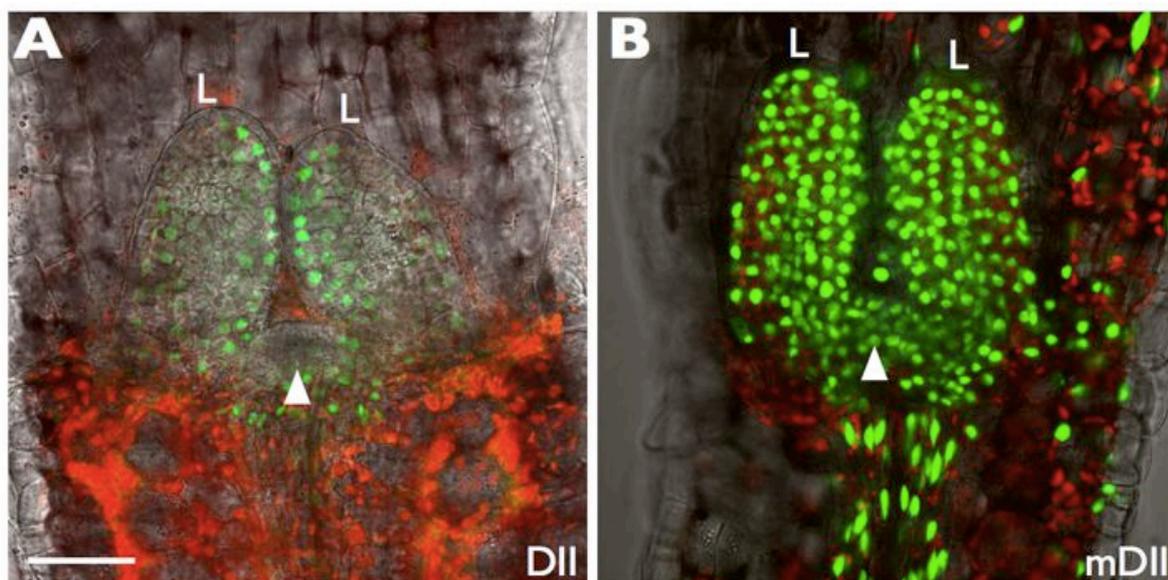


Figure S1.7: Distribution of DII-VENUS and mDII-VENUS in vegetative shoot meristem. DII-VENUS (**A**) and mDII-VENUS (**B**) expression was analyzed in 5 day old plantlets using confocal microscopy. VENUS fluorescence is in green. The red channel visualizes auto-fluorescence. One cotyledon has been peeled out to reveal the meristem (arrowhead). The two first leaves (L) are indicated. Note the absence of DII-VENUS signal at the centre of the meristem and the homogenous distribution of mDII-VENUS. Scale is identical in (A,B). Scale bar: 40 μm .

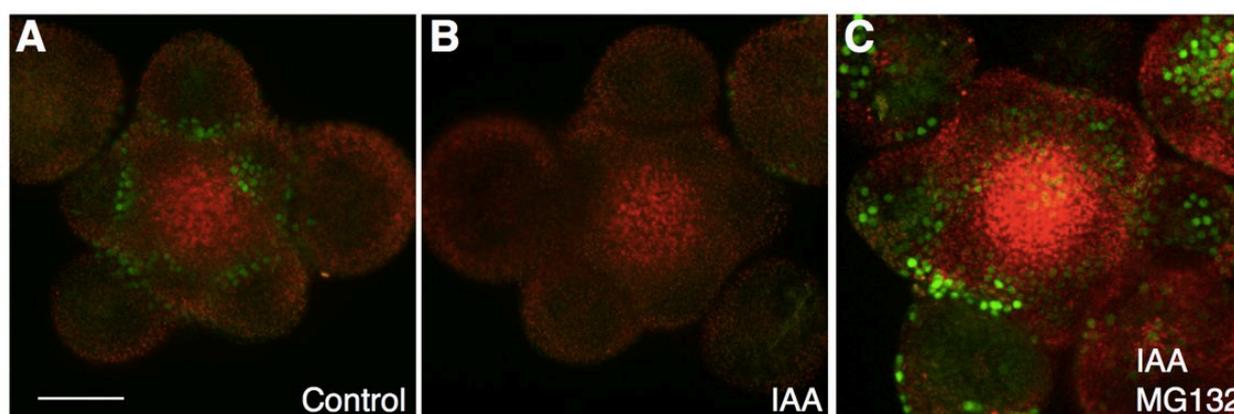


Figure S1.8: Proteasome inhibitors interfere with DII-VENUS auxin-induced degradation in the SAM. Images were obtained by confocal microscopy. DII-VENUS is in green. The red channel visualizes auto-fluorescence in the inflorescence meristem. All the images were taken using the same settings. Plants were treated either with DMSO (control; **A**), 1 μM IAA (**B**) or co-treated with IAA and the MG132 proteasome inhibitor (**C**). Note that the degradation of DII-VENUS induced by auxin is largely blocked by MG132. Scale is the same in (A-C). Scale bar: 40 μm .

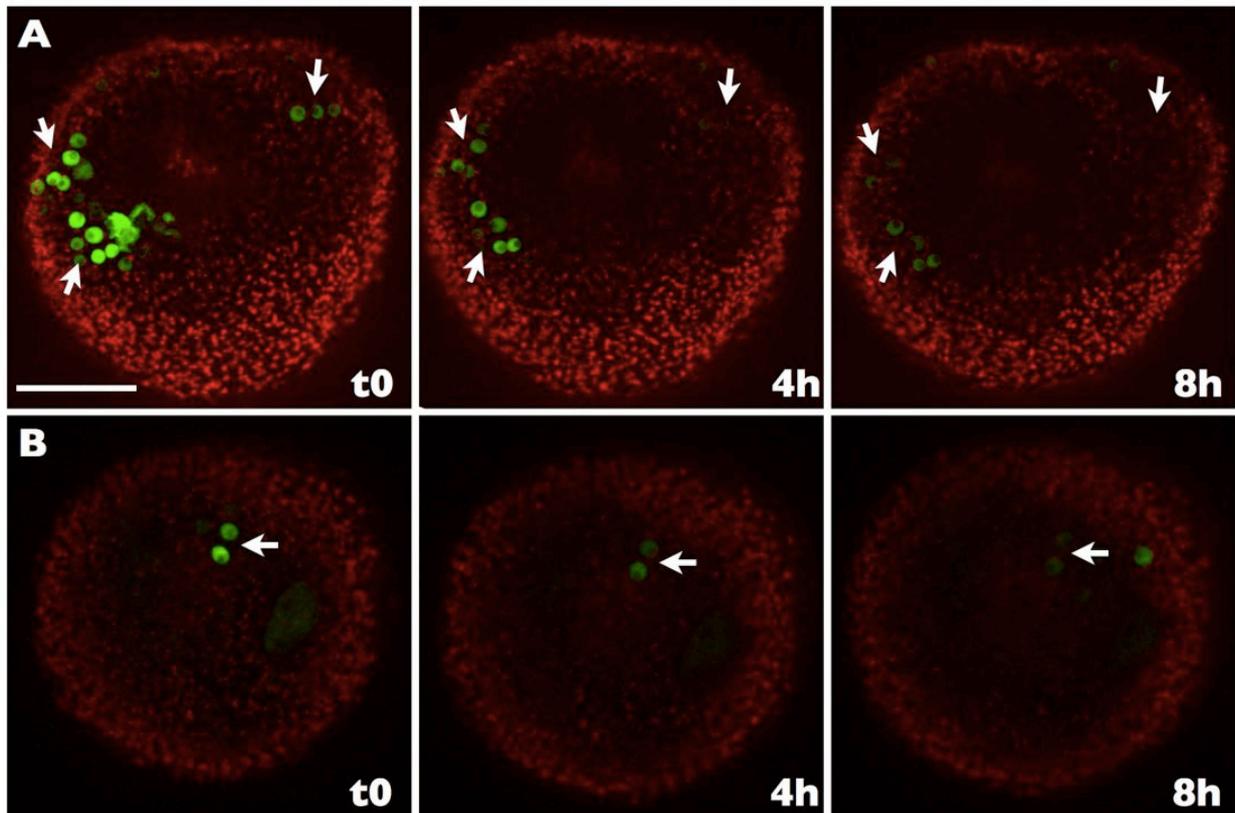


Figure S1.9: Instability of VENUS visualized in DR5::VENUS plants upon chemical inhibition of auxin transport. To verify that the stability of DR5::VENUS fluorescence observed in inflorescence meristem was not solely due to the stability of the VENUS protein, we germinated DR5::VENUS plants on the auxin transport inhibitor NPA to generate pin-like inflorescences in vitro. The plants were then transferred to a new NPA-free medium and DR5::VENUS expression was followed over time as described in Hamant *et al.* [2008]. Due to the perturbation of auxin transport, we often observed in these conditions transient peaks of DR5::VENUS expression that were not followed by organ induction. Representative results from 2 independent time-courses are shown (A,B). Arrows indicate groups of nuclei showing rapid decrease in fluorescence. Note that in each case drastic diminution in VENUS fluorescence could be observed after 4 hours, suggesting a half-life for the protein in that time range. Scale is identical in all images. Scale bar: 20 μm .

Chapter 2

Modelling the influence of phylogenetic distances between dimerisation sequences on the auxin signalling interactome

Contents

Abstract	94
1 Introduction	95
2 Result and Discussion	96
2.1 Y2H experiment and data	97
2.2 Building a binary network from Y2H data	98
2.3 Building a valued network from Y2H data	99
2.4 Network analysis using Bernoulli and Gaussian mixture models . .	100
2.5 Influence of the phylogenetic distance between dimerisation sequences on the interaction network	105
3 Conclusions	111
4 Methods	113
4.1 Construction of the Aux/IAA-ARF interaction network on the basis of the yeast-2-hybrid data.	113
4.2 Network binarisation.	114
4.3 Phylogenetic data	115
4.4 Response distance matrix: standardized distances between transcriptional factors	116
4.5 Phylogenetic distance matrix	117
4.6 Assessing the adequacy of the clustering	117
4.7 Availability of supporting data	118
Author contributions	119
Supplementary Materials	120

Abstract

Auxin is a major phytohormone involved in many developmental processes by controlling gene expression through a network of transcriptional factors. In *Arabidopsis thaliana*, the auxin signalling network is made of 52 potentially interacting transcription factors, activating or repressing gene expression. Since the network was not characterised, we first sought at experimentally testing for all possible interactions. To get a clear picture of the mechanisms at work, we then applied model-based graph clustering methods relying on connectivity profiles between transcription factors. Incorporating phylogenetic distances as explanatory variables, we modelled the influence of the protein dimerisation sequences on the auxin interactome structure using mixture of linear models for random graphs. Our results provide evidence that the network can be simplified into three groups, closely related to biological groups. Ultimately we found that these groups behave differently, depending on their dimerisation sequence similarities, and that the two dimerisation domains might play different roles. We propose here the first pipeline of statistical methods combining Yeast-2-Hybrid data and phylogenetic distances for analysing protein-protein interactions.

1 Introduction

Auxin is a key signal in plant development that regulates organogenesis from embryogenesis onward. This major phytohormone achieves this morphogenetic activity notably by regulating many downstream genes through transcription factors (TFs). In *Arabidopsis thaliana* the control of gene expression in response to auxin involves a complex network of 52 TFs, consisting in 29 AUXIN/INDOLE-3-ACETIC ACID (Aux/IAA) and 23 AUXIN RESPONSE FACTOR (ARF) acting either as transcriptional activators or repressors (for review, see Leyser [2006]; Guilfoyle and Hagen [2007]). The present molecular model of the auxin signaling pathway implicates the forma-

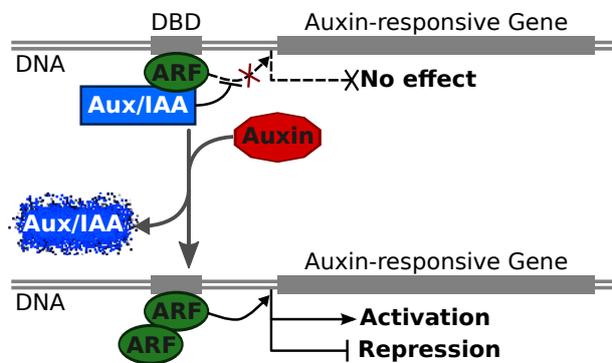


Figure 2.1: Model of auxin transduction pathway. Activation and repression activities depend on ARF middle domain amino-acid composition. DBD indicate the DNA binding domain found upstream of auxin-inducible genes. In absence of auxin Aux/IAA are dimerise with the ARFs, preventing them to exert their activating or repressing activity. When auxin is present, it target the Aux/IAA to the proteasome leaving the ARF free to dimerise and exert their regulating activity. Source: adapted from Chapman and Estelle [2009].

tion of hetero-dimers between ARFs and Aux/IAs in absence of auxin (see figure 2.1). According to Remington *et al.* [2004] these transcriptional regulators interact through a C-terminal dimerisation domain (CTD), made of two conserved sub-sequences known as domains III and IV (figure 2.2). ARFs can bind to DNA thanks to a DNA binding domain (DBD), and depending on the amino-acid composition of their intermediate domain they act as activators (ARF+) or repressors (ARF-) of the auxin-responsive genes transcriptional activity (figure 2.2). It should be noted that Aux/IAs do not have a DBD domain and therefore are thought to be unable to regulate alone the transcription of auxin-responsive genes. When auxin accumulates in cells as a result of the activity of AUX1/LAX [Reinhardt *et al.*, 2003] influx carriers and PIN-FORMED1 (PIN1) [Gälweiler *et al.*, 1998; Vernoux *et al.*, 2000; Reinhardt *et al.*, 2003] efflux carrier or of changes in biosynthesis, auxin perception targets Aux/IAs to the proteasome [Leyser, 2006] leading to their degradation. This subsequently frees the ARFs from the Aux/IAA and leads to the regulation of the gene acting downstream of auxin. It is only recently that global information on the topology of the Aux/IAA-ARF network were obtained. A two-way Yeast-2-Hybrid (Y2H) [Joung *et al.*, 2000] experiment using Aux/IAA and ARF proteins has allowed to test for all possible interactions (with the exception of

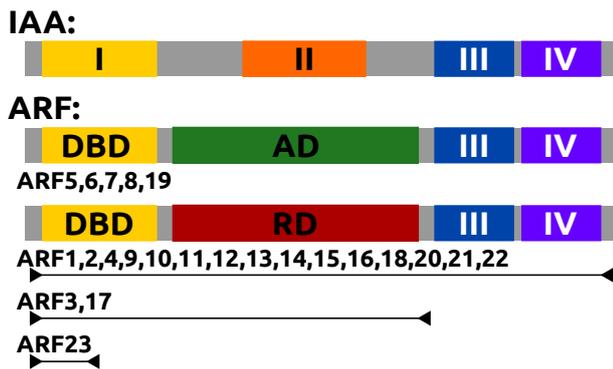


Figure 2.2: Schematic representation of the ARF and Aux/IAA structures as found in *Arabidopsis thaliana*. DBD: DNA binding domain. I: Aux/IAA specific putative homo-dimerisation domain. AD: Activation domain. RD: Repression domain. II: Aux/IAA specific degradation domain. III & IV: protein dimerisation domains. Arrowed lines indicate the extent of each inhibiting ARF structure. Source: adapted from Hagen and Guilfoyle [2002]; Guilfoyle and Hagen [2007] with the notable difference that we found domains III and IV when aligning full length protein sequences for ARF13.

ARF15, 21 and 23, see Vernoux *et al.* [2011] and Methods). To explore the organization of this network, a model-based graph clustering method [Daudin *et al.*, 2007] that groups proteins on the basis of their connectivity profile (i.e. similar interactors) was applied. Three clusters of proteins, that were very close to known biological groups (ARF+, ARF- and Aux/IAs), were found using Bernoulli mixture models for random graphs [Vernoux *et al.*, 2011], thus demonstrating rather stereotypical interaction properties within these three groups.

Here we extended this approach in order to analyse the contribution of the primary sequences of the conserved DIII-DIV dimerisation domains to the interaction capacities between or within known biological groups. To this end, we used a recently proposed generalisation of the mixture models for random graphs that offers the possibility to deal with valued graphs and to include explanatory variables [Mariadassou *et al.*, 2010]. This integrative statistical model constitute the core of a pipeline of methods to analyse how phylogenetic distances between dimerisation domains influence protein-protein interactions.

2 Result and Discussion

	Y2H data	
Pre-processing	Binarisation	Standardisation
Network	Binary	Valued
Structure analysis	BM model	GM, LM models
Informations	Preferential connexion structure	Preferential connexion structure & biological effects

Figure 2.3: Methodological options to analyse Yeast-2-Hybrid and other interaction data.

A binary network is often easier to interpret than a valued one. However, in our case, it does not fully represent the actual biochemical behaviour of the signalling pathway since the “true” biochemical model would depend several properties, such as dimerisation strength, proteins concentration, spatial expression and synthesis/degradation dynamics of the TFs. We will first recall how the binary network was build and analysed in Vernoux *et al.* [2011]. Then we will compare this previous approach with the determination and analysis of a valued network, before investigating the incorporation of phylogenetic distances in such a modelling framework.

2.1 Y2H experiment and data

Our raw experimental data are the Y2H outputs obtained when testing the auxin perception network, made of 52 members belonging to the Aux/IAA and ARF family. The Y2H technique is a bio-engineered tool based on the GAL-4 transcription factor from yeast *Saccharomyces Cerevisiae* [Chien *et al.*, 1991]. It is a powerful tool to test for potential interaction between two proteins, and is now widely used, even though it is an heterologous system in most cases.

In order to be thorough, each interaction have been tested both ways, since the tested proteins were fused to both activation domain (AD) and binding domain (BD) portions of the split Gal-4. In addition, two separate repetitions and two reporter genes have been used (HIS3 and X-Gal). Lastly, we were able to test the interaction of 49 transcription factors, except for three ARFs (15, 21 and 23), thus making a total of 2401 combinations. We give in table 2.1 an example of the results.

Table 2.1: Example of Yeast-2-Hybrid data, with the name of the tested proteins, the side they were attached to and the each output provided by the reporter genes.

Bait(BD)	Prey(AD)	X-Gal	HIS3	Bait(BD)	Prey(AD)	X-Gal	HIS3
BD-ARF1	AD-ARF1	–	12%	BD-ARF1	AD-ARF2	–	14%
BD-ARF2	AD-ARF1	–	14%	BD-ARF1	AD-ARF3	–	13%
BD-ARF3	AD-ARF1	–	15%	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	BD-ARF5	AD-IAA2	+?	119%
BD-IAA2	AD-ARF5	++	90%	BD-ARF5	AD-IAA3	++	121%
BD-IAA3	AD-ARF5	++	90%	BD-ARF5	AD-IAA4	+++	70%
BD-IAA4	AD-ARF5	++	121%	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮				

A preliminary analysis of these 4802 results showed a high correlation between the two tests X-Gal and HIS3 (Spearman rank correlation coefficient $\rho_{sp} = 0.7$). This led us to hypothesize that the two tests provide complementary information regarding protein-protein interactions.

2.2 Building a binary network from Y2H data

2.2.1 Determination of interaction threshold per biological test

Creating a binary network imply to choose thresholds for both tests in order to define them as positive or negative, and to define decision rules to combine the four test results to determine the presence or absence of the interaction. Looking for a structure in the data that could hep us define a threshold for the X-Gal test, we first examined the empirical distributions of the optical density (OD) ratios (HIS3 test) for each mark of the X-Gal test. A gradual shifting of the mean OD ratio for successive marks ‘+?’ , ‘+’ , ‘++’ and ‘+++’ can be observed in supplementary figure S2.1. Moreover, the OD ratios distribution for the ‘+’ mark presents a remarkable bimodal structure, the component of highest mean being rather similar to the OD ratios distribution for ‘++’ (see supplementary figure S2.1). The component of lowest mean should be considered as false positive for the X-Gal test. We thus chose to summarize the output of the X-Gal test as a threshold between the successive marks ‘+?’ and ‘+’.

In order to define a threshold value for the HIS3 test, we then represented the overall OD ratios distribution as cumulative frequencies separating X-Gal marks $< +$ (in red) and X-Gal marks $\geq +$ (in green) as seen in figure 2.4. This histogram shows a hump around 0.15 corresponding to negative interactions followed by a long and flat tail, leading to the idea of distinct populations (at least two) being mixed in the representation. One population clearly defining the absence of interaction (low mean OD ratio) and a less obvious one defining the presence of interaction (high mean OD ratio) in the tail. To separate them, we applied a Gaussian mixture model using the `mclust` R-package on the basis of the overall OD ratio sample. Three components were selected by the Bayesian information criterion (BIC). The two first Gaussian components correspond to almost only X-Gal marks inferior to ‘+’ while the last one corresponds mostly to marks superior or equal to ‘+’. We therefore fixed a threshold at 0.45 for the HIS3 test, which correspond roughly to the limit where the posterior probability of the third component exceeds the posterior probability of the second one and where the proportion of X-Gal marks superior or equal to ‘+’ exceeds the proportion of marks inferior to ‘+’. We also generated the binary graphs with an HIS3 threshold at 0.3 and 0.65. However, they only slightly modify the binary networks (see supplementary figures S2.2 and S2.3).

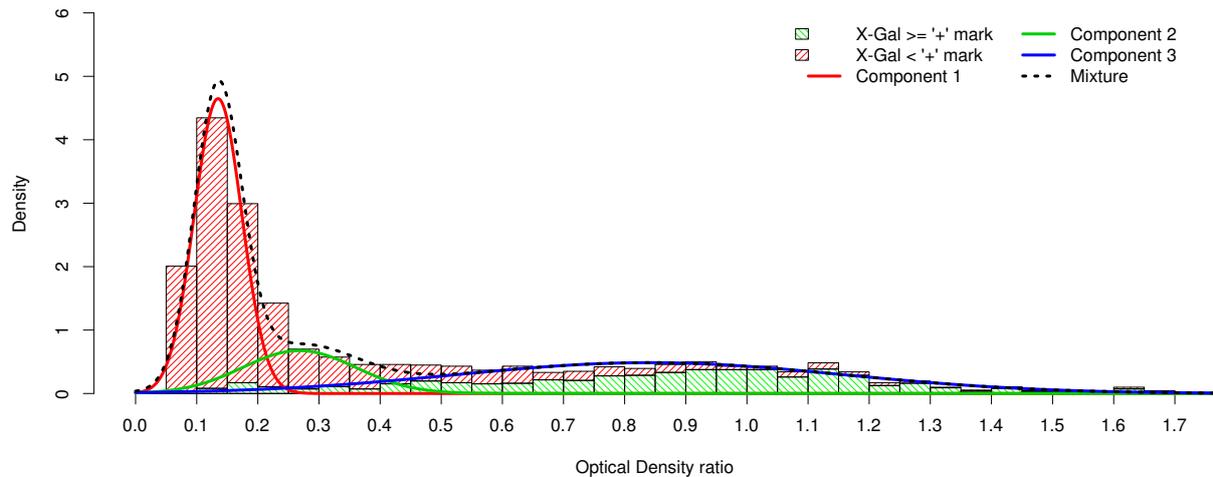


Figure 2.4: Three-component Gaussian mixture model estimated for the the OD ratios. The optical density has here an upper limit of 1.7 for graph readability reasons.

2.2.2 Defining the decision rules

Finally, having two possible configurations for each protein (AD and BD or bait and prey) for the two biological tests (X-Gal and HIS3) we have to define decision rules to combine the four biological tests outputs, to ultimately define the presence or absence of interaction (see Methods 4.2.2). Building five configurations as described in Methods 4.2.2, and relying on four of them, we then obtained a binary network as presented on the form of an adjacency matrix in supplementary figure S2.4.

2.3 Building a valued network from Y2H data

Combining the Y2H experiment outputs in a single interaction distance requires a standardization procedure (see Methods 4.4). The objective of standardization is to suppress the dependency upon elementary distance type and scale. It is important to point out that, in our case, the valued network does not represent affinity between proteins, but reflect the likelihood of interactions. In order to combine the four test outputs into one standardised interaction distance, we tested several weightings of the returned values for the two reporter genes X-Gal and HIS3. To explore the effects of the weighting, we used the three following weighted networks:

- network **A**: $w_{\text{X-Gal}} = 0.75$ and $w_{\text{HIS3}} = 0.25$
- network **B**: $w_{\text{X-Gal}} = 0.5$ and $w_{\text{HIS3}} = 0.5$
- network **C**: $w_{\text{X-Gal}} = 0.25$ and $w_{\text{HIS3}} = 0.75$

To this end, we simply visualized the standardised distance distributions corresponding to no-interaction (red) and interaction (green) according to the previously defined binary assignment (figure 2.5). We could observe that network C tends to spread ‘no-interactions’ on a wide range of standardised interaction distances, thus presenting a rather large overlap with ‘interactions’. Network A tends to concentrate ‘no-interactions’ around a small range of standardised interaction distances, and to separate them from ‘interactions’. Finally, corresponding to the balanced weighting, network B presents acceptable properties in terms of dispersion and overlap between ‘no-interactions’ and ‘interactions’. We will thus present the clustering results for networks A and B. Lastly, we here want to highlight that comparing the distributions for the three weighted networks indicates that, in our case, the X-Gal test seems more reliable than the HIS3 test, probably because of the very long tail corresponding to ‘interaction’ for this test (see figure 2.4).

2.4 Network analysis using Bernoulli and Gaussian mixture models

To gain insight into the network structure, we sought at grouping the TFs on the basis of their connectivity profiles [Vernoux *et al.*, 2011]. We applied the model-based graph clustering method introduced in Mariadassou *et al.* [2010]. The key feature of the mixture model for random graph is to give a probabilistic summary of the connectivity structure by uncovering clusters of TFs that share the same connectivity profiles. The parameters of the model are the cluster weight distribution and the connectivity distributions for each pair of clusters. In the case of a binary adjacency matrix, the connectivity distributions are Bernoulli distributions parametrized by connectivity probabilities, that is the probability for proteins of two clusters to interact.

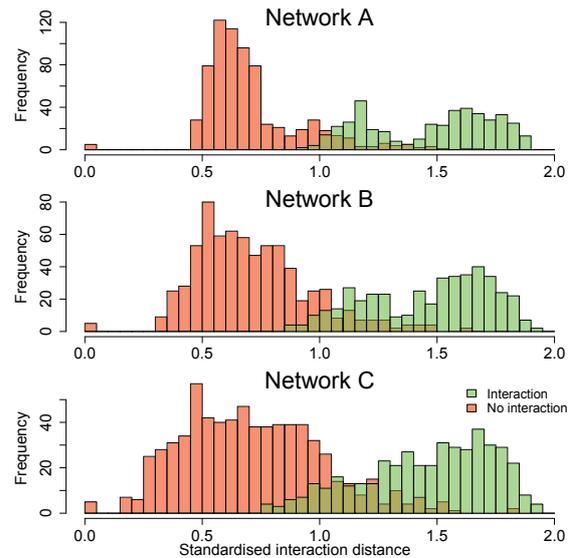


Figure 2.5: Empirical distribution of the standardised interaction distances for the A, B and C (top-down) networks.

$$Z_{ij}|\{i \in C_q, j \in C_\ell\} \sim \mathcal{B}(\pi_{q\ell}), \quad (2.1)$$

where Z is the binary adjacency matrix, Z_{ij} is a connexion between vertices i and j , belonging to clusters q and ℓ respectively. This can be read as follow: the probability of a connexion between vertices i and j knowing that i belong to cluster q and j to ℓ follows a Bernoulli distribution of parameter $\pi_{q,\ell}$.

In the case of a weighted adjacency matrix, the connectivity distributions are Gaussian distributions.

$$X_{ij}|\{i \in C_q, j \in C_\ell\} \sim \mathcal{N}(\mu_{q\ell}, \sigma^2), \quad (2.2)$$

where X is the weighted adjacency matrix.

The inference of such models is not restricted to the estimation of the cluster weight and connectivity distributions but encompasses the inference of the number of clusters using a penalized likelihood criterion. The principle of penalized likelihood criteria such as the ICL criterion consists in making a trade-off between an adequate fitting of the model to the data and a reasonable number of parameters to be estimated. Compared to classical model selection criteria such as BIC, the ICL criterion takes account of the clustering objective and is expected to favour model such that the uncertainty of protein assignment to clusters is low. Jeffreys' rules of thumb [Kass and Raftery, 1995] suggest that a difference of ICL of at least $\log(100) = 4.6$ is needed to deem the model with the higher ICL substantially better.

2.4.1 Binary or valued graph and associated models

The presented clustering results using Bernoulli mixture (BM) model slightly differ from those presented in Vernoux *et al.* [2011]. We here indeed excluded ARFs 3 and 17, because they do not possess the conserved domains III and IV. Therefore we cannot compute phylogenetic distances between the dimerisation sequences in order to use them as covariables in the models presented in section 2.5.

2.4.2 Comparing model selection criterion and clusters composition returned by the BM and GM models

BM models ICL criterion, posterior probability and clusters composition. When estimating BM models on the basis of the 46 protein binary network, the ICL criterion favours the 6-clusters BM model (see Table 2.2) while with 48 proteins, the ICL criterion favours the 4-cluster BM model [Vernoux *et al.*, 2011]. However, in both cases the 4- and 6-clusters BM models were

close in terms of ICL criterion. Moreover for the 46 protein binary network, the ICL difference between the 4- and the 6-clusters BM models was not significant according to Jeffreys' rules of thumb: $\Delta\text{ICL} \leq 2$. Since the ICL criterion is only asymptotically valid (i.e. for large N), the number of clusters given by this criterion should be considered as indicative. After exploring the clusterings outputs for different cluster numbers, we have chose to present the 4-clusters BM model.

Number of Clusters	2	3	4	5	6	7
ICL	-527.3548	-521.8064	-506.7471	-511.0779	-504.9562	-507.8915
Posterior probability	0	0	0.136	0.002	0.818	0.043

Table 2.2: ICL and corresponding posterior model probabilities for the BM model.

Cluster 1	ARF5 (0.19), ARF19 (0.212), ARF8 , ARF7 , ARF6 (0.258), IAA5 (0.299), ARF9, IAA9, IAA34
Cluster 2	ARF14 (0.087), ARF1 (0.096), ARF13, ARF16 (0.115), IAA6, ARF4, ARF10, ARF18, ARF2 (0.137), ARF12 (0.154), ARF20 (0.189)
Cluster 3	IAA3 (0.198), IAA8 (0.205), IAA4 (0.222), IAA2, IAA18, IAA1, IAA16, IAA28 (0.25), IAA15 (0.261), IAA10, IAA12, IAA13, IAA27, IAA19 (0.284), IAA14 (0.296), IAA17, IAA20 (0.307), IAA30 (0.33), IAA7
Cluster 4	IAA11 (0.333), ARF22 (0.337), IAA26, IAA29 (0.348), IAA33 (0.377), IAA32, IAA31 (0.435)

Table 2.3: Composition of the four clusters obtained using the BM model. The ARF activators are in bold. The distance $D(i, q)$ between protein i and cluster q to which its is assigned is given for the most central, the most peripheral and some other proteins of interest for interpretation.

Exploring the composition and the properties of the four clusters (table 2.3), we found three clusters corresponding to biologically meaningful groups (ARF+, ARF- and Aux/IAA respectively) and an “outliers” cluster (see the protein-cluster distances for this cluster compared to the three others in table 2.3) grouping one ARF- and six Aux/IAs. As we will see later, this view of a clustering with three biologically meaningful clusters and an “outliers” cluster is shared by the different models and will therefore be used for clustering comparison returns by the different models.

Number of Clusters	1	2	3	4	5	6	7
ICL	-595.221	-333.666	-283.778	-268.434	-258.972	-260.468	-268.91
Posterior probability	0	0	0	0	0.817	0.183	0

Table 2.4: ICL and corresponding posterior model probabilities for network A based on a GM model.

Number of Clusters	1	2	3	4	5	6	7
ICL	-617.343	-344.357	-306.136	-286.626	-279.985	-265.725	-278.627
Posterior probability	0	0	0	0	0	1	0

Table 2.5: ICL and corresponding posterior model probabilities for network B based on a GM model.

GM models ICL criterion, posterior probability and clusters composition. The ICL criterion values for the GM models are summarized in tables 2.4 and 2.5 for networks A and B respectively. Comparing these tables (2.4 and 2.5) highlights that the X-Gal test is more reliable than the HIS3 test, as foreseen when analysing figure 2.5. By giving more weight to the X-Gal test in network A, than the balanced situation of B, the best GM model estimated for network A (table 2.4) has one less cluster than the best GM model estimated on the basis of network B (table 2.5). We suspect that this may be due to the high dispersion of OD ratio corresponding to protein interaction (see figure 2.4).

Cluster 1	ARF5 (0.01), ARF7 , ARF8 , IAA9, ARF19 , ARF6 (0.019)
Cluster 2	ARF1 (0.012), ARF10, IAA6, IAA11, ARF4 (0.013), ARF14, ARF16, ARF18, IAA29 , ARF20 (0.014), ARF12 (0.015), ARF13, ARF2 (0.016)
Cluster 3	IAA15 (0.028), IAA10, IAA31, IAA2, IAA14, IAA1 (0.031), IAA12, IAA18, IAA4 (0.033), IAA17, IAA27, IAA19 (0.034), IAA3, IAA8, IAA16, IAA28, IAA34, IAA5 (0.036)
Cluster 4	IAA33 (0.019), ARF22 , IAA13 , IAA7 , IAA26 , IAA30 , ARF9 (0.024), IAA20 , IAA32 (0.024)

Table 2.6: Composition of the four clusters obtained using the GM model applied to network A. The ARF activator are in bold. The proteins that are attributed to two clusters in the 5-clusters GM model are in blue and cyan. The distance $D(i, q)$ between protein i and its cluster q is given for the most central, the most peripheral and some other proteins of interest for interpretation purposes. See supplementary figures S2.5 and S2.6 for the distance plot.

One should note a specificity of cluster 3 (IAA) in table 2.6 whose lowest protein to cluster distance (0.028) is greater than the highest protein to cluster distances (0.019, 0.016, 0.024) found in the three other clusters. This very different scale of distances for cluster 3 can be explained by the greater dispersion of the HIS3 test values (OD ratio) for ‘interaction’ with respect to ‘no-interaction’ (see figure 2.4) and is reflected by the higher dispersion of the standardised interaction values for ‘interaction’ with respect to ‘no-interaction’ (see figure 2.5-Network A).

BM and GM models comparison on the basis of the clusters composition. The comparison of cluster composition between networks A and B gives 96% of match (44/46) for four, five and six clusters (see table 2.11). The compositions of the four clusters obtained using the BM model (see table 2.3) and the GM model (see table 2.6) are rather similar: 78% of match between BM and GM-A (10 differences in cluster assignment). It should be noted that the differences in cluster assignment concern almost only peripheral elements in clusters and that the core of the four clusters are very similar.

BM and GM models comparison on the basis of the between-clusters distance matrix. An important criterion to ascertain the validity of a clustering model is the between-clusters distance matrix $D(q, \ell)$. Considering the 4-clusters BM model, the ARF+, ARF- and Aux/IAA clusters present smaller within-cluster distances (diagonal) than between-clusters distances (off diagonal), showing a strong definition of these clusters ($D_{BM}(q, \ell)$ given below, model parameters given in S2.2 and clustered network representation given in S2.7). The “outliers” cluster is mainly composed of Aux/IAs and its within-cluster distance is higher than the within-cluster distance of the ARF+, ARF- and Aux/IAA clusters. In addition, its within-cluster distance is larger than its distance to ARF- group. This configuration can be interpreted in the framework of density-based clustering (Kriegel *et al.* [2011] and references therein) where the ARF+, ARF- and Aux/IAA clusters are characterized by rather high density of elements with respect to the density of elements of the “outliers” cluster. These outliers can be explained for a part by biological noise in the Y2H experiments.

$$D_{BM}(q, \ell) = \begin{matrix} & \begin{matrix} \text{ARF+} & \text{ARF-} & \text{IAA} & \text{Outliers} \end{matrix} \\ \begin{pmatrix} 0.257 & 0.533 & 0.364 & 0.512 \\ 0.533 & 0.124 & 0.524 & 0.314 \\ 0.364 & 0.524 & 0.260 & 0.435 \\ 0.512 & 0.314 & 0.435 & 0.354 \end{pmatrix} \end{matrix}.$$

For the GM model, the models selected by the ICL criterion comprise five clusters. Considering the four clusters given by the GM models, we note that they are poorly defined, since the within-cluster distance for the Aux/IAA cluster is greater than some between-clusters distances (see supplementary table S2.1).

2.4.3 Conclusions on the presented models

The binarisation of the interactions is likely to remove experimental noise but may also introduce errors (false positives or negatives) depending on the used threshold. On the opposite, the standardization is a more objective approach, since it scales the results of the X-Gal and HIS3 tests to make them comparable and limits the loss of information. Although the standardization does not remove experimental noise which here seems to be a shortcoming.

Lastly, one key output of the random graph mixture model to validate the clustering is the

posterior probabilities of protein assignment to clusters. For each protein, these posterior distributions are degenerate (probability of 1 for a given cluster and 0 for the others). This situation eases the interpretation of the model parameters.

2.5 Influence of the phylogenetic distance between dimerisation sequences on the interaction network

2.5.1 Building the phylogenetic distance matrix

To build a distance matrix corresponding to the differences in terms of amino-acid sequences, we started by aligning the full proteic sequence of every TF presenting a C-terminal dimerisation (CTD) domain using CLUSTAL-W [Thompson *et al.*, 1994]. ARFs 3 and 17 having no referenced III and IV domains have been excluded from the analysis. Since it was not possible to test interactions in Y2H experiment for ARFs 15, 21 and 23, and since no interaction was observed for ARF11, these four TFs were also excluded from the analysis. We thus finally worked on a network of 46 potentially interacting TFs. To recover domains III and IV amino-acid sub-sequences we searched for conserved patterns among the aligned sequences using Gblocks [Castresana, 2000]. Two conserved blocks, corresponding to domains III and IV were found at the C-terminal part of the sequences and were thus considered as the two domains (see Methods 4.3 and supplementary figure S2.13). To test for model robustness and to try to include a small omitted portion of the putative domain III sequence we did an expert version of these sequences. The later analysis revealed that there was no significant difference between the two versions, subsequently the expert version will not be presented.

The per-site protein distance matrix has been calculated with PROTDIST (belonging to the PHYLIP package) implementing amino-acid substitution models. Three different substitution models were tested: PAM, JTT and PMB (see Methods 4.5). Since PMB performed very poorly and gave results far from those in Remington *et al.* [2004] we did not use this model. PAM and JTT models, which are very close in their methodology, yielded similar results. We thus chose to focus on protein distance matrices obtained with the PAM substitution model.

In order to test for specific influences of domains III and IV, we also computed two separate distance matrices, corresponding to domains III and IV respectively, and to be used in models with two explanatory variables.

2.5.2 Introduction of phylogenetic distances as an explanatory variable

The linear regression model with a single explanatory variable is written as follows:

$$X_{ij}|\{i \in C_q, j \in C_\ell\} \sim \mathcal{N}(\mu_{q\ell} + \beta_{q\ell}Y_{ij}, \sigma^2), \quad (2.3)$$

where X is the weighted adjacency matrix (response distance matrix) and Y the explanatory phylogenetic distance matrix.

Number of Clusters	1	2	3	4	5	6	7
ICL	-570.028	-343.172	-277.012	-272.276	-282.175	-290.605	-316.304
Posterior probability	0	0	0.009	0.991	0	0	0

Table 2.7: ICL and corresponding posterior model probabilities for network A based on a LM model with III & IV as a single explanatory variable.

Linear regression model with a single explanatory variable. Introducing an explanatory variable enables to reduce the number of clusters selected by the ICL criterion (four clusters for the LM(A) model instead of five clusters for the GM(A) model); see tables 2.4 and 2.7. As for the BM- and GM-models, the posterior probabilities of protein assignment to clusters are degenerate (probability very close to 1 for a cluster and 0 for the others), making model parameters interpretation straightforward.

Cluster 1	ARF5 (0.022), ARF6 , ARF7 , ARF8 , ARF19 (0.024), IAA31 (0.027), IAA7 (0.029), IAA13 (0.029)
Cluster 2	ARF1 (0.012), ARF10, ARF16, IAA6, IAA11, ARF4 (0.013), ARF14, ARF18, ARF2 (0.015), ARF13, ARF12 (0.016)
Cluster 3	IAA10 (0.029), IAA15, IAA14 (0.03), IAA12 (0.031), IAA1, IAA2, IAA18, IAA27 (0.033), IAA17, IAA19, IAA28, IAA4 (0.035), IAA16, IAA34, IAA3, IAA5, IAA8, IAA9 (0.037)
Cluster 4	IAA29 (0.018), ARF22, IAA33, ARF20, IAA26, IAA32, IAA20, IAA30, ARF9 (0.026)

Table 2.8: Proteins clustering for network A based on a 4-clusters LM model with one explanatory variable. The distance $D(i, q)$ between protein i and cluster q to which it is assigned is given for the most central, the most peripheral and some other proteins of interest for interpretation purposes. See supplementary figure S2.8 for the distance plot.

A quick look at table 2.8, presenting the clusters composition, ascertains the fact that we still have biologically meaningful clusters. Moreover matching the elements found in each clusters for the 4-clusters models (table 2.11) highlights the clusters composition coherence regarding the to previously introduced models (87% match with GM and 76% with BM). The estimated slopes of the

linear regression models with a single explanatory variable are as follows (see also supplementary figure S2.9 for a graphical representation):

$$\hat{\beta}_{\text{LM,III\&IV}}(q, \ell) = \begin{pmatrix} \text{ARF+} & \text{ARF-} & \text{IAA} & \text{Outliers} \\ 1.024 & 0.097 & 0.305 & 0.701 \\ 0.097 & -0.092 & -0.057 & 0.119 \\ 0.305 & -0.057 & -0.031 & -0.014 \\ 0.701 & 0.119 & -0.014 & -0.081 \end{pmatrix}$$

In light of the previous matrix, when self-interacting, the TFs found in the ARF+ cluster are under a positive linear influence ($\hat{\beta}(\text{ARF+}, \text{ARF+}) = 1.024$) linking interaction and phylogenetic distance, meaning that the closer the dimerisation sequences, the less ARF+ will self-interact. However, as shown in table 2.8, the ARF+ cluster composition is not only made of ARF+ but also IAA31, 7 and 13. A closer look at supplementary figure S2.9 shows that the linear influence detected for self interaction of cluster ARF+ might come from these IAA, since the plot seems to present two groups. One with a low interaction and phylogenetic distances, for which no interaction was detected and therefore most probably the ARF+ of the cluster, and another one presenting the opposite behaviour, thus most probably the IAA of the cluster. The supplementary figure S2.10 ascertains this observation.

Considering the ARF+ cluster interactions with the Aux/IAAs cluster, there is a weak but positive linear effect ($\hat{\beta}(\text{ARF+}, \text{IAA}) = 0.305$) of the primary proteic sequence on the interaction capacity. Surprisingly, we cannot observe any effect of the phylogenetic distances on Aux/IAA self-interactions likelihoods ($\hat{\beta}(\text{IAA}, \text{IAA}) = -0.031$). Apart for the –difficult to interpret– interaction between ARF+ and outliers clusters, no other influence of the dimerisation sequence upon interaction capacities between TFs can be identified with this model.

Linear regression model with specific domains III & IV explanatory variables:

Using the same linear regression model with two explanatory variables, one for each dimerisation domain, it was possible to test for a differential influence of these two domains. The single covariable model given in equation (2.3) can be re-written as a model with two explanatory variables as follows:

$$X_{ij}|\{i \in C_q, j \in C_\ell\} \sim \mathcal{N}(\mu_{q\ell} + \beta_{\text{III},q\ell}Y_{\text{III},ij} + \beta_{\text{IV},q\ell}Y_{\text{IV},ij}, \sigma^2), \quad (2.4)$$

Number of Clusters	1	2	3	4	5	6
ICL	-532.263	-334.018	-293.711	-295.069	-312.373	-354.551
Posterior probability	0	0	0.795	0.205	0	0

Table 2.9: ICL criterion and corresponding posterior model probabilities for network A based on a LM model with two explanatory variables.

The ICL criterion computed for each clusters number attached to the LM model with two explanatory covariables is given in table 2.9. According to the ICL criterion in this table, the relevant clustering for network A comprises three or four clusters (preferentially three for network A and four for network B, data not shown). Using linear regression models with two explanatory variables, for each conserved domains, therefore improved our understanding of the interaction network by yielding simpler models (i.e. with fewer clusters).

Cluster 1	ARF5 (0.022), ARF6 , ARF7 , ARF8 , ARF19 (0.024), IAA31 (0.027), IAA7 (0.029), IAA13 (0.029)
Cluster 2	ARF1 (0.012), ARF10, IAA6, IAA11, ARF4 (0.013), ARF14, ARF16, ARF18, IAA29, ARF20 (0.014), ARF12 (0.015), ARF13, ARF2 (0.016)
Cluster 3	IAA10 (0.029), IAA15, IAA14 (0.03), IAA12 (0.031), IAA1, IAA2, IAA18, IAA27 (0.033), IAA17, IAA19, IAA28, IAA4 (0.035), IAA16, IAA34, IAA3 (0.036), IAA5, IAA8, IAA9 (0.037)
Cluster 4	IAA33 (0.018), ARF22, IAA30, ARF9, IAA20, IAA26, IAA32 (0.025)

Table 2.10: Proteins clustering for network A based on a 4-clusters LM model with two explanatory variables. The distance $D(i, q)$ between protein i and cluster q to which its is assigned is given for the most central, the most peripheral and some other proteins of interest for interpretation purposes. See supplementary figure S2.11 for the distance plot.

Analysing the clusters composition again proves their biological relevance, since an over-representation of each known biological group is achieved for three clusters. In addition, matching their TFs composition (see table 2.11) shows the strong coherence of this two-covariate 4-clusters model with the single covariate (95%) and the GM(A) (91%) ones. The estimated slope of linear regression models for each domains in case of the network A are as follows (see also supplementary figure S2.12 for a graphical representation):

$$\hat{\beta}_{\text{III}}(q, \ell) = \begin{pmatrix} \text{ARF+} & \text{ARF-} & \text{IAA} & \text{Outliers} \\ 0.021 & 0.079 & 0.294 & 0.569 \\ 0.079 & 0.004 & 0.194 & 0.088 \\ 0.294 & 0.194 & -0.268 & -0.219 \\ 0.569 & 0.088 & -0.219 & -0.050 \end{pmatrix}$$

$$\hat{\beta}_{IV}(q, \ell) = \begin{pmatrix} \text{ARF+} & \text{ARF-} & \text{IAA} & \text{Outliers} \\ 0.887 & 0.109 & 0.052 & 0.069 \\ 0.109 & -0.045 & -0.138 & 0.037 \\ 0.052 & -0.138 & 0.297 & 0.208 \\ 0.069 & 0.037 & 0.208 & 0.004 \end{pmatrix}$$

Considering ARF+ cluster self-interaction capacities, domains IV (DIV) show a positive linear effect of the phylogenetic distances upon interaction likelihoods ($\hat{\beta}_{IV}(\text{ARF+}, \text{ARF+}) = 0.887$), meaning the closer the dimerisation sequences, the less they will interact. However, domains III (DIII) primary sequences seem to have no influence onto interaction capabilities ($\hat{\beta}_{III}(\text{ARF+}, \text{ARF+}) = 0.021$). Therefore, there is a domain specific response for ARF+ homo-dimerisation, where DIV plays a more important role than DIII. The previously detected strong effect of the dimerisation sequences by the LM(A) model with a single explanatory variable ($\hat{\beta}(\text{ARF+}, \text{ARF+}) = 1.024$) was thus attributable to DIV. But again we will qualify this conclusion remembering it applies to all members of the ARF+ cluster, also made of three IAA, again crucial for the detected relation attributed to DIV.

Aux/IAAs cluster on the other hand, when self-interacting, present an antagonistic behaviour for their dimerisation domains: closely related DIII will interact better ($\hat{\beta}_{III}(\text{IAA}, \text{IAA}) = -0.268$), when the farther the DIV sequences are, the more they can interact ($\hat{\beta}_{IV}(\text{IAA}, \text{IAA}) = 0.297$). Again, we observe a domain specific response for Aux/IAA homo-dimerisation capacities regarding their protein primary sequences. This –almost equally– antagonistic behaviour can explain why we could not observe any effect of the phylogenetic distances onto the interaction likelihoods with the single explanatory variable model ($\hat{\beta}(\text{IAA}, \text{IAA}) = -0.031$).

Another domain specific response also appears for the interaction between ARF+ and Aux/IAA clusters. DIII sequences are indeed positively related to interaction ($\hat{\beta}_{III}(\text{ARF+}, \text{IAA}) = 0.294$), but no –or a very limited– effect of DIV is observed ($\hat{\beta}_{IV}(\text{ARF+}, \text{IAA}) = 0.052$). This is in agreement with the detected relationship ($\hat{\beta}(\text{ARF+}, \text{IAA}) = 0.305$) by the single explanatory variable LM model.

Finally, while having a few detected interactions between ARF- and Aux/IAA clusters in the binary network, with the two-covariate LM(A) clustering model we detect a weak antagonistic effect ($\hat{\beta}_{III}(\text{ARF-}, \text{IAA}) = 0.194$ and $\hat{\beta}_{IV}(\text{ARF-}, \text{IAA}) = -0.138$) of each domains relating dimerisation

sequences to interaction capabilities. Again this could not be inferred from the single explanatory variable model ($\hat{\beta}(\text{ARF-}, \text{IAA}) - 0.057$).

	BM	GM A	GM B	LM A 3/4	LM B 3/4	LM A 3/4	LM B 3/4
BM	46	36	34	35	34	35	36
GM A		46	44	40	39	42	37
GM B			46	40	39	42	37
LM A 3/4				46	41	44	38
LM B 3/4					46	43	37
LM A 3/4						46	38
LM B 3/4							46

Table 2.11: Cluster composition matching for 4-clusters models.

2.5.3 Interpretation of the patterns found within the auxin interactome

The initial BM model led us to the conclusion that the auxin transcription network could be simplified in three biologically meaningful clusters (roughly the ARFs+, ARF- and Aux/IAA groups). The strong connectivity between ARFs+ and Aux/IAs was expected from the putative molecular model reviewed in Guilfoyle and Hagen [2007]. This suggest that most Aux/IAs repress the transcriptional activity of ARFs+ when a low concentration of auxin is encountered. However, the weak interaction between ARFs- and Aux/IAs or ARFs- and ARFs+ was a surprising conclusion that was highlighted by the binary graph. We thus propose that ARF- might simply act as competitors for DNA binding against ARF+, and therefore might be auxin independent. Also, the existence of more than one TFs binding site for some auxin-inducible genes might suggest an unknown transcription mechanism as a supplementary control [Cherry and Adler, 2000]. Anyway, further experiments and analyses need to be conducted to unveil the complete transduction mechanism.

Using LM models to investigate the influence of phylogenetic distances on the auxin signalling network is a first attempt to establish a direct link between the proteic primary sequences and the interaction network structures. By first using a single explanatory variable model built from the global protein distance matrix of domains III and IV, we observed contrasted modes of interaction. While TFs within the ARF+ cluster tend to homo-dimerise when their domains are different, Aux/IAs preferentially interact with phylogenetically close relatives. It is quite surprising to observe such an easy auto-dimerisation knowing the repressor role of Aux/IAA against ARF+. Ultimately, ARF+ cluster interact with Aux/IAs cluster when their dimerisation sequences are different.

The two-covariables LM models yielded a more precise view of a very probable domain specific behaviour. Domain III seems indeed to be important for Aux/IAAs to interact with everyone, including themselves. Although the relations might be of few importance for Aux/IAAs dimerisation with ARF-. Domain IV is mostly important for homo-dimerisation of the TFs within the ARF+ cluster, while it presents the particularity to interact through phylogenetically distant domains IV.

We compared compositions of the outliers cluster for the different models (BM, GM A, LM A 34 and LM A 3/4). Four transcription factors (ARF22, Aux/IAA 26, 32 & 33) were systematically assigned to the outliers cluster while three others (ARF9, Aux/IAA20, 30) were assigned to the outliers cluster for models estimated on the basis of the valued graph (GM A, LM A 34 and LM A 3/4). There is thus a high commonality of outliers between the different clusterings models, and this can be interpreted as a consequence of the Y2H experiments.

3 Conclusions

Being able to establish the biological meaning of the auxin interactome based on its connectivity structure, itself extracted from a –noisy– high-throughput experiment like Y2H was not obvious at first. However by first using a model-based clustering method for binary graph (BM) we have highlighted the possibility of simplifying the auxin signalling interactome structure in three clusters [Vernoux *et al.*, 2011]. These three clusters were highly coincidental with known biological groups, but which were established by generalisation of specific knowledge established by working on sub-parts of the network [Hardtke and Berleth, 1998; Hardtke *et al.*, 2004] together with phylogenetic analyses like [Remington *et al.*, 2004].

In this work, using a valued version of the network, instead of a binary graph, we overcome the difficulty of defining thresholds for each Y2H tests, and we could model the –even limited– self-interacting capacities of the ARF+ (which had been removed during the binarisation step). Comparing the outputs of the BM and GM graph clustering models led us to the conclusion that the binarisation is likely to remove experimental noise, therefore helping the model to perform better. However, we would like to stress that this transformation of the data can induce a bias from both thresholds selection and decision rules definition.

In addition, the strength of valued graph approach (GM model), is that it allowed us to make use of –existing– more complex models to explain interaction likelihoods using covariables such as the phylogenetic distances. Using the valued graph and phylogenetic distances as a global covariable,

we could demonstrate an increase of the interaction likelihoods between the ARF+ and Aux/IAAs clusters with an increase of the phylogenetic distance between the III-IV dimerisation domains. These observations lead to the surprising conclusion that interactions between less related proteins are favoured by an increase of the phylogenetic distance between their dimerisation domains. For the cluster to which ARF+ belong, we detected a positive relation between phylogenetic distances and the interaction likelihoods. However this is likely an artefact due to the presence of Aux/IAAs in this cluster, preventing us from drawing conclusion from this observation.

Then, separately analysing the influence of subdomains (DIII and DIV) phylogenetic distances upon the clusters interaction capacities, highlighted that the previously observed phenomena is linked to different conserved blocks of dimerisation sequences. For interactions between ARF+ and Aux/IAA clusters, their likelihoods are modulated by domain IV phylogenetic distances. In the same time, it has been demonstrated by Nanao *et al.* [2014] and Korasick *et al.* [2014] that DIII-DIV mediates interactions between ARF+ and Aux/IAA through two charged interfaces: one face mostly positive and one face mostly negative. In addition, these interactions are bi-directional. DIII contributes principally to the positive face, while DIV contributes to the negative face of these interaction domains. Our results on the impact of phylogenetic distances upon the interaction likelihoods between ARF+ and Aux/IAA clusters suggest that changes in the primary sequence of DIV unlikely modify the corresponding charged face, but rather impact the entire interaction domain, thus contributing to the global interaction capacity of the ARF+ and Aux/IAA clusters.

So far DIII-DIV structures have been obtained only for ARF5 and 7. Obtaining further structures, although challenging, could allow testing the hypotheses emerging from our clustering approach. Other strategies would also allow testing further the link between interaction capacities and phylogenetic distances:

- creating a library of mutated version of DIII-DIV for each elements of the network to artificially enlarge the size of the population;
- generating similar Y2H data for other species, such as rice or tomato, which also posses large families of auxin related transcription factors.

This would be particularly useful for small clusters like ARF+ where there is only few members, limiting the regression model.

4 Methods

4.1 Construction of the Aux/IAA-ARF interaction network on the basis of the yeast-2-hybrid data.

4.1.1 Yeast-2-hybrid protein interaction testing:

We managed to clone full-length cDNA for all members of the Aux/IAA-ARF family except for ARF 15 and 21. We were therefore unable to test for their interaction and conducted the initial screening on 50 transcription factors, representing 1275 interactions to test for. To ensure the robustness of the biological test, all the interactions were tested both ways, as bait and prey (e.g. AD-ARF1 v.s. BD-ARF2 and AD-ARF2 v.s. BD-ARF1). Finally, knowing that this screening method can present false positives, two independent biological tests were conducted for each way, representing a total of 5000 test results. The Y2H experiment is a bio-engineered tool based on the Gal-4 transcription factor from yeast *Saccharomyces Cerevisiae*. The Gal-4 transcription factor protein is made of an N-terminal DNA binding domain (BD) and a C-terminal activation domain (AD). These two parts have been artificially separated so each part can receive two proteins to test for their eventual interaction. In order to be thorough each interaction have been tested both ways, meaning each protein has been append to both AD and BD in two separate repetitions. Also, two reporter genes have been used. The β -galactosidase (β -gal), which is an enzyme hydrolysing the X-Gal (or 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranosid) into a blue compound revealing its activity. The other reporting gene, HIS3, is also coming from *S. Cerevisiae* and allow the yeast to produce histidine, and therefore to survive in a medium without it. The output of the biological test is a ratio of optical densities (OD): $\frac{\text{OD no histidine medium}}{\text{OD histidine rich medium}}$. It can be viewed as an estimation of histidine synthesis capacity upon function recovery.. For detailed explanations on the constructs used in the Y2H screen, see Vernoux *et al.* [2011].

4.1.2 Data description:

The X-Gal test is based on a blue coloration of the media where yeasts are developing. The output of the tests then took the form of a mark chosen among the following ordered marks: -; +; ++; ++++. The HIS3 test is based on the capacity of yeasts to synthesise histidine in a histidine depleted liquid media. To assess for this synthesis capacity a ratio of optical density between yeasts growth in a

media without histidine and with histidine -normalise the OD for each test- was used. This reflect the efficiency of histidine synthesis function recovery. For detailed explanations on the tests outputs used in the Y2H screen, see Vernoux *et al.* [2011].

4.2 Network binarisation.

4.2.1 Mixture model over OD ratios under the X-Gal ‘+’ mark:

We estimated a three-component Gaussian mixture model $\sum_{i=1}^3 \alpha_i f_i(z; \mu_i, \sigma_i^2)$ using the *mclust* R package Fraley and Raftery [2006].

4.2.2 Decision rules:

Because it is a two-way two-reporting-gene experiment (for each potential interaction), there is several possible configurations for which the interaction tests should define the presence or absence of interaction. In the following tables we define a given test as “positive” (+) and “negative” (-) when its result is respectively above or below the defined thresholds:

Configuration 1 : all tests are positive (see table 2.12),

Table 2.12: Table for decision rule 1.

	X-Gal	HIS3
Way 1	+	+
Way 2	+	+

Configuration 2 : only one test is not positive (see table 2.13),

Table 2.13: Tables for decision rule 2.

	X-Gal	HIS3			X-Gal	HIS3
Way 1	-	+	or	Way 1	+	-
Way 2	+	+		Way 2	+	+
or				or		
	X-Gal	HIS3			X-Gal	HIS3
Way 1	+	+	or	Way 1	+	+
Way 2	-	+		Way 2	+	-

Configuration 3 : only one way is positive for both reporter genes (see table 2.14),

Configuration 4 : one reporter gene is positive in each way (see table 2.15),

Configuration 5 : only one reporter gene is positive both ways (see table 2.16),

Table 2.14: Tables for decision rule 3.

	X-Gal	HIS3	or		X-Gal	HIS3
Way 1	–	–		Way 1	+	+
Way 2	+	+		Way 2	–	–

Table 2.15: Tables for decision rule 4.

	X-Gal	HIS3	or		X-Gal	HIS3
Way 1	–	+		Way 1	+	–
Way 2	+	–		Way 2	–	+

Table 2.16: Tables for decision rule 5.

	X-Gal	HIS3	or		X-Gal	HIS3
Way 1	–	+		Way 1	+	–
Way 2	–	+		Way 2	+	–

An analysis -not detailed here- allowed us to state that the fifth configuration (only one reporter gene is positive both ways, table 2.16) is unreliable. We therefore discarded this case when defining the presence or absence of interaction for the binary network.

4.3 Phylogenetic data

Protein sequences were obtained using the 49 accession numbers of Aux/IAA and ARF proteins presenting a III-IV dimerisation domains (ARFs 3, 17 and 23 were thus excluded); see availability of supporting data for list of AGIs. Sub-sequences corresponding to domains III and IV were obtained by first making a multiple alignment of the whole protein sequences using ClustalW Thompson *et al.* [1994], then we searched for highly conserved regions using Gblocks 0.91b Castresana [2000]; see availability of supporting data for list of used parameters. We found in this way 3 conserved regions, the two last corresponding to the III and IV interacting domains (for more information see additional file S2.13).

The detected flanking position of domains III and IV of full Aux/IAs and ARF sequences were respectively [1275-1307] and [1344-1376]. Both selected conserved domains have a length of 32 amino-acids. The sequences for the III-IV domain is made of the concatenation of the two separate domains. We also conducted an analysis with a slightly different flanking position: [1272-1307] and [1344-1376], but this did not lead to significant changes in the analysis outcomes.

4.4 Response distance matrix: standardized distances between transcriptional factors

The Y2H analysis involves two independent tests, the X-Gal and the HIS3 tests. The output of the X-Gal test takes the form of a mark chosen from among the following ordered marks: -; +?; +; ++;+++ . The output of the HIS3 test takes the form of an optical density ratio reflecting the efficiency of recovery (ratio of the OD in the absence of histidine to the OD in the presence of histidine). Each possible interaction was tested in the two possible configurations where each protein was alternatively the bait and the prey protein.

The output of the X-Gal test can be interpreted as a distance defined on an ordinal scale (from no interaction to strong interaction) while the output of the HIS3 test can be interpreted as a distance defined on a ratio scale (between 0 and 1.5). Combining these observed distances requires a standardization procedure. The objective of standardization is to avoid dependency on the elementary distance type and scale. In the case of an ordinal distance (X-Gal test), observed distances are replaced by ranked distances

$$\text{Rank}(y_{ij}) = \frac{1}{2} + \sum_{n=0}^{y_{ij}-1} f_n + \frac{f_{y_{ij}}}{2},$$

where y_{ij} is the output of the X-Gal test for the i th and j th proteins and f_n is the frequency of mark n (the possible marks are assumed to be represented as contiguous positive integers).

In this case, the normalization quantity is the mean rank $(1 + N^2) / 2$, where N is the number of proteins. The ratio-scaled distance (HIS3 test) can be either treated as an interval-scaled distance or as an ordinal distance. Considering that the response curve of the HIS3 test is monotone but highly non-linear and is close to a Michaelis–Menten kinetics, we chose to consider the output of the HIS3 test as a distance defined on an ordinal scale for standardization. Observed distances are replaced by ranked distances and the standardized distances are:

$$x_{ij} = w_{\text{X-Gal}} \frac{\text{Rank}(y_{ij}) + \text{Rank}(y_{ji})}{1 + N^2} + w_{\text{HIS3}} \frac{\text{Rank}(z_{ij}) + \text{Rank}(z_{ji})}{1 + N^2}.$$

It should be noted that a single marginal distribution was considered for each test used in the two possible configurations in order to standardize the distances. In the case of missing test values,

the distances can be straightforwardly adapted. If z_{ji} is missing, we obtain:

$$x_{ij} = w_{\text{X-Gal}} \frac{\text{Rank}(y_{ij}) + \text{Rank}(y_{ji})}{1 + M_{\text{X-Gal}}} + w_{\text{HIS3}} \frac{\text{Rank}(z_{ij})}{(1 + M_{\text{HIS3}})/2},$$

where $M_{\text{X-Gal}}$ is the number of X-Gal test values, M_{HIS3} the number of HIS3 test values.

The distance matrices $\{x_{ij}; i, j = 1, \dots, N\}$ corresponding to $(w_{\text{X-Gal}}, w_{\text{HIS3}}) = (1, 0), (0.75, 0.25), (0.5, 0.5), (0.25, 0.75), (0, 1)$ were built and tested.

4.5 Phylogenetic distance matrix

To use the primary sequence information as an explanatory variable in the mixture model for random graph, we have to define a distance between two protein sequences. ProtDist belong to the PHYLIP package (<http://evolution.genetics.washington.edu/phylip.html>) and allow to compute such distance by using protein substitution models. One can choose between five different models, and we tested three of them PAM, JTT and PMB. PMB which performed poorly was not used in the analyses. For more information about the proteins substitution models, see the PROTDIST documentation (<http://evolution.genetics.washington.edu/phylip/doc/protdist.html>).

4.6 Assessing the adequacy of the clustering

We assessed the adequacy of the clustering obtained by evaluating the separability of the clusters and the dispersion of the proteins within the clusters. Since, in our case, the assignment of proteins to clusters is almost deterministic (i.e. $\tau_{iq} \simeq 1$ for a unique cluster q and $\tau_{i\ell} \simeq 0$ for $\ell \neq q$ where τ_{iq} is the posterior probability of assigning protein i to cluster q), this assignment can be viewed as a partition. The model parameters, which parameterized the edges of the graph, cannot be used directly to define dispersion measures of the proteins assigned to a given cluster. We thus used the adjacency information to derive dissimilarity measures for the proteins. The distance $D(i, j) = \sum_k |x_{ik} - x_{jk}|/N$ between the i th and j th rows of the weighted adjacency matrix $\{x_{ij}; i, j = 1, \dots, N\}$ quantifies the difference in connectivity profile between proteins i and j . In the case of the binary adjacency matrix, this distance is the Sokal-Michener distance between proteins i and j Kaufman and Rousseeuw [1990] $D(i, j) = \sum_k I(x_{ik} \neq x_{jk})/N$, where $I(\cdot)$ denotes the indicator function. This is the proportion of mismatches between the i th and j th rows of the

adjacency matrix.

The distance between protein i and cluster q is given by

$$D(i, q) = \frac{\sum_{j \neq i} \tau_{jq} \sum_k |x_{ik} - x_{jk}|}{\left\{ \sum_{j \neq i} \tau_{jq} \right\} N}.$$

If the proteins are deterministically assigned to a given cluster, this distance simplifies to

$$\begin{aligned} D(i, q) &= \frac{\sum_{j \in q; j \neq i} \sum_k |x_{ik} - x_{jk}|}{(n_q - 1)N} & i \in q, \\ D(i, q) &= \frac{\sum_{j \in q} \sum_k |x_{ik} - x_{jk}|}{n_q N} & i \notin q, \end{aligned}$$

where n_q is the number of proteins assigned to cluster q .

The distance between cluster q and cluster ℓ can be directly derived as

$$\begin{aligned} D(q, q) &= \frac{\sum_{i, j \in q; i \neq j} \sum_k |x_{ik} - x_{jk}|}{n_q(n_q - 1)N}, \\ D(q, \ell) &= \frac{\sum_{i \in q} \sum_{j \in \ell} \sum_k |x_{ik} - x_{jk}|}{n_q n_\ell N} & q \neq \ell. \end{aligned}$$

The within- and between-cluster distances can then be defined as

$$\begin{aligned} D_{\text{within}}(q) &= D(q, q) && \text{within cluster,} \\ D_{\text{between}}(q) &= \frac{\sum_{i \in q} \sum_{j \notin q} \sum_k |x_{ik} - x_{jk}|}{n_q(N - n_q)N} && \text{between cluster.} \end{aligned}$$

4.7 Availability of supporting data

4.7.1 Original Yeast-2-Hybrid data for Aux/IAA - ARFs interaction tests

All Yeast-2-hybrid interaction results for X-Gal and HIS3 reporters are available in supplementary data of Vernoux *et al.* [2011].

4.7.2 Aux/IAA - ARFs protein sequences

Protein sequences can be found within *Arabidopsis thaliana* proteins banks like <http://pfam.sanger.ac.uk> using the following 49 AGIs:

At1g59750, At2g28350, At2g46530, At1g34310, At1g34170, At1g35540, At1g35520, At4g30080, At3g61830, At1g19220, At5g62000, At1g35240, At1g34410, At1g34390, At5g60450, At1g19850,

At1g30330, At5g20730, At5g37020, At4g23980, At4g14560, At1g04100, At4g28640, At1g04550, At2g33310, At4g14550, At1g80390, At3g04730, At1g04250, At1g51950, At3g15540, At3g23030, At2g46990, At3g16500, At4g29080, At5g25890, At4g32280, At1g04240, At3g62100, At3g17600, At2g01200, At5g57420, At1g15050, At5g43700, At1g15580, At1g52830, At3g23050, At2g22670, At5g65670.

4.7.3 Domains III and IV sub-sequences

To obtain protein sub-sequences corresponding to conserved domains III and IV we used the following parameters:

- Minimum Number Of Sequences For A Conserved Position: 25
- Minimum Number Of Sequences For A Flanking Position: 25
- Maximum Number Of Contiguous Nonconserved Positions: 8
- Minimum Length Of A Block: 10
- Allowed Gap Positions: With Half
- Use Similarity Matrices: Yes

See supplementary figure S2.13 for a detailed view of the aligned sequences and selection of conserved sub-sequence.

Author contributions

Together with Yann GUÉDON, I analysed the structure of the interactome. Jean-Benoist LÉGER and Stéphane ROBIN participated to the MixNet and wMixNet models, algorithms and software. Yann GUÉDON, Teva VERNOUX and myself wrote the manuscript with inputs from the other authors.

Supplementary Materials

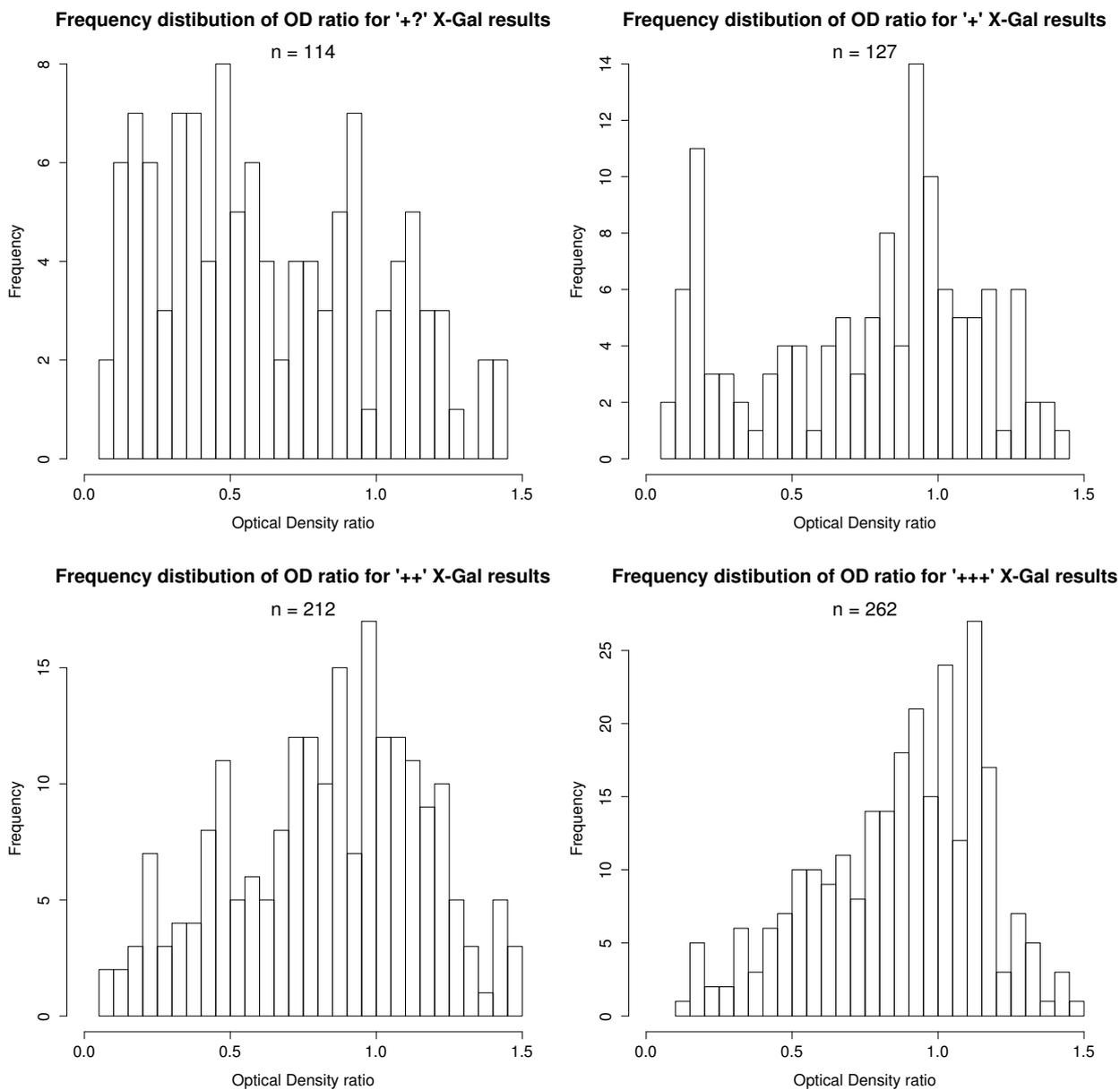


Figure S2.1: Histograms of OD ratios per X-Gal marks for the successive marks '+?', '+', '++' and '+++'. The OD ratio values have an upper limit of 1.5 for histogram readability reasons.

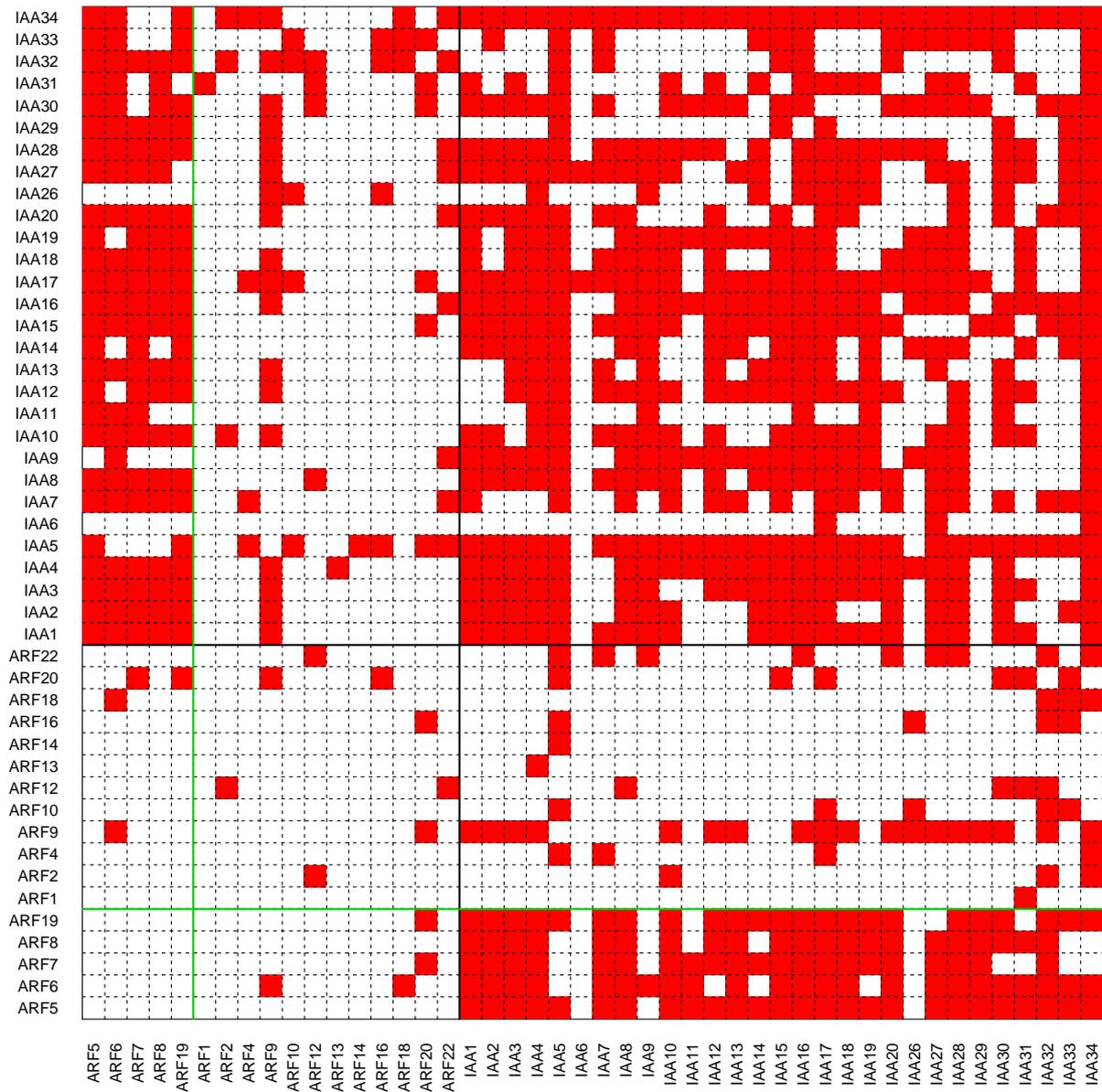


Figure S2.2: Adjacency matrix of detected interaction (red) the the threshods: HIS3> 0,3 and X-Gal>'+'.

Table S2.1: Distance between clusters matrix $D(q, \ell)$ for network A based on a GM model.

$$D_{GM,A}(q, l) = \begin{pmatrix} & \text{ARF+} & \text{ARF-} & \text{IAA} & \text{Outliers} \\ \text{ARF+} & 0.015 & 0.016 & 0.032 & 0.024 \\ \text{ARF-} & 0.016 & 0.013 & 0.016 & 0.016 \\ \text{IAA} & 0.032 & 0.016 & 0.032 & 0.022 \\ \text{Outliers} & 0.024 & 0.016 & 0.022 & 0.022 \end{pmatrix}$$

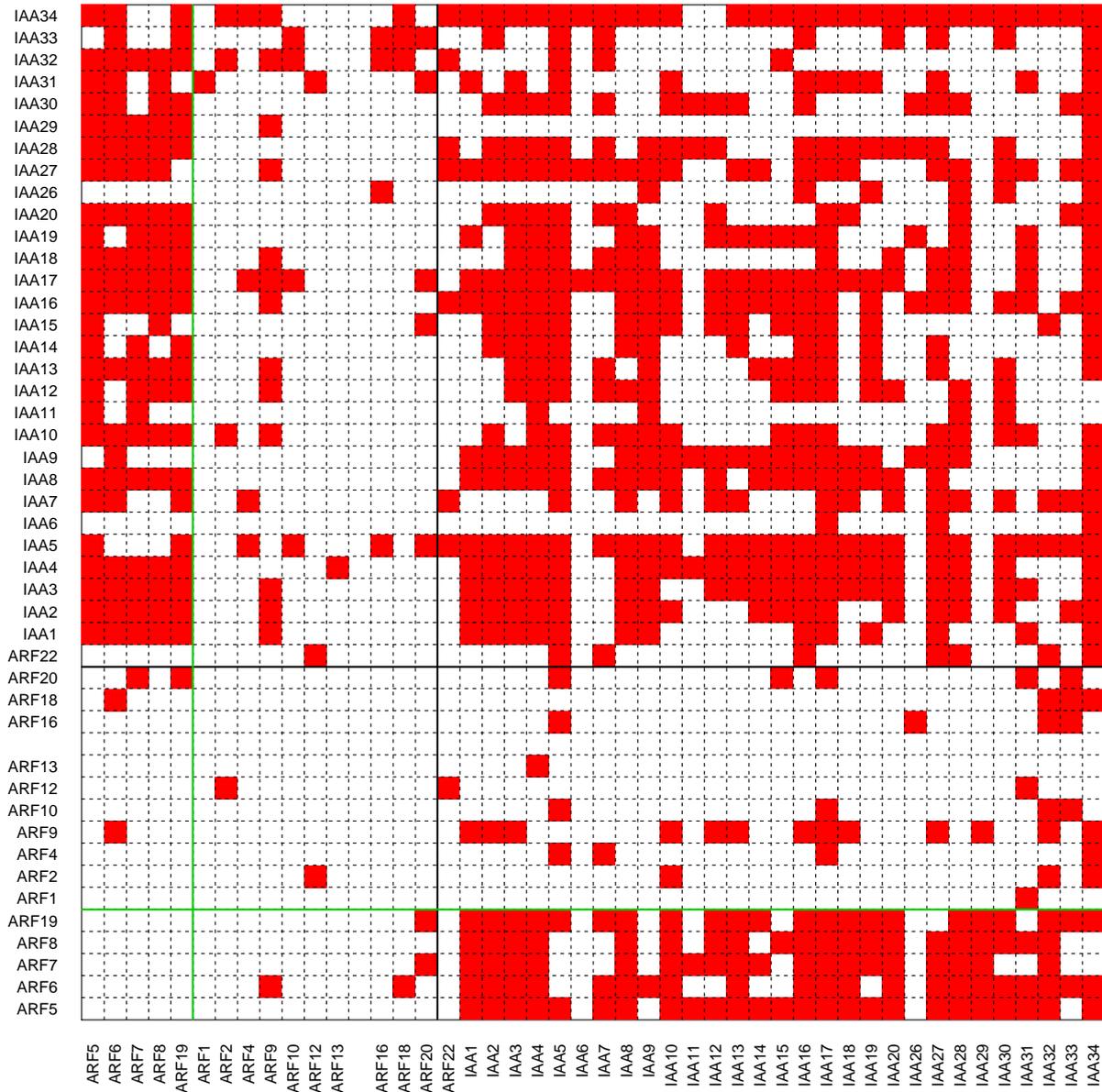


Figure S2.3: Adjacency matrix of detected interaction (red) the the thresholds: HIS3> 0,65 and X-Gal>'+'.

Table S2.2: Estimated model parameters for the 4-cluster BM model. $\hat{\alpha}$ are the cluster proportions and $\hat{\Pi}$ is the connectivity probability matrix.

$$\hat{\alpha} = (0.200, 0.238, 0.400, 0.162), \quad \hat{\Pi} = \begin{pmatrix} \text{ARF+} & \text{ARF-} & \text{IAA} & \text{Outliers} \\ 0.317 & 0.136 & 0.901 & 0.567 \\ 0.136 & 0.036 & 0.052 & 0.186 \\ 0.901 & 0.052 & 0.769 & 0.301 \\ 0.567 & 0.186 & 0.301 & 0.136 \end{pmatrix}.$$

Number of Clusters	1	2	3	4	5	6	7
ICL	-597.206	-354.034	-294.248	-285.443	-290.408	-301.826	-320.95
Posterior probability	0	0	0	0.993	0.007	0	0

Table S2.3: ICL and corresponding posterior model probabilities for network B based on a LM model with III & IV as one global explanatory variable.

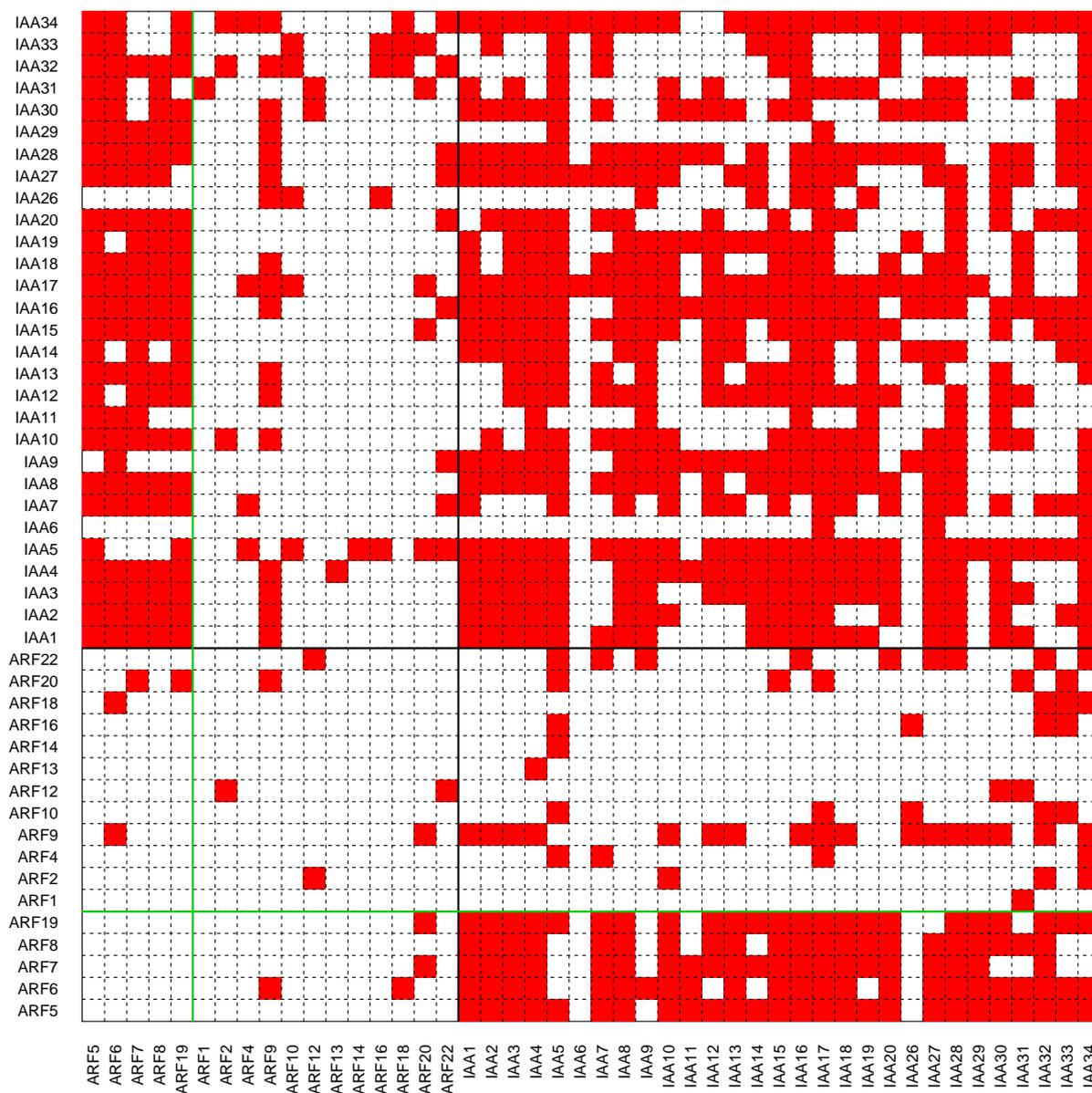


Figure S2.4: Adjacency matrix of detected interaction (red) the the tresholds: $HIS3 > 0,45$ and $X-Gal > +$.

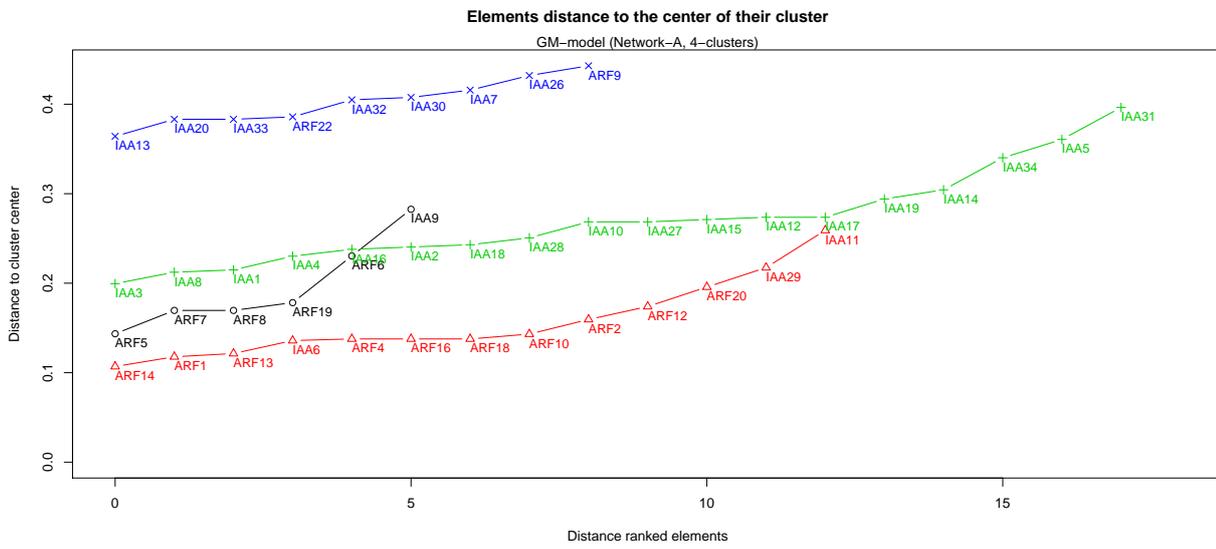


Figure S2.5: Ranked vertex distances to assigned cluster for network A based on a 4-clusters GM model.

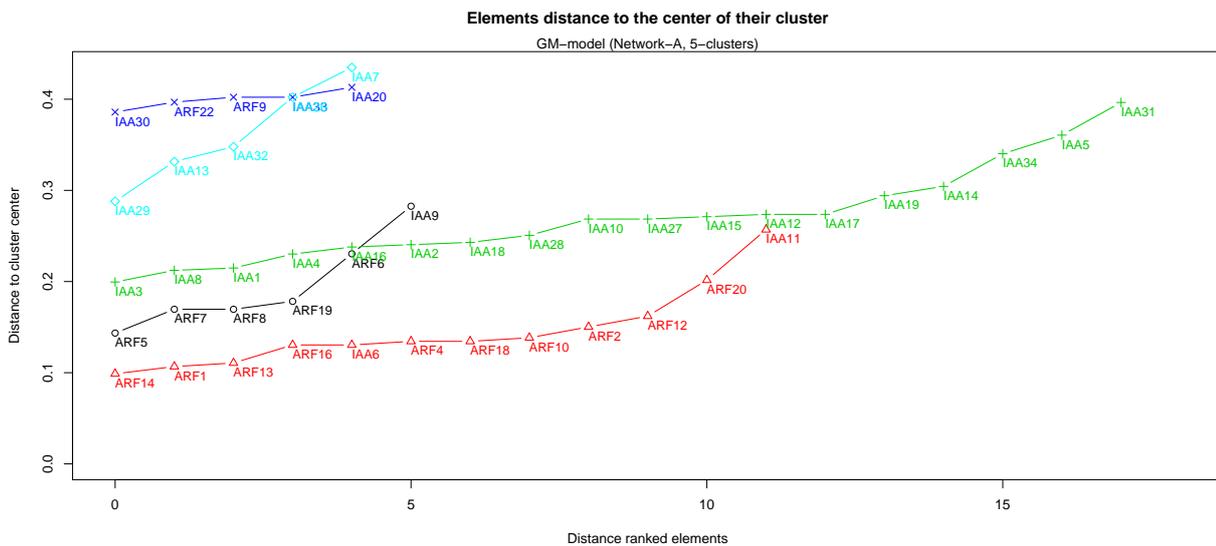


Figure S2.6: Ranked vertex distances to assigned cluster for network A based on a 5-clusters GM model.

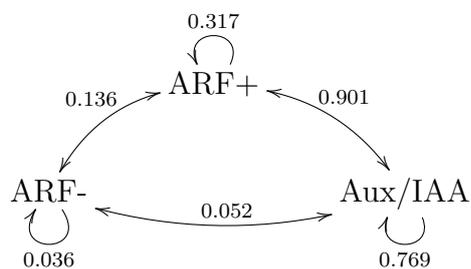


Figure S2.7: Connectivity graph and associated probabilities for the 4-cluster BM model restricted to the ARF+, ARF- and Aux/IAA clusters. The connectivity probability matrix describes the topology of the network at the cluster scale ($\pi_{q\ell}$ is the probability for a protein of cluster q to be connected with a protein of cluster ℓ). This matrix shows the heterogeneous behaviour of the 3 clusters corresponding to ARF+, ARF- and Aux/IAA. Instead of classical graph clustering methods which would present highly intra-connected and poorly inter-connected clusters (high probabilities on the diagonal and low elsewhere), the 3-clusters model captured the latent structure of the binary graph. The connectivity probabilities highlight that ARF- are very poorly connected to the rest of the network, ARF+ being on the other hand well-connected to Aux/IAAs, themselves well-connected to each other.

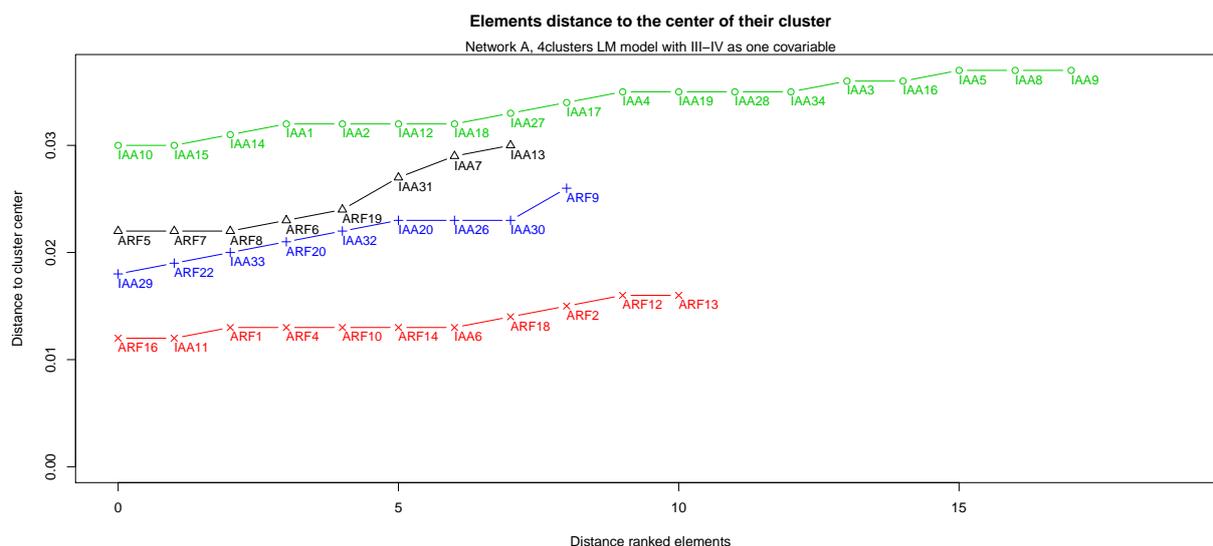


Figure S2.8: Ranked vertex distances to assigned cluster center for network A based on a LM model with one explanatory variable.

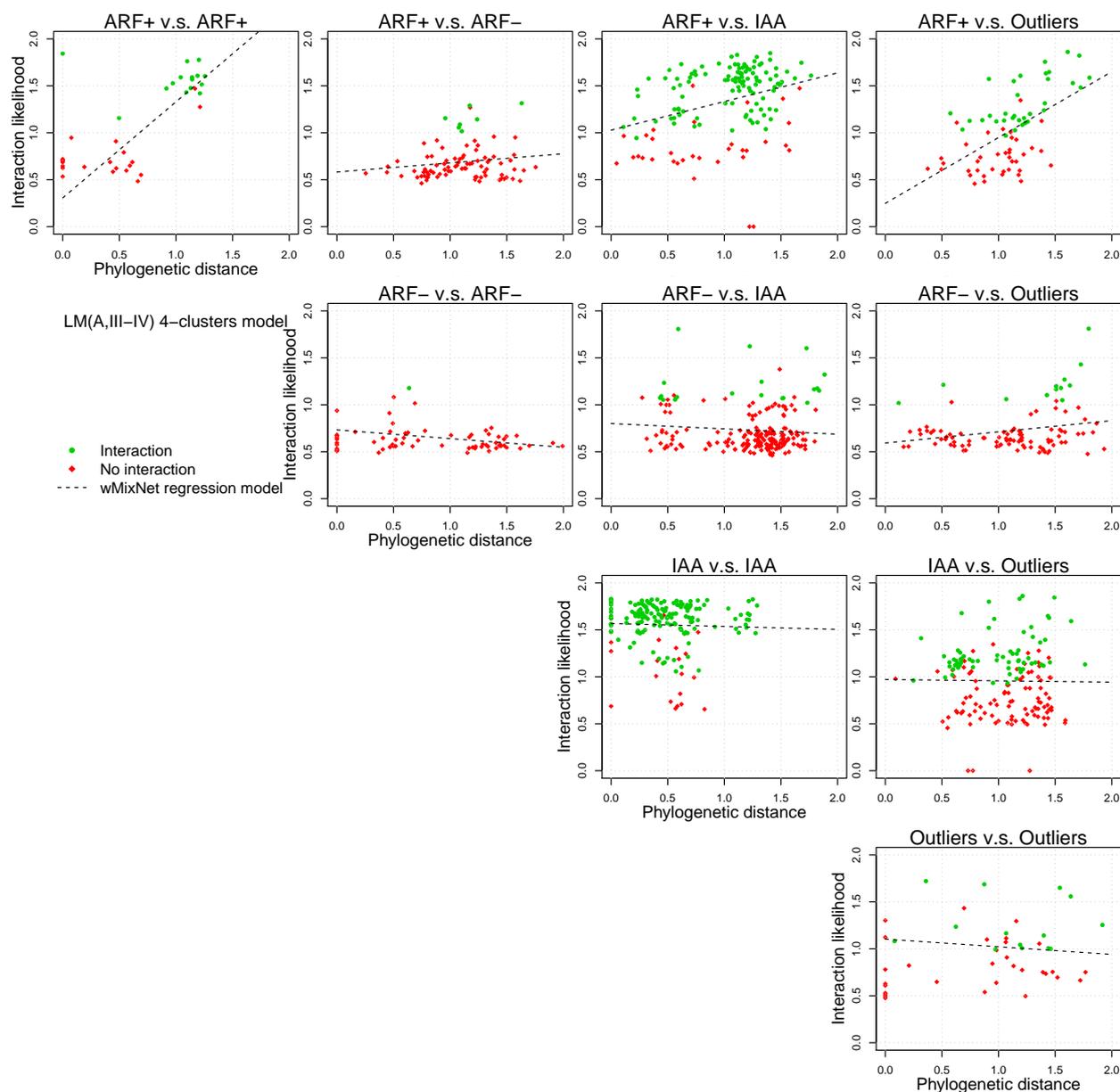


Figure S2.9: Linear regressions for network A based on a 4-clusters model with one explanatory variable for the different pair of groups: ARF+ v.s. ARF+, Aux/IAA v.s. Aux/IAA and ARF+ v.s. Aux/IAA. The outliers cluster is also presented in the graphic.

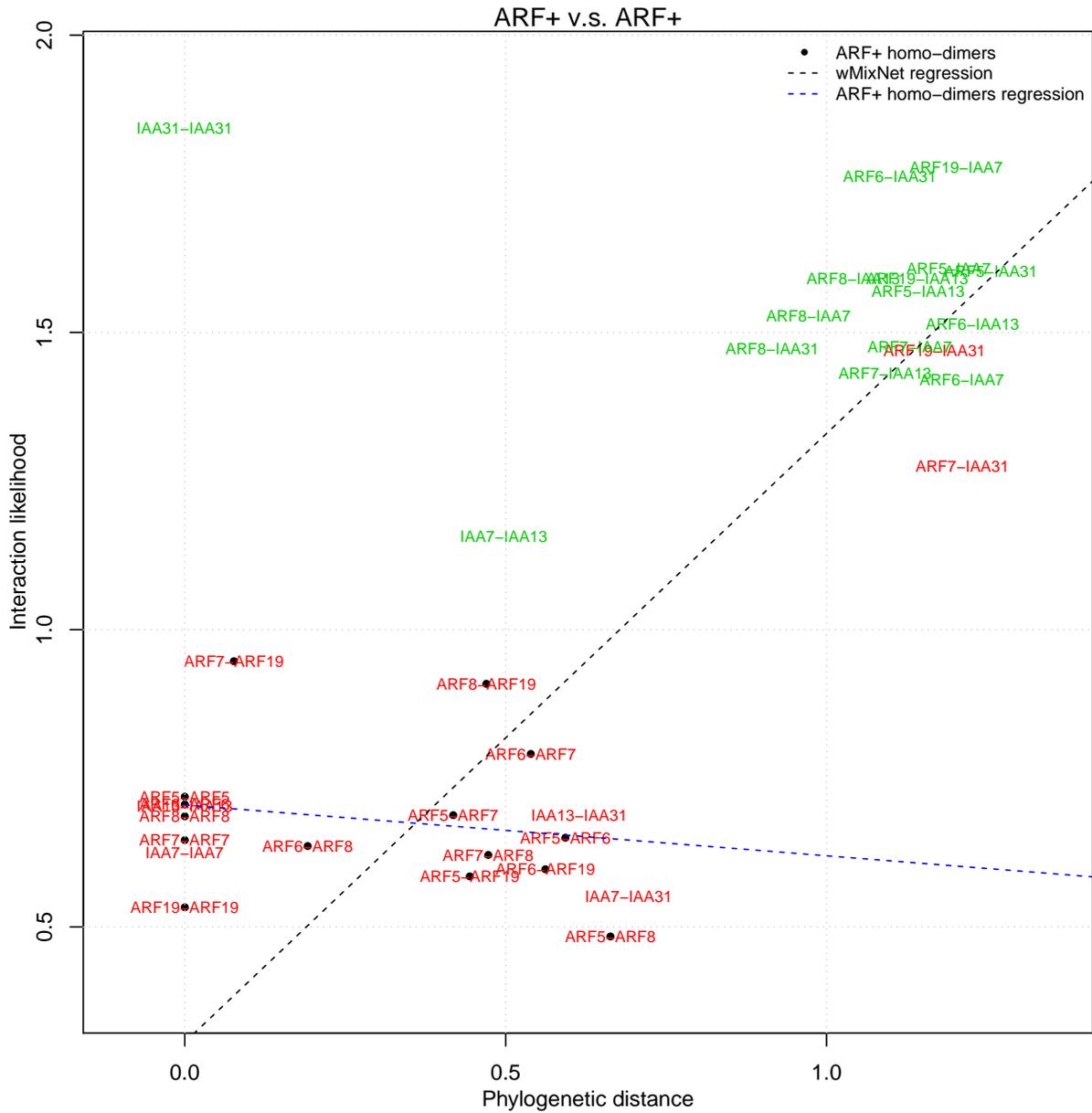


Figure S2.10: Details of the linear regression for network A based on a 4-clusters model with one explanatory variable for the interaction of clusters ARF+ v.s. ARF+.

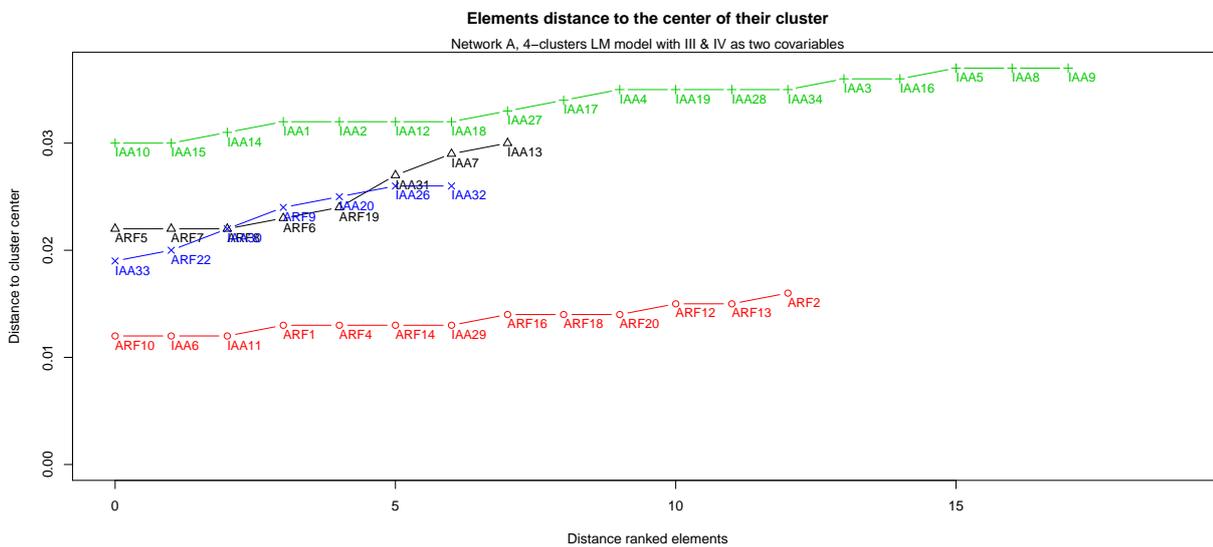


Figure S2.11: Ranked vertex distances to assigned cluster center for network A based on a LM model with two explanatory variables.

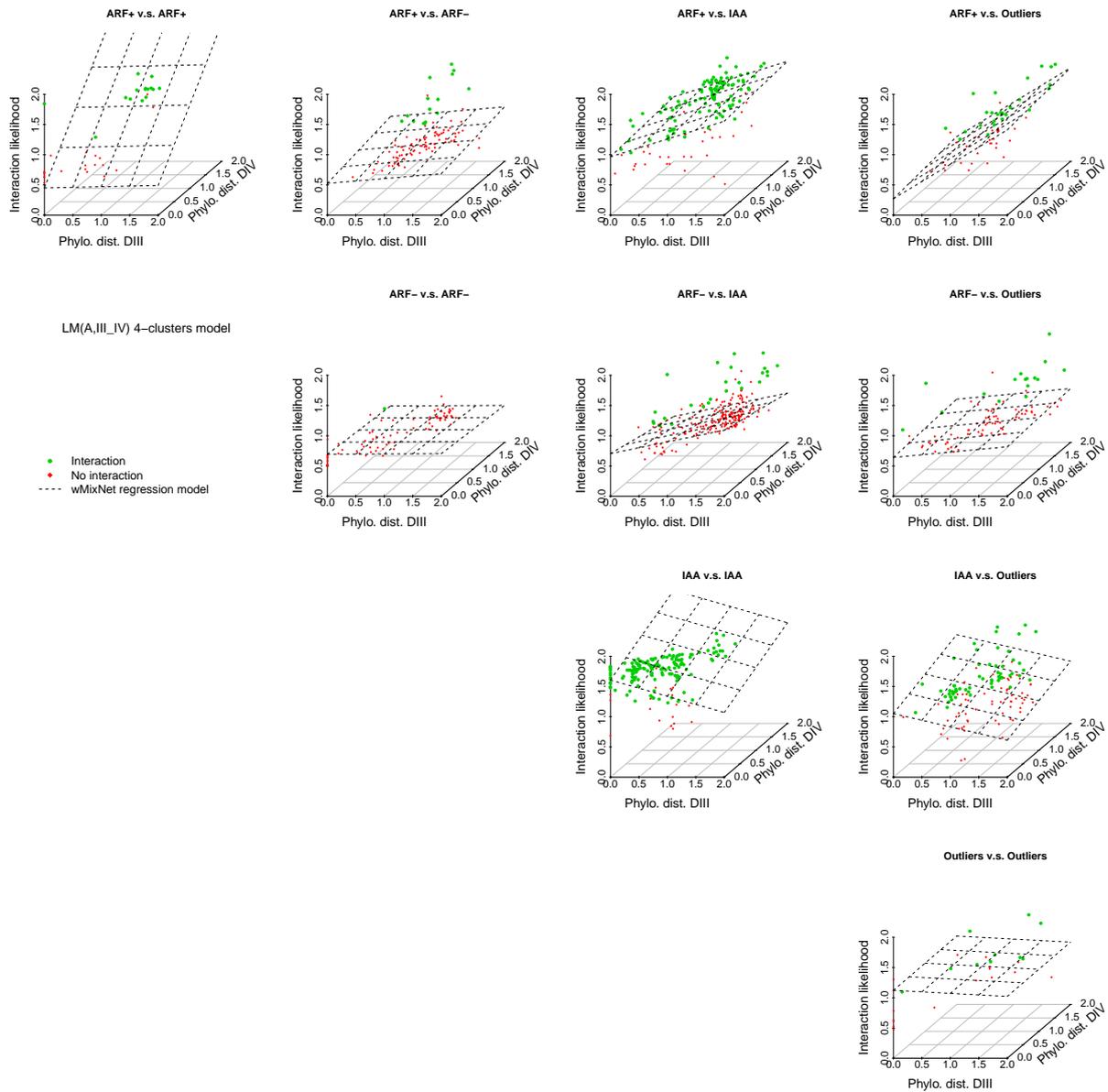


Figure S2.12: Linear regressions for network A based on a 4-clusters model with one explanatory variable for the different pair of groups: ARF+ v.s. ARF+, Aux/IAA v.s. Aux/IAA and ARF+ v.s. Aux/IAA. The outliers cluster is also presented in the graphic.

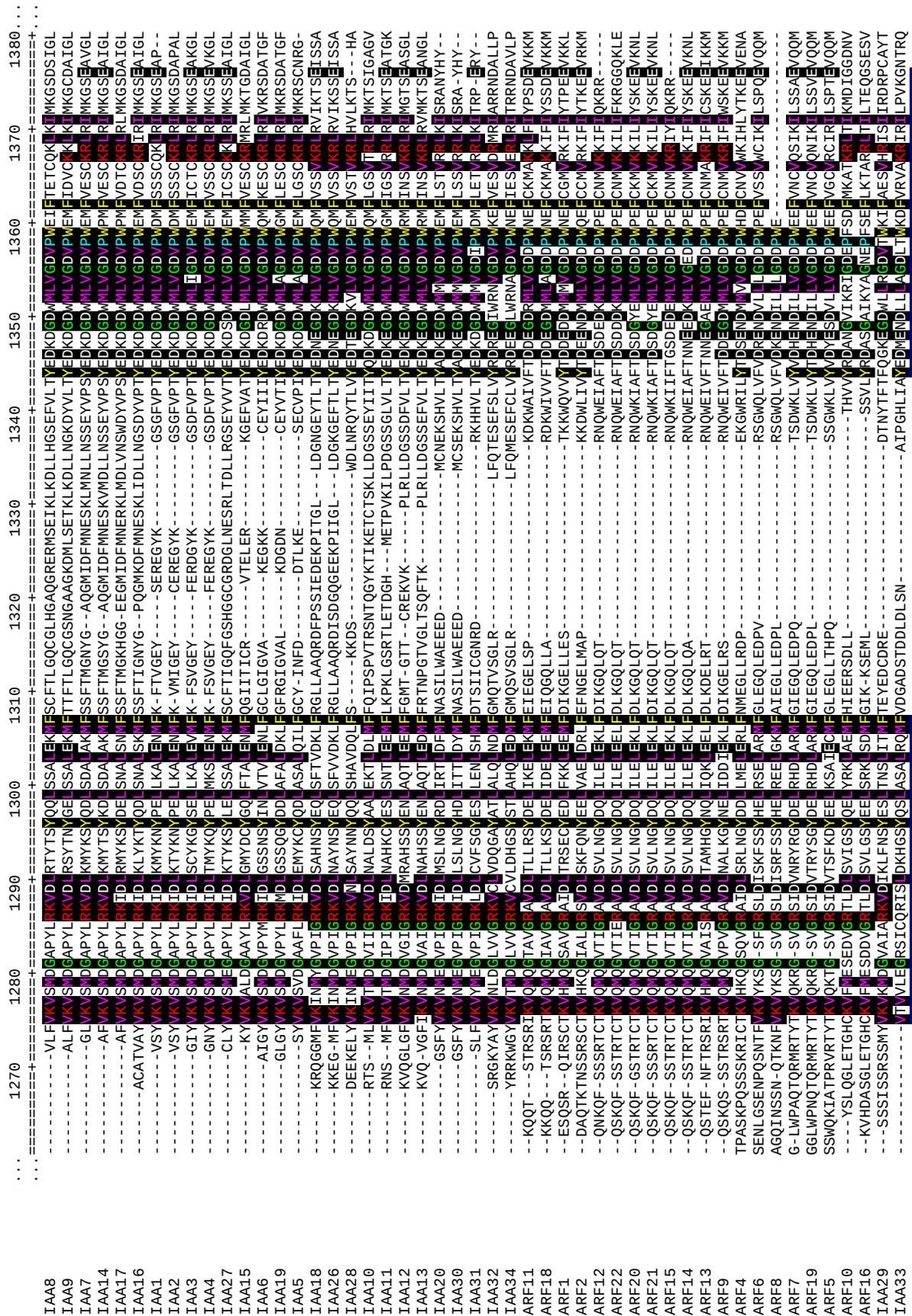


Figure S2.13: Detail of the multiple alignment curated by Glocks showing the domains III and IV. Every proteic sequences of the 49 AGI-accessible Aux/IAA and ARF have been used. The blue boxes indicate the conserved sequences found by Gblocks. Coloured letter represent highly conserved amino-acids. The dashes (-) represent gaps created by the multiple-alignment method (CLUSTAL-W) to align conserved segments of the sequence.

Chapter 3

An attributed spatio-temporal graph framework for characterizing floral morphogenetic patterns in *Arabidopsis thaliana*

Contents

Abstract	132
1 Introduction	133
1.1 Background and requirements	135
1.2 Relation to previous works	137
1.3 Principal contribution of our methods	139
2 Methods	139
2.1 Spatio-temporal graph provides a natural framework for 3D+t data.	140
2.2 Quantifying cellular features of a 3D segmented tissue and enabling graph analysis.	143
2.3 Adaptating the clustering framework to spatio-temporal property data.	162
3 Results	172
3.1 Exploratory analysis of the temporal property graph	172
3.2 Cellular pattern identification on the basis of geometrical features clustering	180
3.3 Cellular pattern identification on the basis of Markov tree analysis of the lineage and geometrical features	192
4 Discussion and perspectives	198
Author contributions	205
Supplementary Materials	206

Abstract

A major challenge in biology is to understand how multicellular tissues can give rise to complex animal or plant architecture. In order to do that we need to explain how the known genetics molecular patterns influence, and are influenced by, the cellular and tissular patterns. It is therefore crucial to be able to quantify and explain these geometrical patterns taking places during morphogenesis. To unravel the complexity of organ morphogenesis, several recent improvements in imaging techniques enabled access to live 3D images (time series), with a resolution sufficient to extract cell features, even for the inner parts of the tissues.

We will address this issue by showing that through a precise quantification of cell features, together with an open-source python based temporal property graph architecture and analysis model, we can identify groups of cells, not only based on their genes differentiation states, but on the outcome of the underlying mechanical, genetical and hormonal events. We will illustrate this methodology during the early stages of *Arabidopsis thaliana* flower development.

1 Introduction

Developmental biology is the study of the processes by which multi-cellular organisms grow and develop. Organogenesis and morphogenesis, meaning the initiation and the shaping of an organ or a tissue, are therefore the two main processes studied. Coordinated and differential growth of a multi-cellular tissue implies active and passive processes to organise morphogenesis to properly shape an organ.

Many experimental protocols have been built by biologist to access (qualitative or quantitative) information about these processes at different scales. For example, quantification of gene expression levels at tissular scale is possible using the microarray technology coupled with dedicated statistical analyses [Schena *et al.*, 1995]. Detection at cellular scale of specific RNA transcripts is possible using *in situ* hybridisation techniques. AFM techniques can measure physical behaviour of the epidermis by quantifying its stiffness at a sub-cellular scale (for review see Kasas *et al.* [1997]).

Recent breakthroughs in microscopy have enabled the live-imaging of developing tissues, providing access to the inner parts of tissues (3D) through time (3D+t). Recently, several contributions to 3D segmentation from membrane-stained live-images (stacks) [Yoo *et al.*, 2002; Fernandez *et al.*, 2010] have enabled the 3D reconstruction of tissues at cellular resolution. Additionally, automatic determination of the lineage of cells between consecutive images has been made possible by the automatic lineage tracking (ALT) algorithm described in Fernandez *et al.* [2010]. It is particularly useful for inner parts of volumetric structures which are difficult to access visually.

A major challenge is to relate the observed gene and hormone activities to the observed tissue shape. It is in Erickson [1976], that was first propose the use of four spatio-temporal parameters to describe growth properties of a 2D tissue: growth rate, anisotropy, direction and rotation. It generalizes to nine for a 3D tissue, but using a combination of the same four parameters [Coen *et al.*, 2004]. Using the somatic clonal technique with appropriate markers, Rolland-Lagan *et al.* [2003]; Coen *et al.* [2004] measured changes of growth and shape in *A. majus* sepals at a supra-cellular level thanks to cloning technique, and incorporated these data in a mechanical growing framework. They then related shape changes to the proposed action of a long-range signal that controls growth directionality, rather than ro local differences in growth rate [Rolland-Lagan *et al.*, 2003; Coen *et al.*, 2004]. Successfully reproducing *in silico* the observed shapes of the petal lobe of *A. majus* is clearly a major advance and demonstrate the necessary use of modelling to study morphogenesis. However, they relied on rather abstract regions (non uniform patches of cells).

To achieve a deeper understanding of morphogenesis, we argue that the whole structure of the object under study should be considered at a cellular level. Then the comparison of quantified spatio-temporal data with those obtained from the growing model will help improving our knowledge about morphogenesis. This requires several steps, starting with the precise quantification of the cellular features in 3D, made possible by segmentation and lineaging of live-images. Several papers describing segmentation algorithms are available (for review), and so far one has introduced automated reconstruction of the lineage [Fernandez *et al.*, 2010]. We will briefly discuss some of them in the next section, but we here propose to accomplish the next step (toward a better understanding of morphogenesis) by proposing a generic framework empowering the analysis of such data. In that view we will introduce the *attributed spatio-temporal graph* representation of the tissue. We believe it is the natural representation of the tissue, since it extracts and relates (spatial and temporal edges) its natural constituents: the cells (vertices). The computable features are of five different orders: enumeration, dimension, position, shape and growth. Attributing them to their related representation (e.g. volume to vertex, wall area to spatial edge and growth to temporal edge) will then complete the necessary step of data processing prior to analysis.

Finally, we propose to identify cellular patterns on the basis of the previously quantified spatio-temporal features using a clustering approach. A cellular pattern will therefore be a group of cells displaying a closely related behaviour regarding defined features. We will introduce a dedicated method to build cell/vertex pairwise distance matrix (used by clustering algorithm), enabling the use of several features at once to define groups of cells. Overall, to use geometrical features to identify cellular patterns and relate them to known gene expression domains would then be an important step toward a direct link between genes and shape.

To illustrate our approach, we applied it on cell-segmented and lineaged data of the early stages of *A. thaliana* flower development. It is indeed a unique feature of plants to be able to produce organs after embryogenesis compared to animals (with the exception of those presenting a larval stage). Entering a reproduction phase, plants develop a new type of organs, the flowers, which we then can observe its early morphogenetic processes. Developing flowers are called floral meristem (FM).

1.1 Background and requirements

By continuously maintaining stem cells at the shoot apical meristem (SAM) and the root apical meristem (RAM), plants possess the ability to initiate and produce different organs throughout their whole life. The meristematic zone is self-maintained and can modulate its developmental activity in response to external cues, providing the necessary flexibility to modulate the aerial plant shape and growth in reaction to environmental conditions. When generating flower, the SAM is also called inflorescence meristem (IM), and can be named interchangeably; see Figure 3.1a.

A. thaliana flowers are made of four distinctive organ types organised in concentric whorls: sepals, petals, stamens and carpels. In addition, flowers display a great robustness in their overall shape, and organ numbers; see Figure 3.1b.

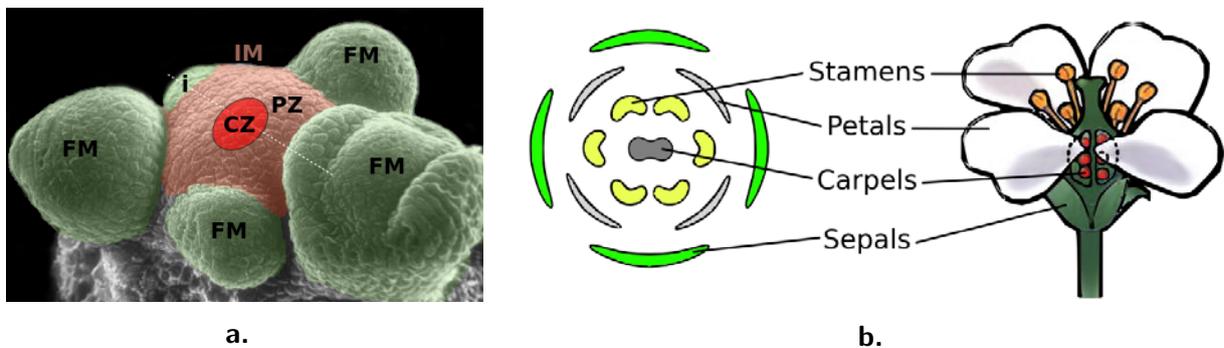


Figure 3.1: *A. thaliana* early and mature flower structure. **a.** IM structural organisations, CZ refers to the Central Zone (organising center) and PZ to the Peripheral Zone; *i* indicate the initium and FM the developing floral meristems. **b.** Whorled organisation of the flower made of 4 sepals, 4 petals, 6 stamens and a gynoecium made of 2 carpels.

1.1.1 Morphological description of the early FM development

The first description of flower development as a series of morphological events was carried out in 1990 by Smyth *et al.*. This started from the initiation of the floral meristem as a small bulge (being a stage 1) on the flank of the IM, and went up to the finale mature flower (being a stage 12) characterising each morphological change as a new stage. Stage 2 starts when the floral bud is separated from the IM by a small crease between the two meristems. Stage 3 is characterized by the emergence of the sepals from the sides of the floral meristem, growing to overlies the primordium. These are therefore the three first morphological events displayed by the young floral primordia during its development. Figure 3.2 presents the first three stages we are interested in.

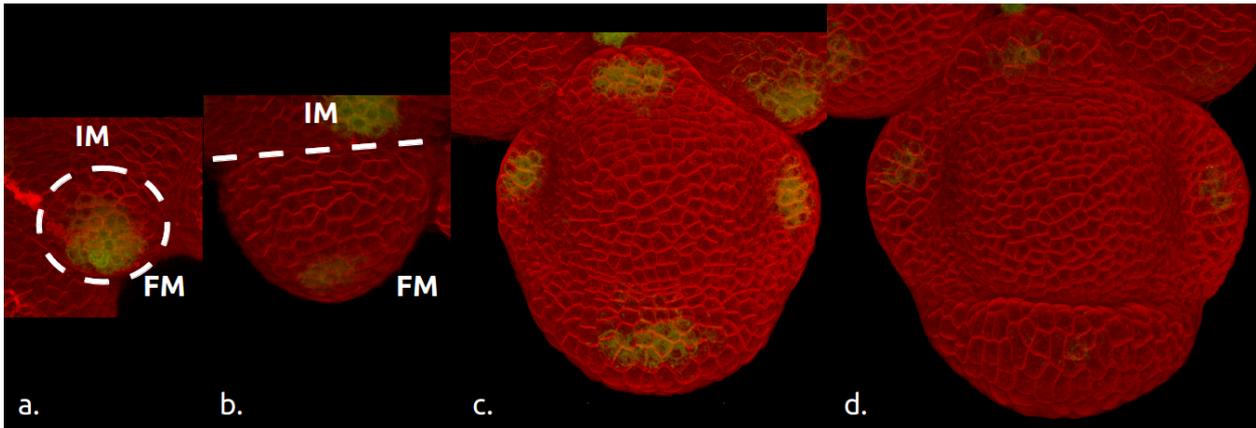


Figure 3.2: The three first stages of *A. th.* FM morphogenesis. **a.** Stage 1, no clear distinction between the FM and the IM; **b.** Stage 2, a clear boundary is visible between the IM and the FM (+26h); **c.** Early stage 3, sepals start emerging at the sides of the FM (+69h); **d.** Stage 3, the sepals are clearly visible (+80h). Vital dye FM4-64 used for wall staining is in red. Green fluorescent marker: *AHP6::erGFP*. Source: Images acquired at the RDP lab by Pradeep DAS.

1.1.2 Quick overview of segmentation and lineage tools.

Several 3D cell-based segmentation algorithms have been developed in the past ten years. Unfortunately most of them are algorithmic libraries like ITK [Yoo *et al.*, 2002] which are not readily usable for non expert programmer.

Some softwares propose semi-automated segmentation methods where you have to manually define the edges of the cells (often by detecting the cells vertices) or using Delaunay triangulations based on the cell's nucleus to recreate their contours. The latter method is obviously biased since it assume that the nucleus is at the center of the cell. Manual detection of the cell edges is accurate but extremely time-consuming, and limited to 2D or surfacic 3D.

In the OpenAlea [Dufour-Kowalski *et al.*, 2007] integrated version of MARS (multi-angle reconstruction and segmentation) [Fernandez *et al.*, 2010], it is possible to combine multiple views of the same object to obtain a 3D stack with the same resolution in the three directions (e.g. from $0.2 \times 0.2 \times 1\mu m$ to $0.2 \times 0.2 \times 0.2\mu m$). This is a unique feature, developed to solve the problem of the lower resolution in z -direction (orthogonal to the focal plane). Combining several images, even with a low tilt angle, is then sufficient to enhance the signal quality of the image used for segmentation. This reconstruction step is semi-automated since manual definition of landmarks is necessary to initiate the rigid and non-linear transformations. However the segmentation is fully automated, although you might have to play with some of the parameters to achieve a reasonable segmentation (seed detection and watershed are sensitive to noise and missing information). All these functions are called through a python command-line terminal.

Another popular software, MorphographX [Kierzkowski *et al.*, 2012] propose a full GUI and the possibility to rapidly edit the segmentation using various correction tools. Its success is also based on the large choice of segmentation algorithms through integration of the ITK library. However only the epidermal layer is accessible since it propose a cell based 3D surfacic reconstruction of the tissue and there is no automatic lineage algorithm yet.

Other integrative platforms like Fiji [Schindelin *et al.*, 2012] or BioImageXD [Kankaanpää *et al.*, 2012] also integrate segmentation tools. The segmentation algorithms are often an integration of separately published methods, but sometimes also dedicated algorithms. For biologists, one key feature to choose between these tools and platforms is the presence of a graphical user interface (GUI), but also the variety of the segmentation algorithms proposed and the possibility to visually check and correct the segmentation.

1.1.3 Obtaining the segmented and lineaged data.

To obtain the 3D segmented images of developing flowers we used the MARS software [Fernandez *et al.*, 2010]. However, using a segmentation algorithm based on a 3D watershed, we faced many errors whenever the walls were too shallow or holes in it. Manual segmentation was therefore necessary to suppress fused cells, but this task was tricky since no well-adapted tools were available for this specific task. In addition watershed can sometimes produce cells whose contours are a bit erratic, and to solve this issue we applied a cell shape correction algorithm whose task was to correct the small bumps at the surface. An overlay view of both segmented and fused image confirm us that the algorithm did not change the morphology of the tissue. Finally, the lineage was generated by hand since we could not get the automated lineage tracking (ALT) working. The various difficulties encountered in obtaining reliable data for the inner layers are the reasons why we are presenting this work on the basis of the epidermal layer (L1). We are confident that our first approach with the L1 will provide evidences of the capacity of our open-source python based temporal property graph architecture and analysis, allowing to go forward with the whole developing flower.

1.2 Relation to previous works

As morphogenesis is an crucial biological process, several biologists, together with physicists, computer scientists and mathematicians have already taken interest in gaining a better understanding

of this process. Several approaches have been developed to tackle the problem in parts or entire organisms as diverse as *Drosophila melanogaster* (fruitfly), *Danio rerio* (zebrafish), *Caenorhabditis elegans* (nematode), *Antirrhinum majus* (snapdragon) or *Arabidopsis thaliana*.

It should be stressed that growth quantification has a long history since measuring overall shape changes, or computing deformations on the basis of natural or manual landmarks (e.g. obtained by addition of a dye), have been attempted for many years. However, cell growth quantification is only accessible since the development of confocal microscopy, first allowing access to in-depth live-imaging of the tissue, when before only external layers could be observed without destroying the object under study [Megason and Fraser, 2007; Oates *et al.*, 2009].

As mentioned before, Rolland-Lagan *et al.* [2003] have successfully reproduced *A. majus* flower lobe development in an *in silico* growing framework. Combining genes network, mechanical deformations and supra-cellular acquired data in their framework, they could test for various growth scenario and concluded to upon the action of a long range signal, rather than local differences in growth rate for being responsible of shape changes. This major advance in morphogenesis illustrate the necessary use of modelling in biology, allowing to use computational models to predict behaviours or test hypothesis. However, we can regret they did not achieve single-cell resolution, and also note that their technique (clonal analysis) is not applicable to all tissues and organisms.

Around the same period, the work of Dumais and Kwiatkowska [2002] has enabled the extraction of surfacic 3D data for the *Anagallis arvensis* SAM. Using successive stereoscopic reconstructions of moulded SAM, they could reconstruct a 3D surface of the SAMs, and then calculate 2D parameters like epidermis cell wall area and curvature. Then using temporal differentiation functions they were able to extract growth features (scalars) and using cell vertices as landmarks they were able to extract deformation features (tensors). Using their analysis pipeline, they could compare a wild-type SAM with a *pin-formed1* (*pin1*) mutant -which does not produce lateral organs such as flowers- and observed a faster surfacic expansion for *pin1*. In addition, they uncovered the presence of regions with distinct deformation rates (strain) in both wild-type and *pin1*, in the latter case corresponding to the geometrically distinguished central and peripheral zones (fig. 3.1a). In addition to these works, many other researches on animal development have shed light on the necessity to quantify growth and morphogenesis.

We have therefore based our approach on these studies, trying to go further doing several contributions. First we propose a generic (open-source) attributed spatio-temporal framework

to analyse tissues from voxel-based segmented images, instead of *ad hoc* framework. Then we introduce 3D+t quantification of the cell features, when 2D+t is yet the commonly used data in morphogenesis studies. Finally, we propose dedicated statistical methods to analyse the obtained 3D+t data, particularly those corresponding to the concept of pattern identification.

1.3 Principal contribution of our methods

Here we present several contributions to the field of quantitative developmental biology, notably by enabling cellular feature and growth event quantification at cellular scale. Such spatio-temporal data being far to complex to be left to intuitive interpretation, we will introduce adapted statistical methodology necessary to their analysis.

The main contribution of our methods are:

- Introduction of the *spatio-temporal graph*, an intuitive structured representation of the developing tissue, allowing us to access the cells spatial and temporal relationships;
- Extraction of many *3D and 3D+t cellular features* from segmented and lineaged images leading to an attributed spatio-temporal graph;
- Introduction of several statistical methods enabling the *mathematical identification and characterisation of biologically relevant cellular patterns* based on the quantified spatio-temporal features (complementary to the classic genetical identity).

The whole framework is open-source and have been developed under python programming language within the OpenAlea platform. It is not bound to any specific segmentation algorithm although it is based on voxelized images, thus meshed images and others would require a transformation script. Finally, the nature of the tissue should be a regular -compact- cell arrangement (e.g. unlike neural tissues).

2 Methods

We will now present the many possibilities our framework offers to identify and characterise cellular patterns from the rich and complex spatio-temporal data that arise from developing tissues reconstructed at cellular level.

2.1 Spatio-temporal graph provides a natural framework for 3D+t data.

As explained earlier, the data consists of segmented and lineaged (at cellular level) time-series of images of early stage floral meristems (figure 3.3). These 3D+t data, observable in most multi-cellular tissues, are highly complex and present structure and dependencies in space and time. In this context, it is important to first develop an abstracted model of the tissue that would describe and organise the natural constituents of a multi-cellular tissue. Such a structure model would be considered interesting if efficiently representing the *in silico*.

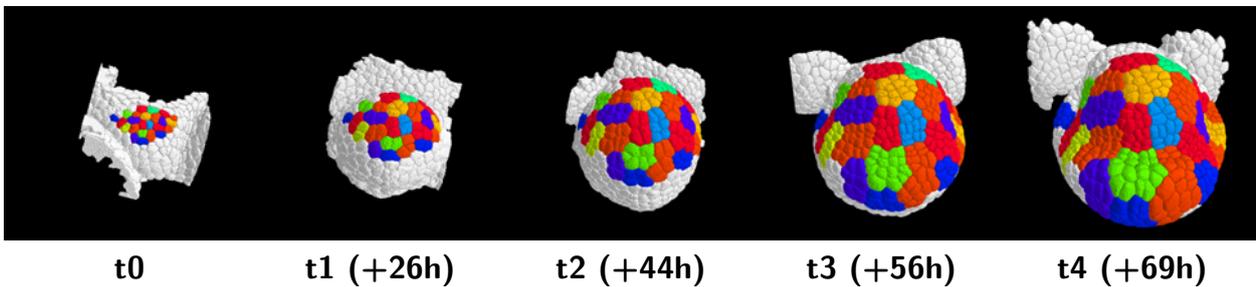
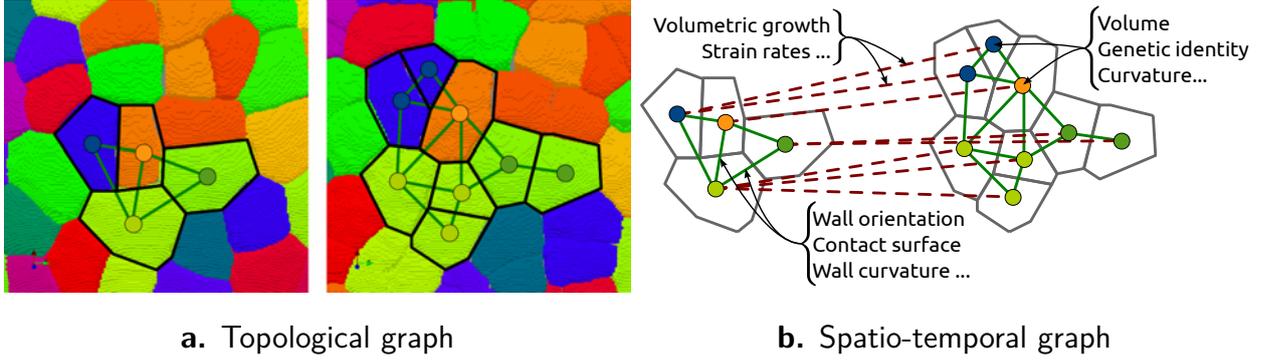


Figure 3.3: Example of segmented and lineaged time-series. Each color giving the lineage from a mother cell at the initial time point passed on to all its daughter up until the last segmented and lineaged time point. Only cells with a complete lineage from the initial to the last time point have been coloured.

Due to acquisition as well as computational limitations (regarding the maximal size of the object), we restricted our analysis of the FM development to the 3 first morphological stages introduced in figure 3.2. As explained earlier, this early stages displays morphological changes. Therefore, the acquired data set, introduced in its segmented and lineaged form in figure 3.3, will be use to develop, test and illustrate our ideas.

Structuring a segmented tissue refers to defining the spatial and temporal topology of its constituents. This can be achieved in two steps: first by decomposing the object into its natural constituents (cells), then by determining how these constituents (vertex) are connected between each others in both space and time (edges).

Spatial graph. As an input, we have a set of N segmented images S_i with $i \in [0, N]$ that gives information at cellular level. Having first abstracted cells as vertices, we then characterise their spatial relationships (neighborhood) by adding adjacency relations as undirected edges, therefore obtaining a spatial graph. A spatial edge will thus be added between pairs of cells if they are physically in contact.



a. Topological graph **b. Spatio-temporal graph**
Figure 3.4: Illustration of the spatio-temporal graph creation from segmented and lineaged images. **a.** Topological graph extraction from segmented images; Coloured circles represent the vertices; Plain green links are the topological relations between cells. **b.** Spatio-temporal graph framework; Dashed red links are temporal relations between cells (lineage); In black we indicate the graph structured that will receive the cell features introduced later.

Such a graph is formalized as $G_{s,i} = (V_i, E_{s,i})$ with V_i a set of vertices representing the different cells at time i , and $E_{s,i}$ the set of undirected spatial edges that link vertices of V_i to represent their spatial relationships at time i .

Temporal graph. The temporal or lineage information present a structured information about the cells organisation in time. This is a directed tree-like representation where each initial parent cell is linked to its descendants.

The tree analogy is applicable because the vertices are only linked by their common root (in our case the unknown initial cell). Such a relation is modelled by a graph $T = (V, E_t)$ that relates each parent vertex V to its descendants by directed edges E_t . It gives a partial ordering relationships between the vertices. Such a graph is called a forest since its made of a set of directed trees.

Spatio-temporal graph. Merging both views, we obtain a model where the cells are abstracted as vertices V and are related by undirected spatial and directed temporal edges. Doing so, we obtain a structured version of the developing tissue, while creating a conveniently organised *in silico* database.

This can be formalized as a graph $G = (V_i, E_{s,i}, \pi, \phi)$ with $E_{s,i}$ representing the spatial relationship between vertices V_i at time i , π the temporal relationship between vertices and ϕ is defined as a surjective mapping $\phi : V \rightarrow [t_0, \dots, t_N]$ that map vertices to their original time points. Note that temporal constraints can be formalized on such structure. In particular, if two cells have spatial relation, we can check that they have a common parent or their distinct parents have also a spatial

relation:

$$\forall e_i = (v_i, v_j) \in E \Leftrightarrow \begin{cases} \exists e_k = (\pi^{-1}(v_i), \pi^{-1}(v_j)) \in E; \\ \pi^{-1}(v_i) = \pi^{-1}(v_j); \end{cases} \quad (3.1)$$

Such constraints helps us checking the integrity of the structure.

Spatio-temporal graph with attributes. After defining the spatio-temporal relations between the cells, the obtained spatio-temporal graph can be used to store basic spatial cell features such as their volume (example are given in table 3.1). Attributing the basic features to their related vertex will then keep the positional and geometrical information about the cells. As described later in 2.2.2 and 2.2.4, it is possible to compute other features using differentiation and transformation functions applied on these basic information. Spatial edges can also receive related information like contact surfaces areas, for example, or transporters directionality. Temporal edges will account for temporally differentiated features (described in 2.2.3).

Thus certain temporal information (e.g. division or siblings numbers) and spatial relations (e.g. ranked distance or neighbourhood) can then be inferred on the fly from the graph structure itself and does not need to be saved explicitly.

However, it is important to note that the lineage can be incomplete. Indeed, it is subject to missing data, for example when descendants cells partially or fully escape the acquisition frame so some t_n cell have no descendants at t_{n+1} . This is an experimental problem and it should not be mistaken for cell apoptosis (which has not been detected yet during flower organogenesis). It is also possible that the acquisition windows has been extended at t_n compared to t_{n-1} and some lineage will not present root at the initial time-point t_0 . Again, it shall not be confused with “appearance” of a cell file.

Finally, it should be noted that, due to acquisition time-intervals, the division process may not be binary (i.e. two descendants per parents). Multiple divisions of a parent cell can indeed occur between successive acquisitions. To solve this and obtain a full binary tree we manually annotated the lineage to reconstruct a sub-lineage corresponding to the binary process of division. Since only a maximum of four descendant cells (two parallel or three division occurrences) were found in the data, we were able to recover every binary division event. This reconstruction of the L1 cells binary tree has been achieved manually using epidermis wall areas, cell volumes and positional informations (topology and division wall orientation). We hypothesized that the division plane of a cell is orthogonal to its main axis of inertia and the resulting division yields two daughters of

roughly the same volume. In case of the Markov tree approach, where the only structure used is the set of temporal relationships between vertices (forest), it was a necessary step.

2.2 Quantifying cellular features of a 3D segmented tissue and enabling graph analysis.

It is noteworthy that the geometrical descriptors we use have been previously introduced in the literature. However, the methods and formulations we use sometimes differs from those used in Dumais and Kwiatkowska [2002], Coen *et al.* [2004] and others since we are working in 3D on voxel-based data. We could have first transformed our data in a meshed structure like used in Dumais and Kwiatkowska [2002], however, meshing a discrete spatial representation comes with a whole suite of problems by itself.

2.2.1 Cellular features data

The basic cellular features that are accessible from static, voxel-based segmented images are essentially the following enumeration, positional, dimensional, and shape descriptors:

1. enumeration, such as cell numbers, are scalar data usually self-contained within the spatio-temporal graph;
2. Dimension descriptors, such as cell volumes, are also scalars;
3. Position descriptors, such as the cell barycenters, are 3D vectors;
4. Shape descriptors, such as the inertia tensor, are rank-3.

In table 3.1 I list two examples per type of variable.

Enumeration	Dimension	Position	Shape descriptor (tensors)
Cells	Area	Barycenters	Inertia axis
Neighbours	Volume	Landmarks	Principal curvatures and normal (wall)

Table 3.1: Examples of basic spatial information extractable from segmented image.

Basics enumeration descriptors

Number of cells on the whole image is easily defined as the number of unique labels (ids) attributed to each cell by the segmentation algorithm.

Number of neighbours is defined as the number of cells sharing a wall with a reference cell. It can be filtered according to a minimum contact area.

Basics dimensional descriptors

Area is computed in a specific fashion due to the voxel-based (discrete) nature of the images. Biologically, we can make a difference between three types of surface:

- *Contact areas*: there are as many contact areas as cells in contact with the reference cell;
- *Epidermis area*: can be considered as a particular type of contact surface since it is defined for the outer layer of cells (also called the 'L1') that are in contact with the surrounding environment;
- *Cell area*: the sum of all contact surface areas.

Volume is easily computed due to the voxel-based (discrete) nature of the images, by counting voxels of each label. To obtain the volume in the metric system (real world units) one just need to multiply the previous count of voxels by the image resolution.

Basics positional descriptors

Center of mass defines the relative position of the cell in the image with respect to the referential of the image. Since we cannot infer the mass distribution within each cell, we will suppose that it is homogeneous, therefore it will be equivalent to the *barycentre*.

We used the function `center_of_mass` available in `scipy.ndimage.measurements` from the `scipy` library [Oliphant, 2007].

Landmarks are a set of reference points, suchs as cells vertices, that can be associated in time. Extracting cell vertex positions in space and associate them through time allows us to compute the deformation undergone by cells while growing in a constrained environment like a tissue.

However, as we will illustrate with examples later, cell vertices are not robust landmarks because of potential segmentation errors. The origin of the limited robustness of cells vertices to segmentation errors comes from their geometrical definition in a 3D space. To better understand this, we give the vertices, edges, walls and cells geometrical definition in 3D. Remembering that a segmented image is a 3D array of voxels whose values correspond to the cell they belong to, in 3D we define:

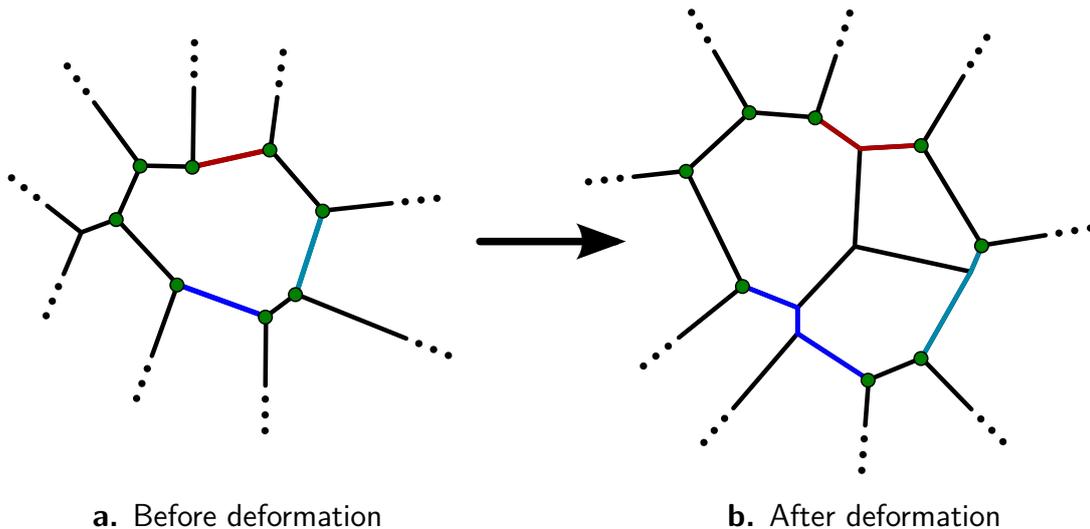


Figure 3.5: 2D example of cell vertices used as landmarks (green points) associated between time points t_n (a) -before- and t_{n+1} (b) -after observed growth. The black lines indicates the cell outlines. Coloured lines indicate the surrounding cells that have divided and the resulting new adjacent walls. This representation is a simplification of a real tissue.

- a *cell* as the set of voxels with the same value;
- a *wall* as the set of coordinates where two voxels with different values (i.e two cells) are in contact at a minimum of one voxel face;
- an *edge* as the set of coordinates where three voxels with different values (i.e two walls or three cells) are in contact at a minimum of one voxel edge;
- a *cell vertex* as the coordinate where four voxels with different values (i.e three edges or four cells) are in contact at their common vertex;

It is to note that segmentation algorithm are not precise at the scale of individual voxels. Indeed, some voxels may have different values in alternative segmentations obtained with slightly different segmentation parameters (seeds, h-min, ...) in particular at the frontiers of the cells. A change of value of one voxel (from cell A to B, its neighbour) will not greatly alter the definition of the cell (usually made up of thousands of voxels) or of the wall (made of hundreds of voxels), but will have a slight effect on the associated edge and an even greater effect on the cell vertex position.

Due to this sensitivity to segmentation imprecision, it is difficult to automatically define reliable landmarks using cell vertices on a time series of segmentations. Indeed various topological incoherences such as :

- Changes of topology occur when for instance (see figure 3.6) two cells, A and D, are at t_n in contact with B and C (which are themselves not in contact), whereas at t_{n+1} that relation is inverted (i.e. B and C are now in contact with A and B, themselves not in contact any more).

Plants have rigid walls that forbid cell rearrangement in space and therefore constraint the topology to be conservative. Using figure 3.6 as an example, vertices i, j are defined as $i : \{A, B, D\}$ and $j : \{A, C, D\}$. Without topological errors we would have automatically associated them with $i' : \{A', B', D'\}$ and $j' : \{A', C', D'\}$, however we have $k : \{A', B', C'\}$ and $l : \{B', C', D'\}$;

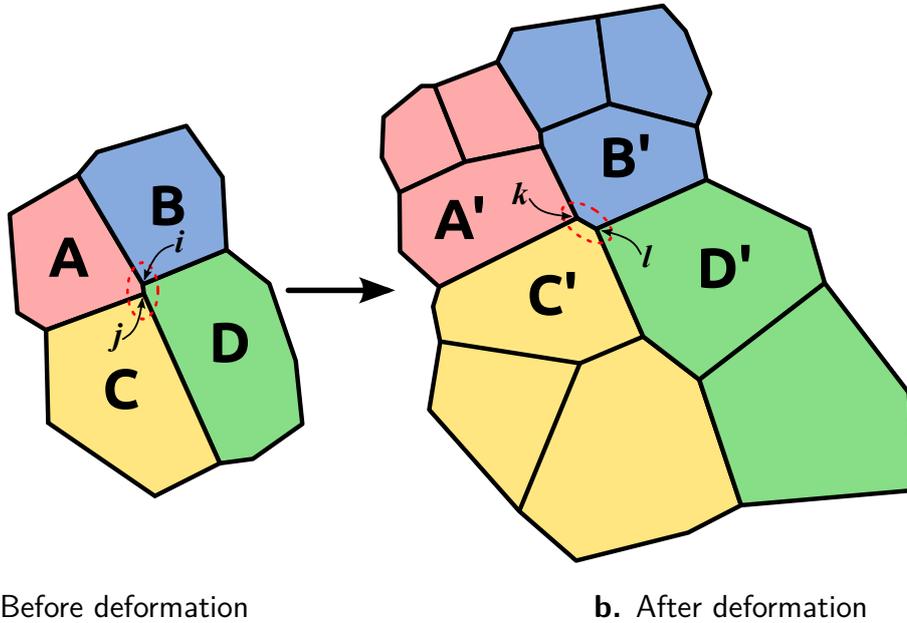


Figure 3.6: 2D example of segmentation error leading to impossibility to associate landmarks at time $t_n: (i, j)$ with those at $t_{n+1}: (k, l)$. Coloured cells indicate their lineage. The black lines indicates the cell outlines. This representation is a simplification of a real tissue.

- Segmentation presenting “holes”, in our case in the epidermis layer (L1), can cause the failure of vertex detection (fig. 3.7 white arrows) if the hole is close to supposed vertices;
- “Fused” vertices are another case where the epidermis might present a specific problem when two separate cells vertices are too close. This is a specific to the external side of the L1. An example is shown in figure 3.8, where four cells have a common vertex. Counting the outside world makes five jointed 3D shape defining one vertex, instead of four as presented earlier, then breaking the imposed geometrical rules;

Instead of developing specific methods to correct detected errors, we thought of using the cell wall medians in 3D, as well as the edge geometrical median in 2D as presented in figure 3.9, for the quantification of cell deformation. As mentioned above, these geometrical reference points fulfil the pre-requisite of a landmark (notably coherent positioning through time) and less susceptible to segmentation errors.

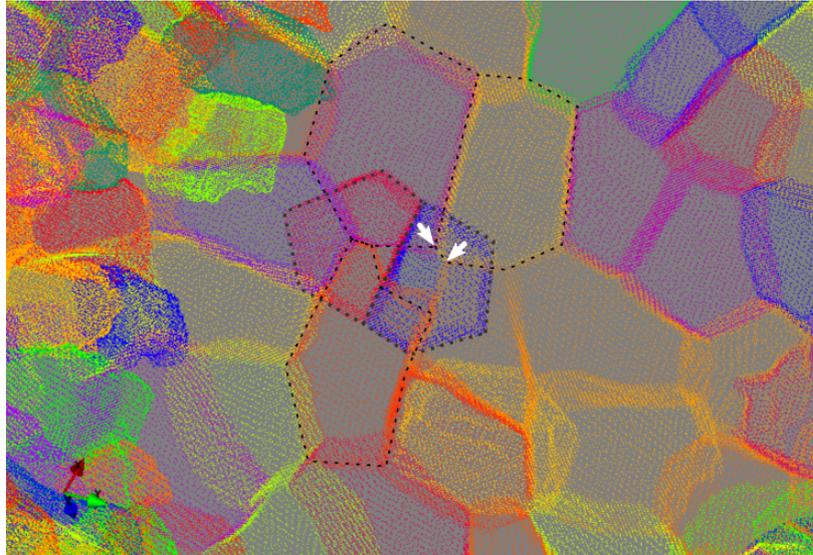


Figure 3.7: Example of a segmentation error leading to the indeterminacy of vertices. Red and blue cells in the center of the image (gray dots) are below the epidermal layer. Since the orange cell present a missing portion we cannot detect the vertices marked by white arrows. This representation is a real tissue where only wall associated voxels are presented as coloured points indicating cells.

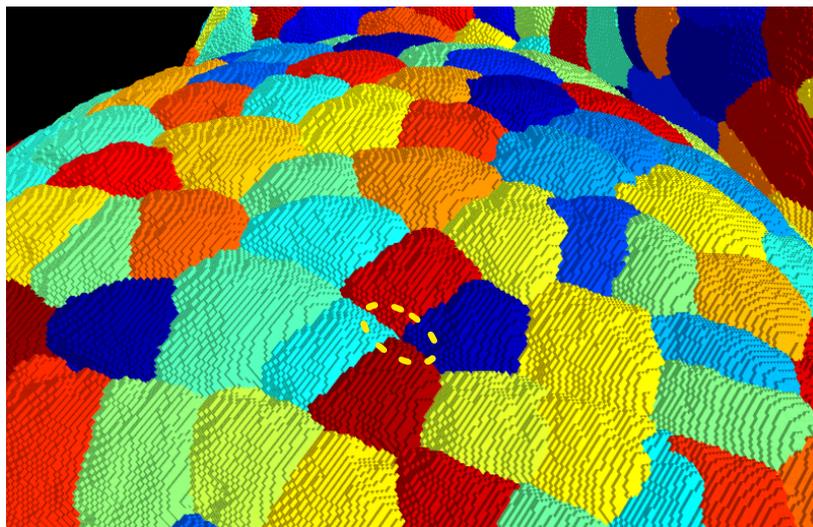
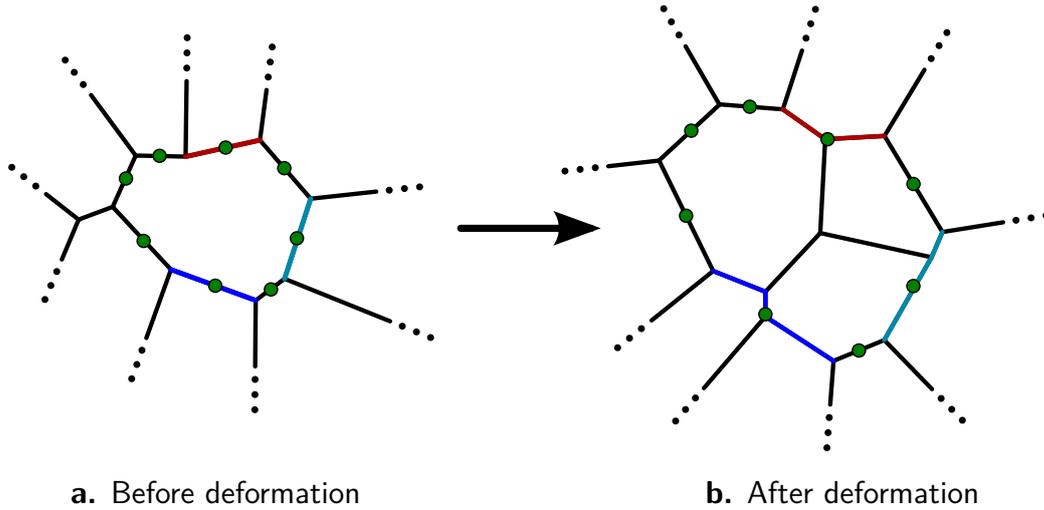


Figure 3.8: Example of a segmentation error leading to the indeterminacy of two vertices by presenting only one. The indeterminacy is shown in the yellow dotted ellipse. This representation is a real tissue where only wall associated voxels are presented as coloured cubes indicating cells.

Topological distance define the distance in terms of walls separating two cells. It is computed according to a shortest path methods since there is several “paths” joining two cells by moving between their neighbours. We have implemented a version of *Dijkstra’s algorithm*. It is a graph search algorithm that solves the single-source shortest path problem for a graph with non-negative edge path costs, producing a shortest path tree.

Euclidean distance is a metric distance here defined between two cell barycenters in 3D or geometric median of their epidermis wall for surfacic 3D. It can be computed by using the norm



a. Before deformation

b. After deformation

Figure 3.9: 2D example of cell wall median used as landmarks (green points) associated between time points t_n (a) -before- and t_{n+1} (b) -after observed growth. The black lines indicates the cell outlines. Coloured lines indicate the surrounding cells that have divided and the resulting new adjacent walls. This representation is a simplification of a real tissue.

between their two barycentres (centres of mass) or geometric medians. The Euclidean distance between two points p and q in \mathbb{R}^3 is the length of the line segment connecting them (\overline{pq}). In \mathbb{R}^3 with Cartesian coordinates, p and q are commonly defined as $p = (x_p, y_p, z_p)$ and $q = (x_q, y_q, z_q)$. Then the distance from p to q , or from q to p is given by:

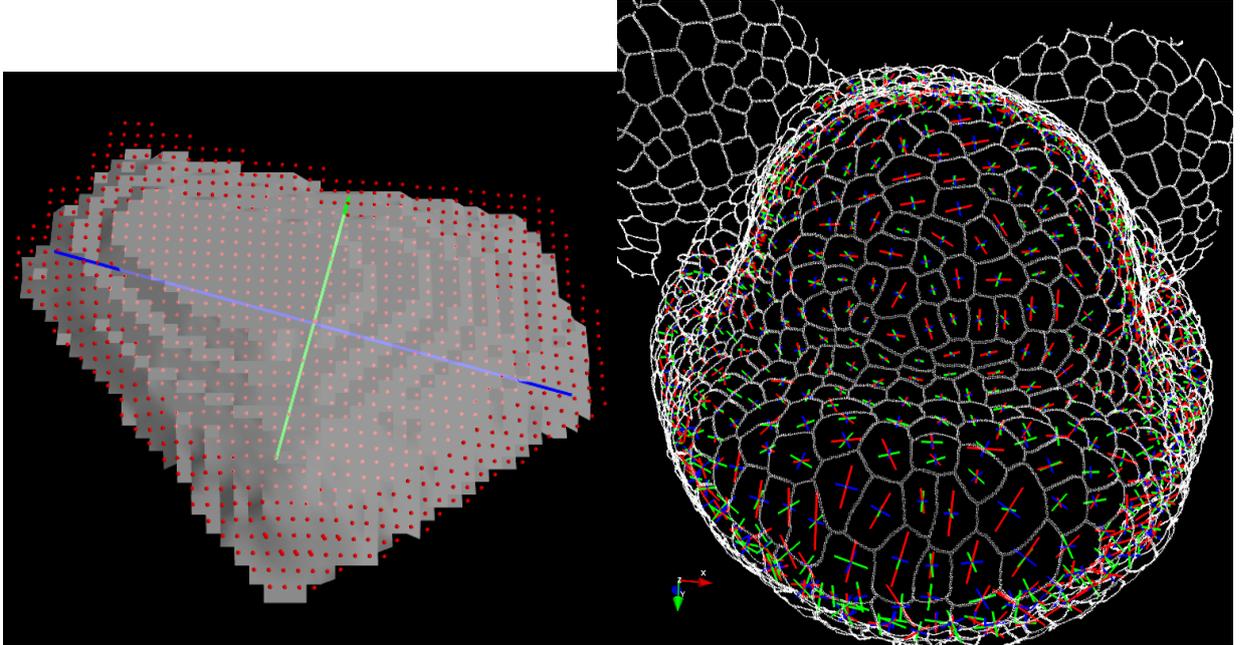
$$d(p, q) = d(q, p) = \sqrt{(x_q - x_p)^2 + (y_q - y_p)^2 + (z_q - z_p)^2}$$

The topological distance is a metric distance between two cells. It can be computed by using the norm between their two barycentres (centres of mass) or geometric medians.

Basics shape descriptors I fully introduce two shape descriptors: the inertia tensors (fig. 3.10a) and principal curvatures (fig. 3.11a). These two tensors are also given for the whole object at t_4 for temporally linked vertices, in figures 3.10b for inertia tensor and 3.11b for the principal curvature estimated on a local radius of 70 voxels around the geometric median point of the epidermis wall (origin of curvature estimation).

Inertia tensors have their origin at the barycentre of the cell and the length and direction of the axis describe the shape in three directions. To compute this tensor we assumed that the density of the cell is homogeneous. The main axis of inertia can be viewed as the axis that would present the maximum resistance to increasing the rotation speed of a body about that axis. Assuming the

body density is homogeneous, this corresponds roughly to the axis presenting the maximal length of a convex body. A detailed 2D illustration on one cell wall is given in figure 3.10a, with the main axis in blue and secondary in green. The epidermis wall in gray is first projected in a 2D subspace (flattened into a plane).



a. Detailed example of inertia axis (2D)

b. Cell inertia axis at t_4 (3D)

Figure 3.10: Illustration of the inertia tensor shape descriptors: detailed illustration of a 2D inertia tensor for a cell wall (a) and the representation of computed 3D inertia tensors at t_4 (b). **a.** Representation of the 2D inertia tensor with the main and secondary axes in blue and green respectively, obtained using the rank-2 subspace projection (plane defined by the red dots) of the 3D cell wall (gray). **b.** Inertia tensor with decreasing sorted vectors in red, green and blue; the tensors are centred on the barycentre of the cells. The epidermis cells outlines is given in white.

Inertia tensor computation use a *Singular Value Decomposition* of the variance-covariance matrix. This assume that the cell density is homogeneous, but since we have no information about cell density, we will admit this approximation.

We have a voxel based segmented image and therefore, each cell is made of a set of N points X_i in \mathbb{R}^3 . The variance-covariance matrix is given by:

$$C = \frac{1}{N} \bar{X} \bar{X}^T,$$

where \bar{X} is the the barycenter of the X_i . Applying an SVD on C yields:

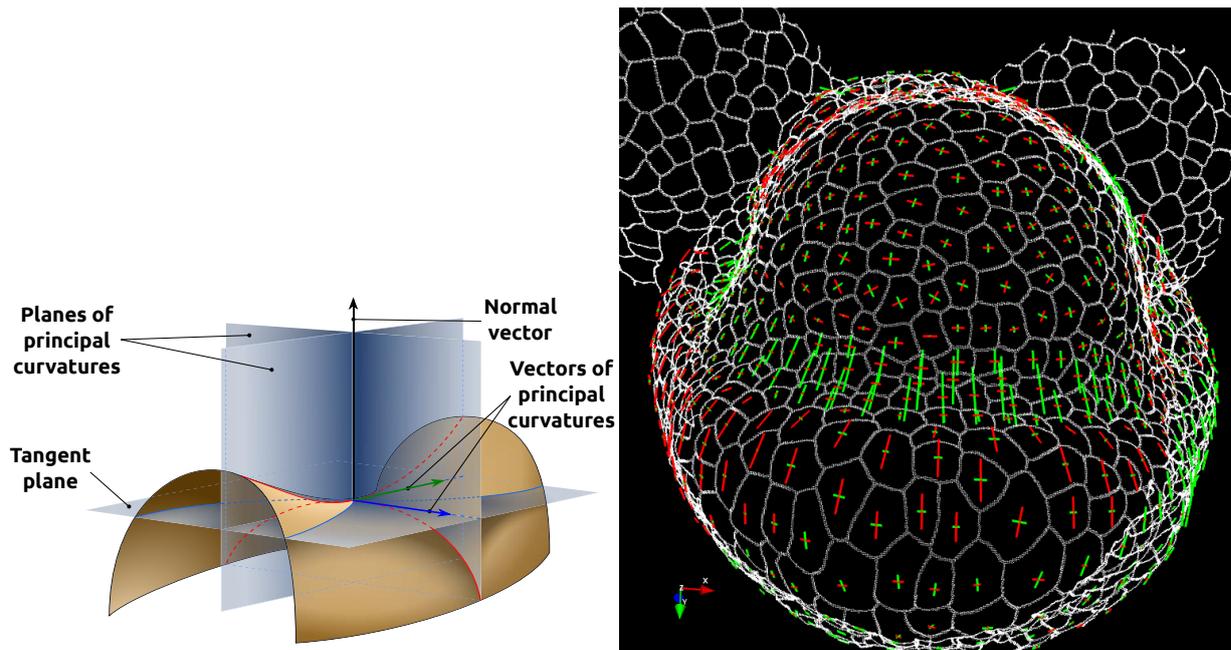
$$C = U \Lambda V^T,$$

where the diagonal terms of Λ are the eigenvalues and U and V the left and right eigenvector respectively.

Principal curvatures describe the directions and extent of how a surface is bent at a given point using the geometrical median of the wall. At this point, the normal vector to the surface (fig. 3.11a black arrow) define the tangent plane in which the vectors of principal curvature will be defined. The curvature is said to be positive (fig. 3.11a blue arrow) if the curve turns in the same direction as the surface’s chosen normal, and otherwise negative (fig. 3.11a green arrow). In addition to the point were to realise the curvature estimation, a sampling surface also have to be provided. In case of an internal wall, providing all the voxel positions defining the wall, thus the estimated surface will be “wall-defined”. However, for the epidermis surface, it is possible to extend the sampling surface used to do the estimation since epidermis walls are contiguous. Then it is possible to define a regular sampling surface by selecting all epidermis voxels within a given distance from the estimation origin point. A good approximation would be to imagine a circle fitted to the surface, centred on the origin point. We believe the regular aspect of the sampling surface will provide a less biased curvature estimator than the “wall-defined” sampling, especially for cells with a highly anisotropic shape.

Principal curvatures estimation has been possible using the CGAL library [Fabri and Pion, 2009] in which there is an appropriate method for principal curvature estimation. The idea is to fit a degree-two curved plane to the set of point defining a wall. Then it is possible to extract the main axis of curvature, defined by the direction presenting the maximum curvature and a secondary curvature value, which direction is orthogonal to the main axis. We then obtain two principal curvature directions and their associated values k_1 and k_2 with $k_1 > k_2$. CGAL also return the normal orientation to the surface.

To select which points to use for the estimation, several possibilities arise depending on the type of estimation desired (local or cell wall based) and were we are in the object (epidermal layer or inner calls). The first obvious option for this estimation, is to use all voxels defining a cell wall. This make the estimation doable for all walls, but the “sample” of voxel used will therefore change depending on the area they define. This potentially affected the quality of the estimation depending on the size of the wall. The quality of the estimation will then also depend on the image



a. Detailed example of principal curvatures **b.** Epidermis principal curvatures at t_4 (2D)

Figure 3.11: Detailed illustration of principal curvature descriptors: the principal curvatures of a saddled surface (a). **a.** Saddle illustrating the principle curvature estimation; the origin point where to do the estimation must be provided; the red lines indicate the direction of maximal curvature in both directions, defining therefore a rank-3 tensor made of the two vectors of principal curvature (blue and green arrows) and the normal vector (black arrow). **b.** Epidermis curvature computed for a radius of 70 voxels from the epidermis wall median (origin of the vectors); the normal to the surface is not represented here for readability reasons; the epidermis cells outlines are given in white in both figures.

resolution, which directly defines the number of voxels per wall. The quality of the segmentation, that depends on many parameters (raw image quality, segmentation algorithm, ...), will also be crucial.

In the particular case of the epidermis, it is possible to extend the estimation area by selecting the voxels belonging to neighbours at rank-1 or more. However, we think it could be a biased way of doing the estimation since the selected area will depend on the areas of the cell and its selected neighbours.

In order to overcome this problem we selected the voxels at a maximal distance of 70 voxels from the geometric median (see section 2.3.8) of the epidermis wall as presented in figure 3.12. With our resolution of $0.2 \times 0.2 \times 0.2 \mu\text{m} \cdot \text{voxel}^{-1}$, this define a circle of $616 \mu\text{m}^2$. The observed mean epidermal cell area is around $70 \mu\text{m}^2$ and in average a cell is surrounded by 8 neighbours, therefore the rank-1 neighbours define roughly an area around $560 \mu\text{m}^2$

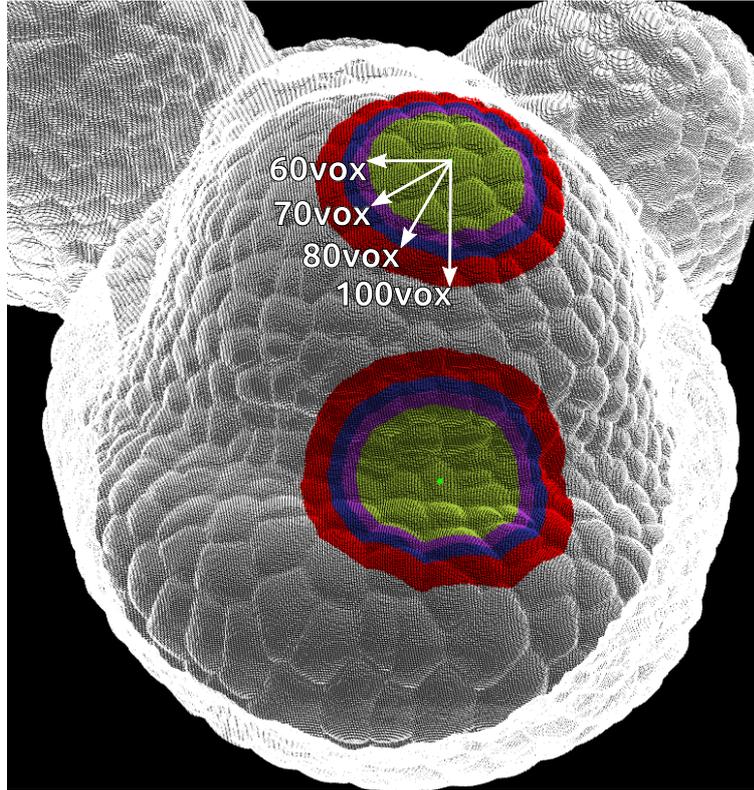


Figure 3.12: Illustration of four different local sampling radius around the geometric median (green cube) of two epidermal cell walls. The yellow, magenta, blue and red discs are voxels selected at a distance of 60, 70, 80 and 100 voxels respectively.

Wall orientation. We used the CGAL library for curvature estimation with a degree 0 fitted plane. It then return the plane orientation axis and the normal to the plane, therefore defining the plane orientation.

2.2.2 Spatial differentiation functions

Scalar spatial features can be spatially differentiated through spatial filters like *laplacian* (eq. 3.2) or *mean absolute deviation* (eq. 3.3). Their interpretation is usually straightforward and will give an idea of local variations in cell features, usually given a topological rank distance of 1 (direct neighbours), though rank- n computation is also possible (see supplementary figure S3.6). Several other functions can be implemented and designed, but the more complex their formulation is, the more complex their interpretation becomes. For sake of simplicity, we chose the following two definitions.

For a cell c surrounded by N neighbours, its *laplacian* (L) and *mean absolute deviation* (MAD) for any scalar feature x are defined as:

$$L(c) = x_c - \frac{1}{N} \sum_{n=1}^N x_n, \quad (3.2)$$

$$MAD(c) = \frac{1}{N} \sum_{n=1}^N |x_c - x_n|, \quad (3.3)$$

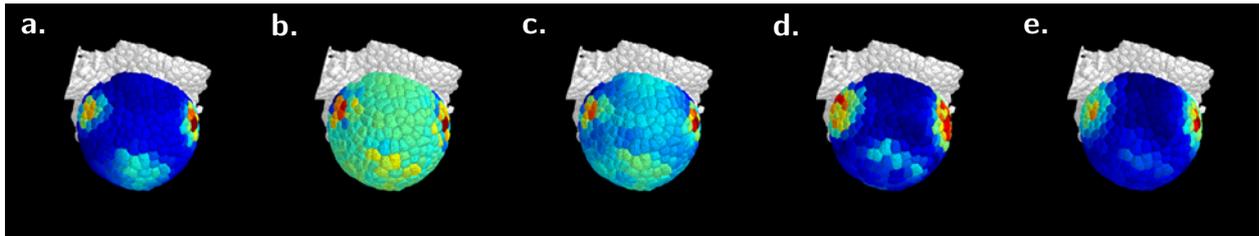


Figure 3.13: Examples of spatial differentiation functions *Laplacian* and *mean absolute deviation* applied on AHP6 signal (genetic marker) for rank-1 and 2 (topological distance) at t2 (+44h). Top view projection of **a.** AHP6 signal feature; **b.** rank-1 laplacian; **c.** rank-2 laplacian; **d.** rank-1 mean absolute deviation; **e.** rank-2 mean absolute deviation. The colormaps of each illustration (a. to e.) are scale ranging their max and min values over the whole time series.

Adapting such differentiation functions on a graph $G = (V, E)$ with vertices V and edges E , lets consider a function $\phi: V \rightarrow R$ associating a value to each vertices of the graph. The discrete Laplacian Δ acting on ϕ at rank 1, is then defined as:

$$(\Delta\phi)(v) = \sum_{w: d(w,v)=1} [\phi(w) - \phi(v)] \quad (3.4)$$

where $d(w, v) = 1$ is a spatial distance of 1 between vertices w and v . Thus the Laplacian sum is taken over the nearest rank- d neighbours of the vertex v , hence we compare it value to their mean. This function is often used in graph theory, simulation or physics to study diffusion processes, as it highlights spatial fluctuations of the variable of interest. As an example, in figure 3.13b we cannot observe any gradient, but in figure 3.13c, using the rank-2 discrete Laplacian, we observe a diffusive decreasing gradient of the AHP6 signal from the abaxial and lateral sides to the top of the dome.

2.2.3 Temporal differentiation functions

Using the known parameters of acquisitions, especially the time intervals between each acquisitions, the temporal differentiation of scalar cell features is straightforward and will yield growth-related metrics like volumetric growth if applied on volumetric information. However, it is possible to compute various “growth functions” like *absolute* or *relative temporal changes* and others. the following illustrate four different growth functions and their differences, so the decision of which one to select will be adapted to the context (i.e. what conclusions I want to draw or what analysis I will do).

Absolute and relative temporal change.

For each parent cell c having $j = \{1, \dots, D\}$ descendants, the *temporal change* (3.5) and the

relative temporal change (3.6) are defined as:

$$TC(c) = \frac{1}{\Delta t} \left(\sum_{j=1}^D x_j - x_c \right), \quad (3.5) \quad RTC(c) = \frac{1}{\Delta t} \left(\frac{\sum_{j=1}^D x_j - x_c}{x_c} \right), \quad (3.6)$$

where x_i is the value of cell i for any scalar feature attached to a cell.

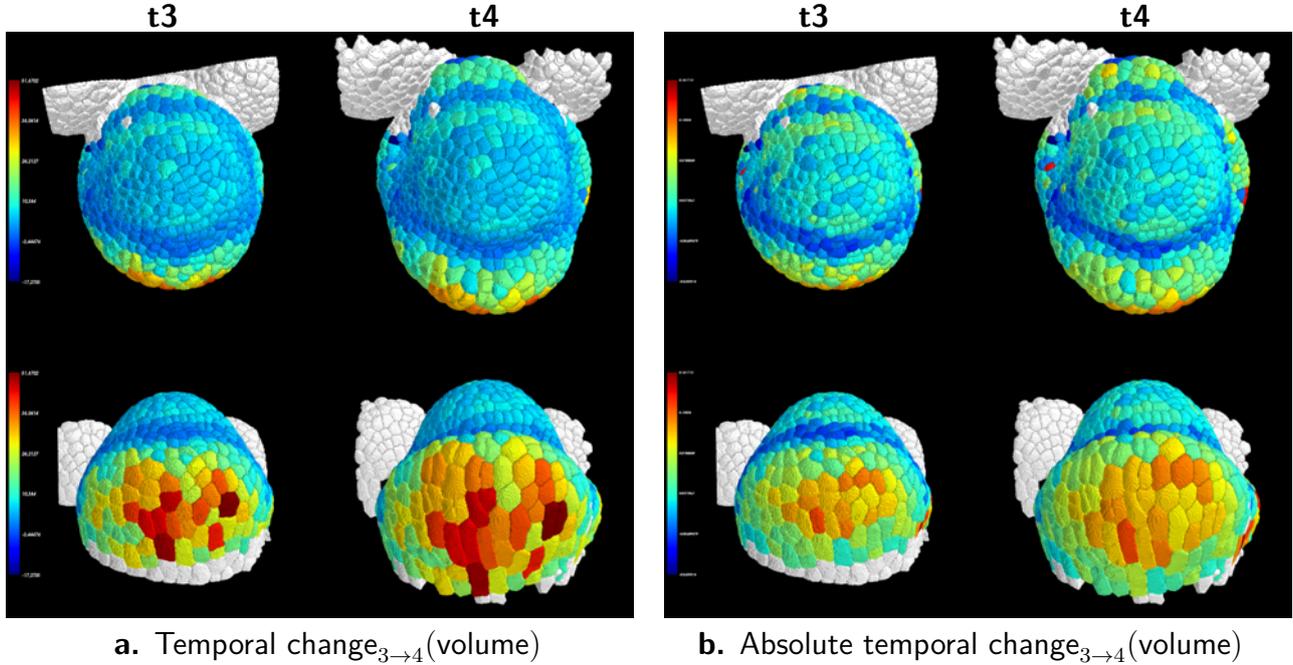


Figure 3.14: Examples of temporal differentiation functions *temporal change* (a) and *relative temporal change* (b) applied on 'volume' between t_3 and t_4 , represented at t_n and t_{n+1} (same data). When using the relative temporal change instead of the temporal change, the extrema of the observed changes (volumetric growth) are highlighted, here showing the boundary between the central dome of the FM and its emerging sepals. The colormaps for (a) and (b) both range between their respective minima and maxima. **a.** Observed *temporal change* in volume. **b.** Observed *relative temporal change* in volume.

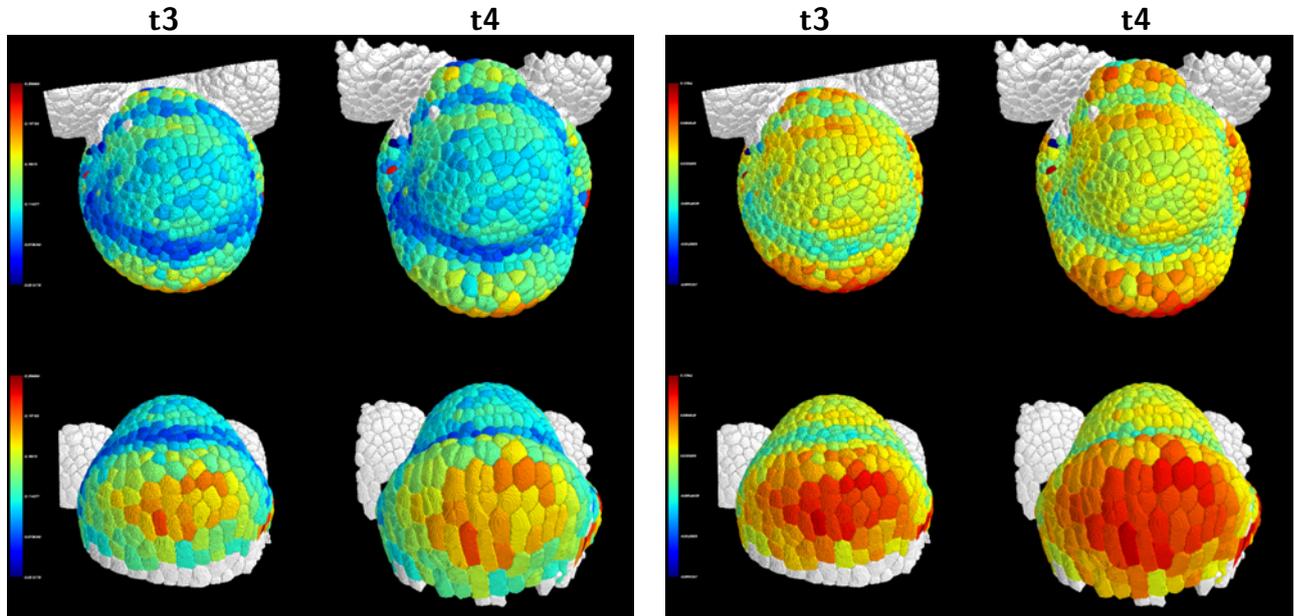
As we can observe in figure 3.14, dividing the difference between the sum of descendants cell volumes and that of the parent by the initial value of the parent allows to suppress the bias of the initial state. Comparing the standardized version of each observed distribution (fig. 3.16a) clearly illustrate the fact that these two function return different information. Dividing by the initial volume using the RTC formulation seems to redistribute the growth related values on a wider range when the TC formulation return value accumulating around the mode.

Temporal rate and log temporal rate.

For each parent cell c having $j = \{1, \dots, D\}$ descendants, the *temporal rate* (3.7) and *log temporal rate* (3.8) are defined as:

$$TR(c) = \frac{1}{\Delta t} \frac{\sum_{j=1}^D x_j}{x_c}, \quad (3.7) \quad LTR(c) = \frac{1}{\Delta t} \ln \left(\frac{\sum_{j=1}^D x_j}{x_c} \right), \quad (3.8)$$

where x_i is the value of cell i for any scalar feature attached to a cell.



a. Temporal rate_{3→4}(volume)

b. Log temporal rate_{3→4}(volume)

Figure 3.15: Examples of temporal differentiation functions *temporal rate* (a) and *log temporal rate* (b) applied on 'volume' between t_3 and t_4 , represented at t_n and t_{n+1} (same data). Using the log ratio impose a symmetry between increase and decrease growth rate for temporally differentiated values. The log temporal rate(b) representation seems less interesting, but only because of the biased repartition of the values around the mean value. The colormaps for (a) and (b) both range between their respective minima and maxima. **a.** Observed *temporal rate* in volume. **b.** Observed *log temporal rate* in volume.

Using the temporal rate, the values associated to growth increase are defined in $]1, +\infty[$, when the decrease is defined in $]0, 1[$. The values are thus defined on a larger range in the case of an increase than for a decrease, which makes their comparison harder. Taking the log of the ratio has the effect of imposing a symmetry between growth increase ($\in]0, +\infty[$) and decrease ($\in]-\infty, 0[$). In the given representation of both function (fig. 3.15), the log temporal rate values projection (fig. 3.15b) seems to return higher growth rate values than the temporal rate values projection (fig. 3.15a), but it is only a matter of the colormap representing the values distribution (scaled by the min and max values of the data).

Figure 3.16b clearly show the symmetry restoration using the log values:

- when using the temporal rate, the standardized observed data are mostly between $[-2, 5]$, with a mode below the mean;
- when using the log temporal change, the standardized observed data are mostly between

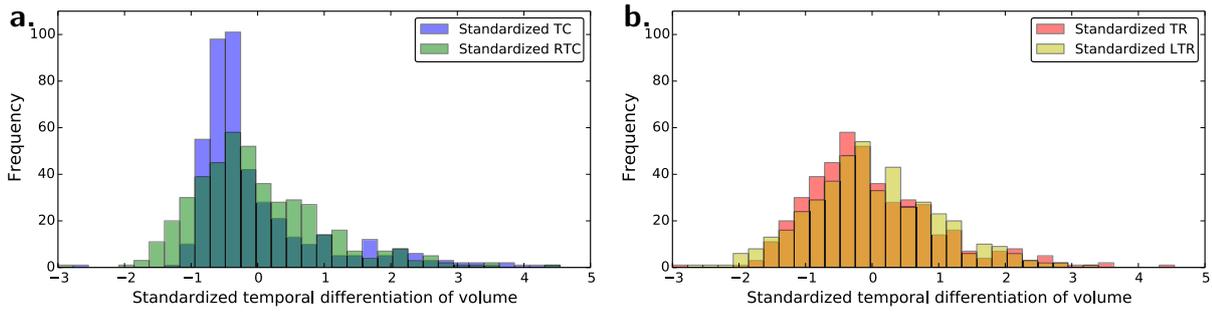


Figure 3.16: Observed standardized distributions of temporally differentiated volume values computed with the given four temporal differentiation functions. **a.** Comparing *temporal change* and *relative temporal change* highlight the redistribution of the values by dividing by the initial values allowing to compare the effective volumetric growth independently for this initial values. **b.** Comparing *temporal rate* and *log temporal rate* show that the log function restore a symmetry in the distribution allowing for a better relation between increase and decrease.

$[-3, 3]$, with a mode close to the mean;

These equations are not fully independent, since they are based on the same variables, and they provide slightly different information. Therefore depending on their mathematical formulation and their properties (symmetry, mode, variance, limits...), the interpretation of the returned values should be adapted.

Division Rate.

For each cell c having $j = \{1, \dots, D\}$ descendants, the division rate is defined by:

$$DR(c) = \frac{D}{\Delta t}$$

2.2.4 Features transformation functions

We hereafter introduce functions applied on features of higher dimension than scalars like vectors of tensors, either to express them as scalar (at cost of loss of information) or to combine them so they express other high dimensional features.

Dimensionality reduction functions transforms tensor shape descriptors into scalar. An example is transforming the inertia axis into a scalar describing shape anisotropy using the orthogonal axis norms. These transformation functions reduce the dimensionality of the space in which the feature is defined, though some information is lost in the process. It is therefore crucial to understand the limits of such reduction functions to choose wisely.

Shape anisotropy scalar descriptor. Describing with one scalar the shape anisotropy is not easy, and there is many physical and mathematical definitions, mostly based on inertia-like properties. Here we propose to use the magnitudes of the inertia vectors to compute a 3D shape anisotropy scalar called *fractional anisotropy* [Basser and Pierpaoli, 1996]. As described before (section 2.2.1) the inertia tensor is obtained by SVD of the variance-covariance matrix.

Using the three eigenvalues $\lambda_1, \lambda_2, \lambda_3$ obtained by SVD, we can extract the *fractional anisotropy* defined in $[0, 1]$ as:

$$\text{FA} = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \hat{\lambda})^2 + (\lambda_2 - \hat{\lambda})^2 + (\lambda_3 - \hat{\lambda})^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}},$$

with the trace $\hat{\lambda} = (\lambda_1 + \lambda_2 + \lambda_3)/3$.

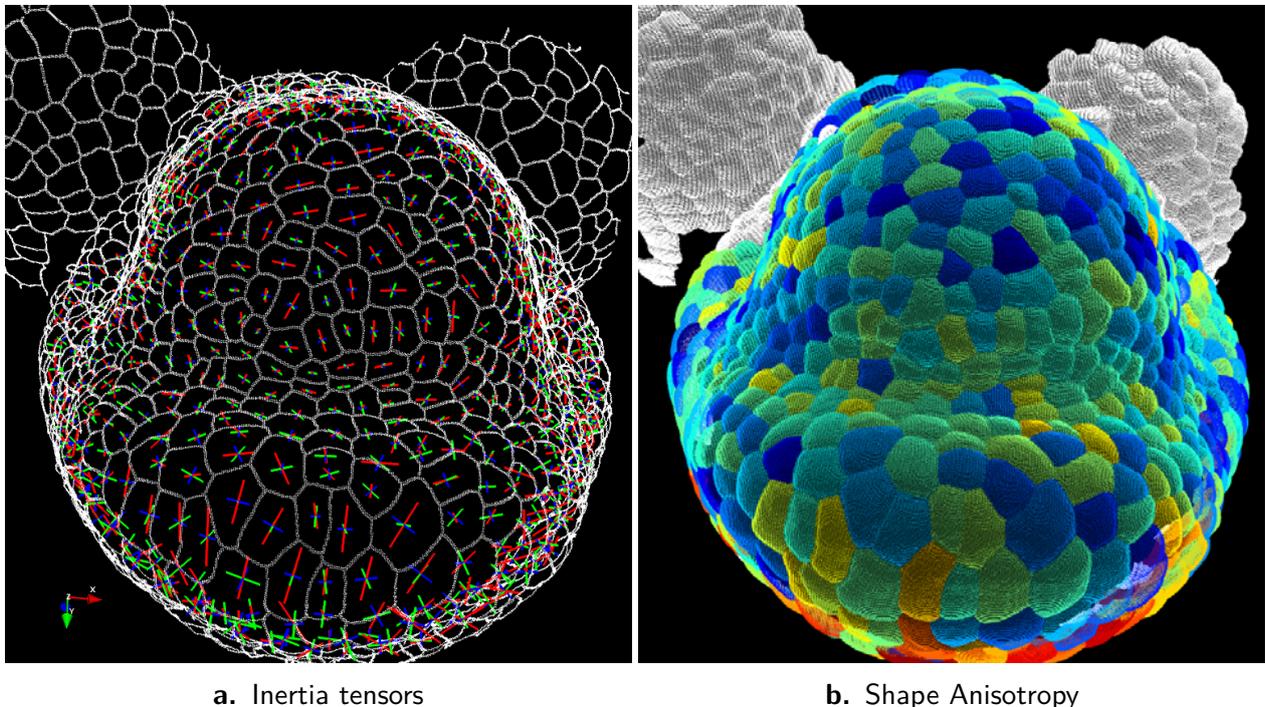


Figure 3.17: Example of transformation of a tensor (3D) shape descriptors into a scalar (1D) shape descriptor for t_4 segmented image. **a.** Inertia tensor with decreasing sorted vectors in red, green and blue; the tensors are centered on the barycenter of the cells. **b.** Fractional anisotropy describing shape anisotropy defined in $[0, 1]$.

Another and probably better solution would be to directly compare the tensors, though this require a common spatial reference system for all observations. Since between successive acquisitions the object under study may move (by itself or to be put back in growing conditions) it is not always possible to define this common reference system. This could be solved using biological landmarks, or computed from registration of all images to define a self reference system. The last solution being as good as the observed deformations between successive time points are small, it is

again not an universal solution. In our case, we have no means yet to do such tensor comparisons since no biological landmark exists (in the acquisition frame) and the deformations are important between successive images.

Curvature estimators are based on the previously defined principals curvature values k_1 and k_2 , it is possible to compute several estimators. Here we give the formula for two of them, the Gaussian and mean curvature.

The *Gaussian curvature* of a cell c is the product of both curvature values:

$$GC(c) = k_1 k_2$$

The *mean curvature* of a cell c is the mean of both curvature values:

$$MC(c) = \frac{k_1 + k_2}{2}$$

Subspace projection. For each cell, we therefore have a set of N landmarks X_i in \mathbb{R}^3 . Finding the best representation of those points on a 2D hyperplane found by least square fitting is equivalent to the *Singular Value Decomposition* of X :

$$X = USV^T,$$

Here U is an $N \times 3$ orthogonal matrix ($U^T U = \mathbb{I}_3$) whose columns u_j are the *left singular vectors*; V is a 3×3 orthogonal matrix ($V^T V = \mathbb{I}_3$) whose rows v_j the *right singular vectors*, and S is a 3×3 diagonal matrix, with diagonal elements $s_{1,1} \geq s_{2,2} \geq s_{3,3} \geq 0$ are the *singular values* (and off-diagonal elements equal to zero).

If we denote X_2 the projection of X onto its rank-2 subspace, then for each point X_i their orthogonal projection on the plane is: $X_2 = V_{:,2} \otimes V_{:,2}^T X_i = H_2 X_i, \forall i \in [1, N]$, where $V_{:,2}$ is the 2 first columns of V ($v_j, j \in [1, 2]$), and \otimes the dot product.

Landmark definitions and transformations are often used to compute deformation and stress associated descriptors.

Below I detail the stretch matrix computation leading to the strain tensor characterisation as

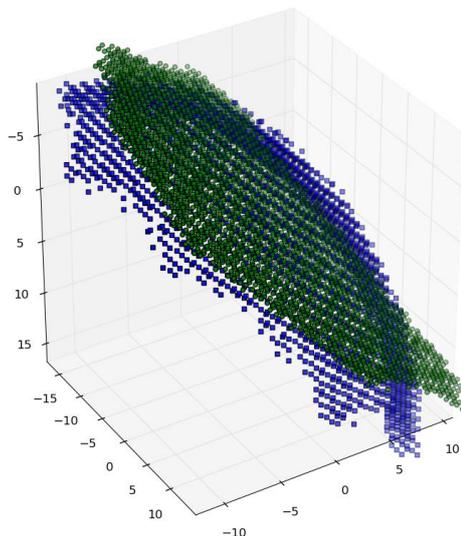


Figure 3.18: Example of a rank-2 projection (green) computed from the set of voxels defining the real initial shape of the wall (blue). In blue are the voxels defining the true wall. In green are the voxels position computed through rank-2 subspace projection of the blue ones. This representation has been generated from rank-2 computation and representation of real wall from our data.

well as the formal definition of the landmarks.

Defining the landmarks as the cell wall medians. For each cell, we therefore have a set of N landmarks X_i in \mathbb{R}^3 computed from the geometric median function (see Methods 2.3.8). We will denote X and Y the $N \times 3$ matrix of coordinates before and after deformation respectively. We also define \bar{X} and \bar{Y} the barycenters of the X_i and Y_i for $i \in [1, N]$.

An important question to characterize cell geometry dynamics is to quantify the deformation that undergo its walls. To compute the deformation matrix of a wall in a rank-2 subspace (to obtain a 2D deformation), we present a general version of the stretch matrix computation [Hastie *et al.*, 2009]. It is notably useful to compute the *areal strain rate* of a wall by getting rid of its potential “bending” due to osmotic pressure, and essentially act as a plane regression (see figure 3.18). For the details about subspace projection see 2.2.4.

Stretch matrix computation. I consider the observed deformation as affine and introduce *least-square regression* for its estimation. With d being the dimensionality (here $d = 2$ or 3), the affine deformation is:

$$\bar{Y}_d = \bar{X}_d A_d + \epsilon_d, \quad (3.9)$$

where A_d is the $N \times d$ matrix of deformation or stretch matrix, and ϵ_d the $N \times d$ matrix of errors. The estimation of the affine transformation matrix A in equation (3.9) and the error matrix ϵ is done using a *least-square estimation* [Hastie *et al.*, 2009].

In figure 3.19 we give an illustration of the method. This enables the computation of the strain tensor (axis and values) as well as the rotation undergone by the cell. This is illustrated in figure 3.19, where after deformation, the initial centered black cross, now in dashed grey, has been affected by a rotation \vec{R} and a dilation d_1 along the main deformation axis and d_2 along an axis orthogonal to \vec{d}_1 .

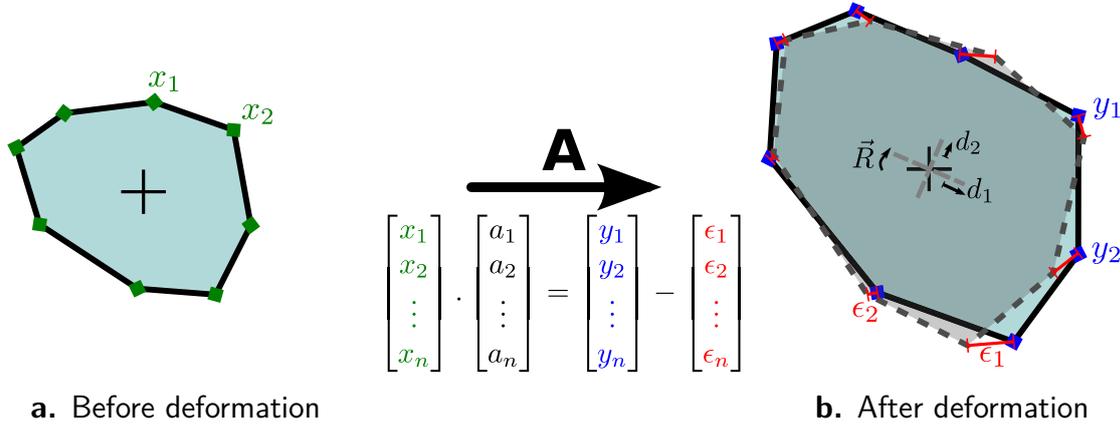


Figure 3.19: Stretch matrix 2D illustration with cell wall medians used as landmarks before deformation at t_n (green) and after at t_{n+1} (blue). The red line illustrate the error made by the affine regression (rotation and non-isotropic dilation) that may not be able to match perfectly t_n and t_{n+1} landmarks (obviously in the case of a non-affine deformation). After deformation, the initial centered black cross, now in dashed grey, has been affected by a rotation \vec{R} and a dilation d_1 along the main deformation axis and d_2 along an orthogonal axis to \vec{d}_1 . The black lines does not indicates the cell outlines, but represent the closed shape. Coloured point indicate the wall medians. This representation is a simplification of a real tissue.

Strain magnitude and directions. The *singular value decomposition* of the previously computed stretch matrix A gives also the main axis and magnitude of the deformation. Writing the SVD of A as:

$$A = RDQ^T,$$

this decomposition expresses the stretch matrix A as the product of three matrices: **two rotation matrices** R and Q^T , and the **dilation matrix** D (with off-diagonal elements equal to zero). Columns of R and Q^T are the strain directions before and after deformation, respectively.

The d_k diagonal entries of D are the stretch ratios, that can be related to the *strain rates* by the following equation [Dumais and Kwiatkowska, 2002]:

$$\sigma_k = \frac{1}{\Delta t} \ln(d_k), k \in [1, 3] \text{ in } \mathbb{R}^3 \quad (3.10)$$

Areal strain rate define the amount of areal dilation (if positive) or contraction (if negative) of a given surface (wall) of the object. It is defined per cell wall and their comparison make sense for the epidermis wall (L1). We define it as the sum of strain rates as defined in equation (3.10) if projected into a 2D subspace (because then $\sigma_3 = 0$):

$$\Sigma_{\mathbb{R}^2} = \sigma_1 + \sigma_2 \quad (3.11)$$

Another direct way of computing the *areal strain rate* is to use the log temporal change formula (eq. 3.8) with the epidermis area:

$$\Sigma_{\mathbb{R}^2} = \frac{1}{\Delta t} \ln \left(\frac{A_{t_{n+1}}}{A_{t_n}} \right), \quad (3.12)$$

where A_{t_n} and $A_{t_{n+1}}$ are the cell area before and after deformation.

To compare the two formulation, we plotted the returned values against each other in figure 3.20. As we can see the two functions are correlated as expected. The values returned by the

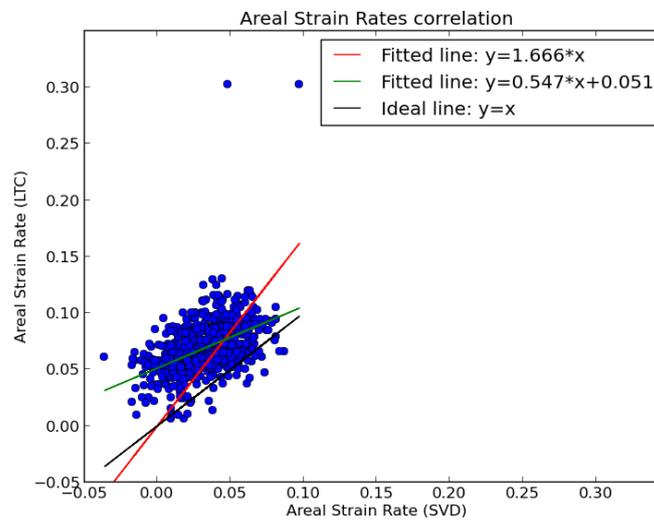


Figure 3.20: Areal strain rate computational methods comparison between the ‘LTC version’ (eq. 3.12) and the ‘SVD version’ (eq. 3.11). The ‘LTC version’ slightly over-estimating the areal strain rate since using non-flattened surfaces of the external wall of epidermal cells.

log temporal change is slightly higher than those from the SVD formulation since they use ‘non-flattened’ surface, therefore slightly over-estimating the areal strain rate. This ‘LTC version’ (eq. 3.12) is however much simpler and faster than the ‘SVD version’ (eq. 3.11).

Volumetric strain rate define the amount of volumetric dilation (if positive) or contraction (if negative) of an object. It is defined for all cells and a “true” 3D strain computation is necessary. We define it as the sum of strain rates defined in equation (3.10):

$$\Sigma_{\mathbb{R}^3} = \sigma_1 + \sigma_2 + \sigma_3$$

Again another direct way of computing the Volumetric Strain Rate is to use the log temporal change formula (eq. 3.8) with the volume:

$$\Sigma_{\mathbb{R}^2} = \frac{1}{\Delta t} \ln \left(\frac{V_{t_{n+1}}}{V_{t_n}} \right),$$

where V_{t_n} and $V_{t_{n+1}}$ are the cell volumes before and after deformation.

2.3 Adapting the clustering framework to spatio-temporal property data.

The identity of a cell is defined on the basis of its differential genetic state, under the assumption that cells in the same state will behave the same way. Looked at differently, I propose that groups of cells that behave similarly with respect to specific morphogenetic features may have similar genetic state.

To identify relevant cell behaviour, we use clustering methods that consists of grouping individuals behaving the same way and separating those that do not, depending on a chosen metric. It is based on a matrix of pairwise distances between individuals. We aim to define a distance combining count dimension, shape and growth features. Such distance first requires data standardisation to make them of the same magnitude and therefore comparable.

2.3.1 Creating a standardised pairwise distance matrix.

Creating a pairwise distance matrix consists of computing distances between each pair of cells. The same way knowing all the distance amongst a set of cities allows us to find their relative positions in space, the obtained pairwise distance matrix is a representation of the neighbouring or distant cells.

To compute distances between cells, we sought to combine variables attached to cells, temporally differentiated variables, and either the spatial ranked distance between cells, or the Euclidian distance between either centroids or geometrical medians of cells.

We aimed at applying a global clustering to all vertices defining tissues acquired at successive time points. This entailed:

- a global standardization of all the cells of one (or several) meristem(s) along time,
- taking into account topological or Euclidian distances between cells belonging to a given meristem at a given time t ,
- managing side effects induced by missing data potentially arising in several cases:
 - for temporally differentiated variables: depending on the convention for attaching these variables to cells at t_n instead of t_{n+1} there will be no variables at the last observation date;
 - inability to compute a variable (e.g. inability to auto-associate landmarks leading to no areal strain rate), or to remove it (e.g. in case of outliers);
 - non-existence of the variable (e.g. no topological nor Euclidian distances between cells that are not at the same time-point);

This distance then compares between all selected quantified features of the two cells i and j as follows [Guédon *et al.*, 2003; Hastie *et al.*, 2009]:

$$D_{ij} = w_{topo} \frac{d_{ij}}{\text{absd}_{topo}} + \sum_e w_e \frac{\delta(x_{i,e}, x_{j,e})}{\text{absd}_e} + \sum_f w_f \frac{|\Delta x_{i,f}/\Delta t_i - \Delta x_{j,f}/\Delta t_j|}{\text{absd}_f} \quad (3.13)$$

where:

- w_{topo} , w_e and w_f are the weights of topological, spatial and temporal information, respectively;
- with: $w_{topo} + \sum_e w_e + \sum_f w_f = 1$;
- d_{ij} (if it exists) is the topological distance between cells i and j or the Euclidian distance between centroids or geometrical medians (for robustness) of cells i and j expressing cell spatialisation;
- $\delta(x_{i,e}, x_{j,e})$ is the distance for the e^{th} spatial variable;
- $\Delta x_{i,f}/\Delta t_i - \Delta x_{j,f}/\Delta t_j$ is the distance for the f^{th} temporally differentiated variable.

The aim of standardization is to make the local distances independent of the type of the elementary variables (nominal, ordinal, interval-scale, ratio-scale, circular ...) and also independent of the choice of the measurement units. See section 2.3.2 for definition of dispersion measures absd_{topo} , absd_e and absd_f .

Some weights that are chosen in a subjective way on the basis of application knowledge can be introduced. We here propose to chose the weight for each variable category (i.e. w_{topo} , $\sum_e w_e$ and $\sum_f w_f$) and to assume that the different variables have the same weight within each category (see section 2.3.3). In addition, handling missing data is done through re-weighting, respecting $w_{topo} + \sum_e w_e + \sum_f w_f = 1$ as explained in section 2.3.3.

2.3.2 Data standardisation.

Seeking to combine variables attached to cells, temporally differentiated variables and topological relations between cells, we need to standardise the data. The L_1 metric approach is less sensitive to outliers than the L_2 metric approach since the basic quantities are not squared in the computations (equations (3.14) and (3.15)).

e -th variable: Depending on the type of variable $\delta(x_{i,e}, x_{j,e})$ is defined as follow:

- for a nominal variable: $\delta(x_{i,e}, x_{j,e}) = I(x_{i,e} \neq x_{j,e})$, with $I()$ the indicator function;
- for an interval-scale variable and the L_1 metric: $\delta(x_{i,e}, x_{j,e}) = |x_{i,e} - x_{j,e}|$

L_1 metric and pairwise distances: We therefore used the L_1 metric (mean absolute difference):

$$\text{absd}_e = \frac{1}{N(N-1)} \sum_{i \neq j} |\delta(x_{i,e}, x_{j,e})| \quad (3.14)$$

where $\delta(x_{i,e}, x_{j,e})$ is the distance between cells i and j for the e -th variable.

For the L_1 metric, the standardized distances D_{ij} between cells i and j is written as :

$$D_{ij} = w_{topo} \frac{d_{ij}}{\text{absd}_{topo}} + \sum_e w_e \frac{\delta(x_{i,e}, x_{j,e})}{\text{absd}_e} + \sum_f w_f \frac{|\Delta x_{i,f}/\Delta t_i - \Delta x_{j,f}/\Delta t_j|}{\text{absd}_f}$$

where :

- w_{topo} , w_e and w_f are the weights of -respectively- topological, spatial and temporal information,

- all those weight sum up to 1: $w_{topo} + \sum_e w_e + \sum_f w_f = 1$,
- d_{ij} (if it exists) is the topological distance between cells i and j or the Euclidian distance between centroids or geometrical medians (for robustness) of cells i and j expressing cell spatialisation,
- $\delta(x_{i,e}, x_{j,e})$ is the distance for the e^{th} spatial variable,
- $\Delta x_{i,f}/\Delta t_i - \Delta x_{j,f}/\Delta t_j$ is the distance for the f^{th} temporally differentiated variable.

L_2 metric and pairwise distances: One could choose the L_2 metric (mean square difference):

$$\text{quad}_e = \frac{1}{N(N-1)} \sum_{i \neq j} (\delta(x_{i,e}, x_{j,e}))^2 \quad (3.15)$$

where $\delta(x_{i,e}, x_{j,e})$ is the distance between cells i and j for the e -th variable.

For the L_2 metric, the standardized distances D_{ij} between cells i and j is written as :

$$D_{ij} = \sqrt{w_{topo} \frac{d_{ij}^2}{\text{quad}_{topo}} + \sum_e w_e \frac{\delta(x_{i,e}, x_{j,e})^2}{\text{quad}_e} + \sum_f w_f \frac{(\Delta x_{i,f}/\Delta t_i - \Delta x_{j,f}/\Delta t_j)^2}{\text{quad}_f}}, \quad (3.16)$$

where:

- $\text{quad}_{topo} = 1/(N(N-1)) \sum_{i \neq j} d_{ij}^2$ is the measure of dispersion for the standardization of the topological variable.
- $\text{quad}_e = 1/(N(N-1)) \sum_{i \neq j} \delta(x_{i,e}, x_{j,e})^2$ is the measure of dispersion for the standardization of the e -th elementary variable
- $\text{quad}_f = 1/(N(N-1)) \sum_{i \neq j} (\Delta x_{i,f}/\Delta t_i - \Delta x_{j,f}/\Delta t_j)^2$ is the measure of dispersion for the standardization of the temporal variable.

2.3.3 Re-weighting and handling missing data.

Missing topological variables: If cells i and j belong to different time points, there is no topological distance d_{ij} , therefore it is the same as if $w_{topo} = 0$ and that no longer fulfil the constrain $w_{topo} + \sum_e w_e + \sum_f w_f = 1$, hence a re-weighting is necessary. In such case, the standardized distance

(3.13) can be rewritten as:

$$D_{ij} = \sum_e \frac{w_e}{1 - w_{topo}} \frac{\delta(x_{i,e}, x_{j,e})}{\text{absd}_e} + \sum_f \frac{w_f}{1 - w_{topo}} \frac{|\Delta x_{i,f}/\Delta t_{t,i} - \Delta x_{j,f}/\Delta t_{t,j}|}{\text{absd}_f} \quad (3.17)$$

Missing temporal variables: If cells i or j belong to the first observation date assuming that the temporally differentiated variables are attached to the final date (and not to the initial date), the standardized distance (3.13) can be rewritten as:

$$D_{ij} = \frac{w_{topo}}{1 - \sum_f w_f} \frac{d_{ij}}{\text{absd}_{topo}} + \sum_e \frac{w_e}{1 - \sum_f w_f} \frac{\delta(x_{i,e}, x_{j,e})}{\text{absd}_e} \quad (3.18)$$

The two rewritings (3.17) and (3.18) can be seen as two examples of managing missing variables in the computation of standardized distances between cells.

2.3.4 Clustering a pairwise distance matrix.

The aim of cluster analysis is to group data into homogeneous groups or clusters [Kaufman and Rousseeuw, 1990; Gordon, 1999]. Cluster analysis can be viewed as an exploratory tool revealing the structure embedded in a pairwise matrix of distances between individuals. The two main categories of clustering methods are partitioning methods and hierarchical methods, whose advantages and drawbacks are thoroughly discussed in Kaufman and Rousseeuw [1990]. For partitioning methods, the resulting set of clusters, whose number is fixed by the user, satisfies the requirements of a partition: each cluster must contain at least one individual; each individual must belong to exactly one cluster. Considering hierarchical methods, a hierarchically-nested set of partitions is constructed. These hierarchical methods are either agglomerative or divisive. Agglomerative methods start with each individual in a unique cluster and iteratively merge the two closest clusters. Divisive methods work the other way around by splitting the cluster presenting the highest within-cluster distance into two clusters. Both methods are therefore computationally expensive since they consider at every step all the possible merges or divisions.

We chose to make use of the clustering algorithms available in Scikit-learn [Pedregosa *et al.*, 2012], with a particular focus on selecting those presenting the capacity to detect non-convex clusters. Therefore we selected two methods available in the software:

- **Ward** clustering is an agglomerative hierarchical method that *minimizes the sum of squared differences* within all clusters. It is thus a variance-minimizing approach, similar to the k-

means objective function but not limited to convex cluster detection and with the possibility to add connectivity constrains [Pedregosa *et al.*, 2012];

- **Spectral** clustering performs a low-dimension embedding of the pairwise distance matrix, followed by a K-means in the low dimensional space. K-means is bound to convex cluster detection, however the dimensionality reduction might change that and non-convex clusters and even embedded (i.e. clusters within) clusters can be obtained when coming back to the original space [Pedregosa *et al.*, 2012].

Assessment and post-processing methods (e.g. multidimensional scaling methods) will be used as presented later in 2.3.6.

Finally, it is important to stress that in our context, clustering methods focus mainly on the spatial dimension of data, the temporal dimension being achieved by the inclusion of temporally differentiated variables assigned to vertex, either at t_n or t_{n+1} .

2.3.5 Managing outliers for the standardisation step.

Both the clustering and the standardisation step are sensitive to outliers. As presented in section 2.3.2 and in equation 3.13, rely on a measure of data dispersion. By modifying data dispersion, outliers contribute modifying the distance distribution between variables.

It is possible to select these outliers in two ways: manual definition of a threshold for each cell feature, or automatic detection. The first one is simple and straightforward, however it may not be repeatable. On the opposite, a reliable automatic detection of outliers is hard to achieve.

Finally, handling those detected outliers is also a tricky task. Removing them from the standardisation value computation process will preserve distances between core elements (those close to the mean value), but will make the outliers even more obvious for the clustering. On the other hand keeping them when standardizing will artificially make outliers comparable to the absolute largest values of other features. One solution would be to exclude them from standardisation and clustering analysis beforehand (i.e. form the pairwise distance matrix) and either add them as an extra cluster or group them with the cluster presenting the highest (or lowest if negative values) values for the property where these outliers comes from. This will also be challenging when using multivariate clustering. The safest solution would therefore be to exclude them from the analysis altogether and tag those data as “impossible to explain” with the selected model.

Some clustering methods such as the density-based clustering method DBSCAN [Ester *et al.*, 1996] apply such ideas by labelling and excluding outliers when detected. Another great strength of this method is its capacity to automatically define the number of clusters. However we were unsuccessful when testing this clustering method with our data since the two parameters of this method (maximum distance between two individuals and the number of individuals in a neighbourhood for a point to be considered as a core point) defining the “density” were difficult to adjust properly. As noted in Ester *et al.* [1996], it works well for data that contains clusters of similar density, but we have no *a priori* knowledge regarding that property for our data.

2.3.6 Clusters assessment and inspection.

When a clustering is achieved, either from a manual annotation or from an algorithmic method, the quality of the clusters must first be assessed using data and clusters representations. Since clustering is computed based on a standardized pairwise distance matrix, therefore even in the case of manually defined clusters it is necessary to select which features to use and their respective weights, to build this matrix before interpreting the clusters.

This inspection step is important to assess the validity of the clustering, but also to gain insight into how the clusters are made and what limitations to apply to our mathematical and biological conclusions.

Vertex to cluster distance allows the ordering of the vertex by ascending average distance $D(i, q)$ to other cells assigned to the same cluster i.e. from the least (most central) to the most outlying. For a vertex i , its distance to cluster q is computed as follow:

$$D(i, q) = \frac{\sum_{j \in q; j \neq i} D_{ij}}{N_q - 1} \quad \text{if } i \in q,$$

$$D(i, q) = \frac{\sum_{j \in q} D_{ij}}{N_q} \quad \text{if } i \notin q,$$

where N_q is the number of cells assigned to cluster q .

In a representation, the vertices are usually indexed (rank) by their distance to their cluster. We then can readily detect the well-classified individuals inside the cluster since they will show only a small increase in their distance compared to the previous ranked element. On the contrary, those who present a drastic increase in term of distance (highlighted by the ranked distribution in the other axis) can be declared as poorly classified. The presence of a large group of poorly classified

elements in one or more clusters can indicate the necessity of adding supplementary clusters to obtain a potentially better defined clustering. These poorly classified elements are always found at the margin of a meaningful cluster.

Within- and between-cluster distances are important criteria to consider since they quantify cluster-compactness (within-cluster distances) and cluster-separation (between-cluster distances).

$$D_{\text{within}}(q) = D(q, q) = \frac{\sum_{i,j \in q; i \neq j} D_{ij}}{N_q(N_q - 1)},$$

$$D_{\text{between}}(q) = D(q, \ell) = \frac{\sum_{i \in q; j \in \ell} D_{ij}}{N_q N_\ell} \quad \text{if } q \neq \ell.$$

where $N = \sum_q N_q$ is the total number of cells, N_q and N^ℓ the number of cells in clusters q and ℓ .

The within/between-cluster distance matrix is therefore a crucial tool to assess the quality of a clustering. Indeed a good clustering should have at least smaller within-cluster distances (diagonal) than between-cluster distances (off diagonal). This ascertain the absence of overlapping clusters in their original high-dimensional space. See methods 2.3.6 for the definition of global between- or within-cluster distances.

Global between- or within-cluster distances can be computed from the $D(q, q)$ within-clusters and $D(q, \ell)$ between-clusters distances as follow:

$$D_{\text{within}} = \frac{\sum_q \sum_{i,j \in q; i \neq j} D_{ij}}{\sum_q N_q(N_q - 1)} = \sum_q \frac{N_q(N_q - 1)}{\sum_\ell N_\ell(N_\ell - 1)} \frac{\sum_{i,j \in q; i \neq j} D_{ij}}{N_q(N_q - 1)} = \sum_q \frac{N_q(N_q - 1)}{\sum_\ell N_\ell(N_\ell - 1)} D_{\text{within}}(q),$$

$$D_{\text{between}} = \frac{\sum_q \sum_{i \in q; j \notin q} D_{ij}}{\sum_q N_q(N - N_q)} = \sum_q \frac{N_q(N - N_q)}{\sum_\ell N_\ell(N - N_\ell)} \frac{\sum_{i \in q; j \notin q} D_{ij}}{N_q(N - N_q)} = \sum_q \frac{N_q(N - N_q)}{\sum_\ell N_\ell(N - N_\ell)} D_{\text{between}}(q).$$

These indicators will be relevant if the clusters are convex. Otherwise, other indicators (e.g. the average distance to the k -nearest neighbors within the cluster that take account of the density-based clustering assumption) should be proposed.

Multi-dimensional scaling (MDS) can give a representation of the original high-dimensional distance problem (tackled by the clustering) by reducing it to a 2D or 3D representation. This dimensionality reduction method is comparable to other multi-dimensional methods like principal component analysis (PCA) or singular value decomposition (SVD). They indeed present the par-

ticularity to compute and select the most representative dimensional axes of a given dataset. By selecting 2 or 3 of the most representative axes it is then possible to represent the data in a lower dimension (i.e. observable dimensions like 2D or 3D).

MDS is different from PCA and SVD since it is specifically designed to preserve the distance metric separating the elements simultaneously to completing the representative axes research. However, it is possible that the 2 or 3 first selected axes represent only a small percentage of the overall information (sum of the firsts eigenvalues divided by the sum of all eigenvalues). In that case, the MDS representation is not meaningful and should be discarded. Plotting the elements of the pairwise distance matrix with a color code corresponding to cluster association can give a representation of the spatial distribution within the cluster (elements at the core or margins of the clusters). This representation will be as close to the within/between-cluster distance matrix as the selected axis are representative of the overall information.

Features distribution per cluster highlight feature contribution to their respective cluster therefore enabling clusters characterisation (e.g clusters with high, medium and low volumes). It can also be helpful in detecting a split or absence in the data distribution property by adding or removing a cluster in a range of clusters for the same feature combination. It is coherent with a view where the clustering algorithm could yield a mixture-like structure when attributing cells to groups. Looking at the properties distribution per cluster could then assess the presence of well defined mixture component by showing their overlap or separation.

Projecting clusters on the segmented image allow us to picture the output of the clustering in a spatial fashion. It is a simple representation, but very important in order to correlate and analyse this information in their biological context.

Clustered spatio-temporal graph representations provide a synthetic view of the temporal evolution of the patterns at a cluster level. There are actually two ways to look at it: where do the cells come from and where are they going? These ideas and representations are a good door to more detailed analysis of the temporal process at a sub-tissular level (groups of cells behaving the same way temporally is an identity specific pattern) by hidden Markov tree.

2.3.7 Clusterings comparison with bipartite matching of cluster composition.

To compare clustering we sought at matching cluster composition to get a clearer picture of the differences. In the case of nested-clusters obtained from hierarchical clustering methods, this is even more pertinent since we can then keep coherent clusters ids for a given range of clusters number. For instance, comparing a 3- and 4-clusters model with bipartite matching gives us the id of the new cluster.

Cost function matching elements of ensemble i to those of ensemble j is defined as follow:

$$\text{cost function}(\{i\}, \{j\}) = \frac{2 * \{i\} \cap \{j\}}{n_i + n_j} ,$$

where $\{i\}$ means ensemble i (i.e. set of vertex ids in cluster i), n_i is the number of elements in ensemble i and \cap means ensemble intersection.

Bipartite matching is then obtained by computing the cost function for each ensembles $i \in I$ with respect to all ensembles $j \in J$. For each ensemble from i we then take the sum of the cost function minimising the overall sum.

2.3.8 Supplementary function definitions.

Geometric Median The geometric median of a discrete set of N sample points X_i in a Euclidean space (\mathbb{R}^3) is the point minimizing the sum of distances to the sample points. This generalizes the 1D-median (minimize the sum of distances for one-dimensional data) as follow:

$$\text{Geometric Median} = \arg \min_{y \in \mathbb{R}^n} \sum_{i=1}^m \|x_i - y\|^2 , \quad (3.19)$$

where $\arg \min_{y \in \mathbb{R}^n}$ is the value of the argument y which minimizes the sum, meaning the point y from where the sum of all Euclidean distances to the X_i is minimum.

3 Results

3.1 Exploratory analysis of the temporal property graph

I now introduce some example of possible exploratory analyses. Starting with a simple test of collinearity between vectors, I will show that it is possible to verify the “universal rule for symmetric division” demonstrated in Besson and Dumais [2011]. Then, on the basis of a manual annotation (morphology oriented) of the last time point, I will highlight the necessity of using unsupervised methods to define relevant regions on the basis of geometrical features. Finally, using the clustered version of an acquired genetical signal, I will present a first evidence of a link between gene and shape.

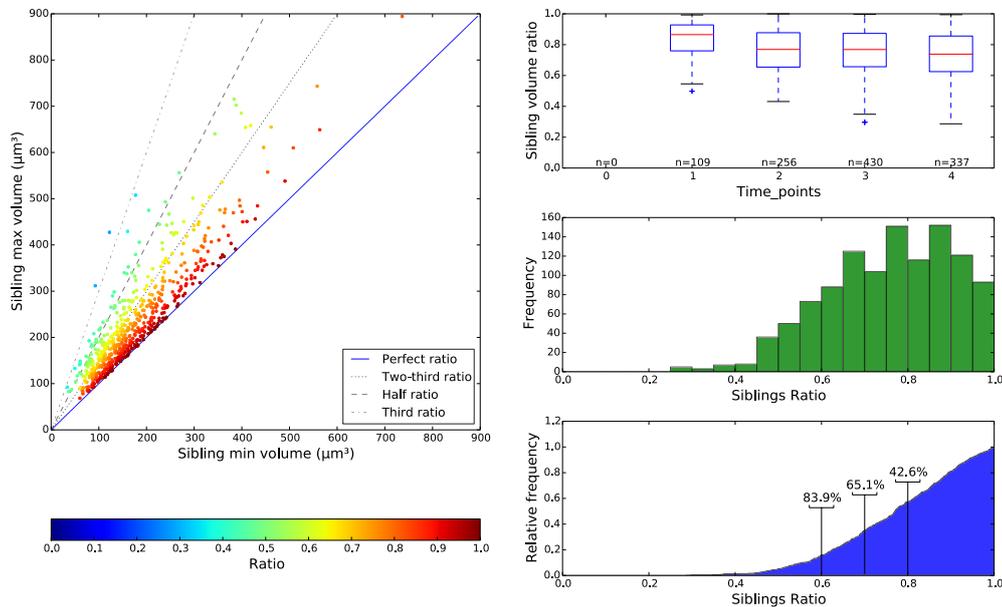
3.1.1 Taking the “universal rule for symmetric division” presented in Besson and Dumais [2011] to the third dimension

Here I consider the case were one is interested in using the spatio-temporal graph framework for the inspection of known or putative properties of the tissue. Empirical evidences suggest that tensional forces within the cytoskeleton provoke division plane to align along the plane that minimise its area. This empirical property of cell division, known as the Errera’s rule, thus yields two daughters of equal size. Using a mechanistic model of division and a maximum entropy formulation, Besson and Dumais [2011] confirmed that the probability of observing a particular division configuration increases inversely with its relative area, independently of the cell size and shape. In addition, they confirmed that this properties was imputable to tensional forces within the cytoskeleton who is positioning the cell nucleus during mitosis.

Size symmetry among the siblings.

Besson and Dumais [2011] tested this hypothesis and built their model using 2D features and 2D observation of cells. Later on Yoshida *et al.* [2014] analysed in 3D the division patterns found in the developing embryo of *A. thaliana*, but to my knowledge no-one have ever looked it with 3D+t data. I do not pretend to demonstrate in 3D+t the rule, but merely test if the size symmetry among the siblings holds when using the observed division statistics of 3D+t data. To that end, I have first compared the observed volumes of sibling cells, as well as their ratios. We here need to assume that, between the successive time-points, the descendant did not undergone too important

deformation after the establishment of the division wall. In order to be thorough I thus computed the ratio for sibling cells coming from a binary division (i.e. only two descendants) as well as those detected within the sublineage (hand-made) to compute volumetric ratios for all cells presenting a division even in the case of parallel divisions (4 cells).



a. Bi-plot of sibling against sibling **b.** Distribution of siblings volume ratios

Figure 3.21: Comparison of sibling volumes similarity computed over the whole time-series. **a.** Each coloured point represent the coordinate defined by the two siblings volumes; the blue line represent the coordinates for a perfect symmetry in terms of volume. The abscissa present the sibling with the minimum volume and the ordinate the one with the maximum volume, thus all the information is on the same side of the 'perfect ratio line'. However, the orthogonal distance to this 'perfect ratio line' is not linear for the sibling volumes ratios, since bigger cells can be further away from this line than smaller cells but present the same ratio. I thus coloured the points according to their respective volume ratios and given three others ratios lines: $1/3$, $1/2$ and $2/3$ to help reading the illustration. **b.** Boxplot per time-point (top), distribution (middle) and cumulative distribution (bottom) of the sibling volume ratios over the whole time-series; the vertical lines indicate the percentage of sibling under the curve starting from 1 (the perfect ratio) to the line defined ratio (e.g. 42.6% of the ratios are superior to 0.8). The box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data as follow: $Q3 + 1.5 * IQR$, where $IQR = \text{interquartile range}$ (i.e. $Q3 - Q1$). Flier points are those past the end of the whiskers.

Using figure 3.21a plotting siblings coordinates on the basis of their volumes, we observe that the smaller cells tend to present more divergence in their volumetric ratio than bigger cells. Although this conclusion is limited by the difference in terms of individuals within each population (large for small and limited for big cells). The boxplot I present at the top row of figure 3.21b indicate that the dispersion of the interquartile value does not change over time. Although a slight decrease in the median can be observed, this seems to be the result of a greater variability of the 'extreme' values highlighted by the extension of the whiskers over time. This can be interpreted

either as an increase in segmentation errors over time, or by the presence of more divisions going against Errera’s rule. Analysing middle and bottom illustrations of figure 3.21b giving sibling volume ratios, highlight a correct adequacy of the observed ratios to Errera’s rule. Indeed, nearly 84% of the siblings present a volume ratio superior to 0.6, and nearly half (42.6%) of these ratios are above 0.8. This is quite good considering the potential error rate of the segmentation algorithm.

All these representation and their interpretations are limited by the absence of spatial information. To further analyse the ratio, I thus give in figure 3.22 a spatial projection of the ratio onto the segmented time-series. Observing these illustration, we can observe that the ratios do not present a spatial bias, meaning that no obvious spatial regularity can be found in the spatial distribution of the ratios. In addition, as shown by the top figure 3.21b, t1 seems to present a high coherence to Errera’s rule, but for t2 to t4 the spatial information does not highlight any temporal changes in the sibling volumetric ratio (except a limited amount of data for t4, but this is consistent with the low division number observed between t3 and t4; see supplementary figure S3.5).

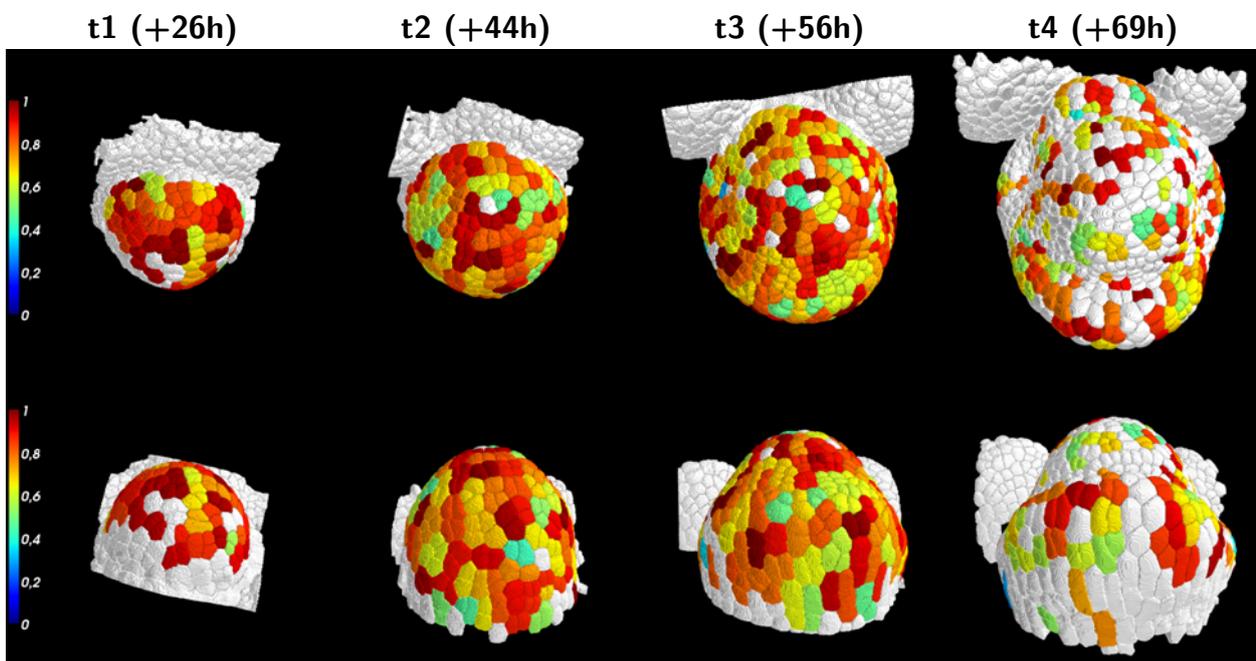
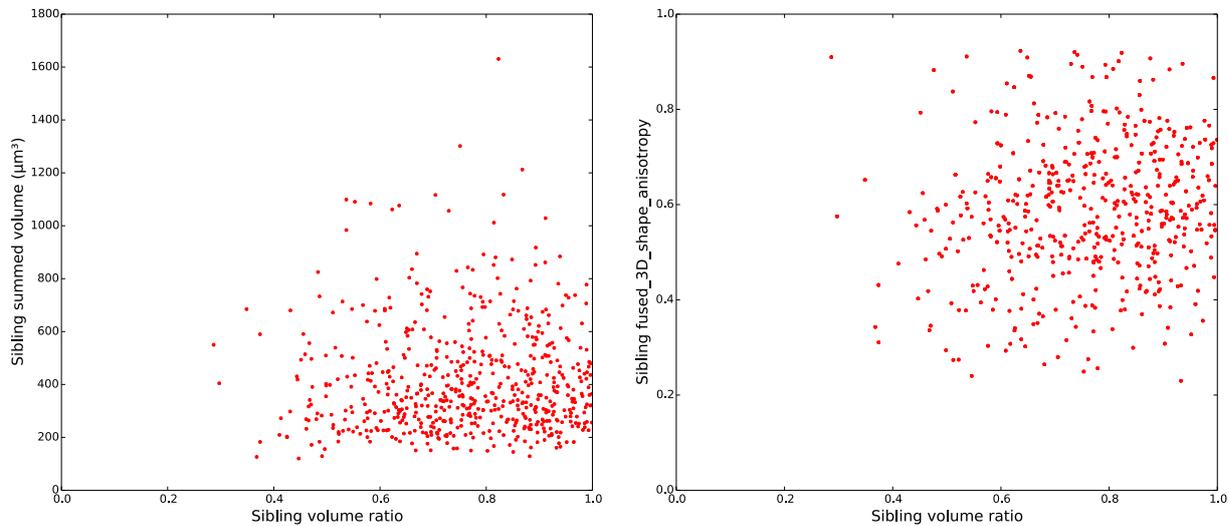


Figure 3.22: Spatial projection (top and abaxial views) of the siblings volume ratio at $t_n + 1$. In order to be thorough we uses the sublineage (hand-made) to compute volumetric ratios for all cells presenting a division The t4 present a limited amount of data, but this is consistent with the low division number observed between t3 and t4 (see supplementary figure S3.5).

I may now finally test the independence of the volumetric ratio to cell size and shape, and to that end we present in figure 3.23 two representations of the volumetric ratio against each of the sibling

volume sums (figure 3.23a) and the 3D shape anisotropy obtained from the inertia tensor computed with the fused-siblings (figure 3.23b). Analysing figure 3.23a, we observe a slight tendency of small



a. Sibling volume sums

b. Sibling fused 3D shape anisotropy

Figure 3.23: Representation of the sibling volume ratio correlation to **(a.)** the sibling volume sums and **(b.)** the 3D shape anisotropy obtained from the inertia tensor computed with the fused-siblings.

cells to sometimes present a low ratio. Apart from that the dispersion of the rest of the data do not display any correlation. In addition, observing figure 3.23b, we note that it is also impossible to distinguish any correlation between the volumetric ratio and the shape anisotropy. We can therefore conclude to the independence of this volumetric ratio respectively to size and shape.

Division plane orientation respectively to the main axis of inertia.

The key statement of the “universal rule for symmetric division” is that plant cell walls are suppose to divide a parent cell along a plane that minimizes its area. This is for the tensional forces within the cytoskeleton to be minimized, positioning the cell nucleus close to the barycenter of the cell, which is the origin of the tensor of inertia. The division plane should thus be orthogonal to the main axis of inertia, since the main vector is oriented along the maximum length of the cell and the two others form an orthogonal plane close to the one of minimal area.

Possessing the division wall orientations and being able to compute the main axis of inertia, I thought at computing their collinearity in a 3D space. Since the division wall can only be observed at t_{n+1} , we have to compute the inertia tensors at this time point too. Therefore I thought at “fusing” the descendants before computing their inertia tensor. I here had to hypothesise that the descendants shapes are close to the mother shape right before division.

The illustrations of figure 3.24 present the spatial projection of the vectors collinearity results as computed at t_{n+1} (see supplementary figure S3.2 for the same data projected at t_n). Analysing those results, the developing L1 of the acquired FM seems to follow the proposed pattern of Besson and Dumais [2011]. Indeed, most of the descendants cells present a division wall orthogonal to the main inertia axis computed on the fused siblings, since 92% of the vectors present a collinearity of at least 0.8 (see figure 3.25).

As for the sibling volume ratios, no temporal bias can be observed although for the last time-point (t_4) it seem that more cells deviate from the –globally observed– high collinearity (see figure 3.25, left illustration). However, this may be due to a possible spatial dependency displayed by the vectors collinearity measure that seems to present more orthogonal main axis of inertia and division plane at the bottom of the abaxial side (figure 3.24).

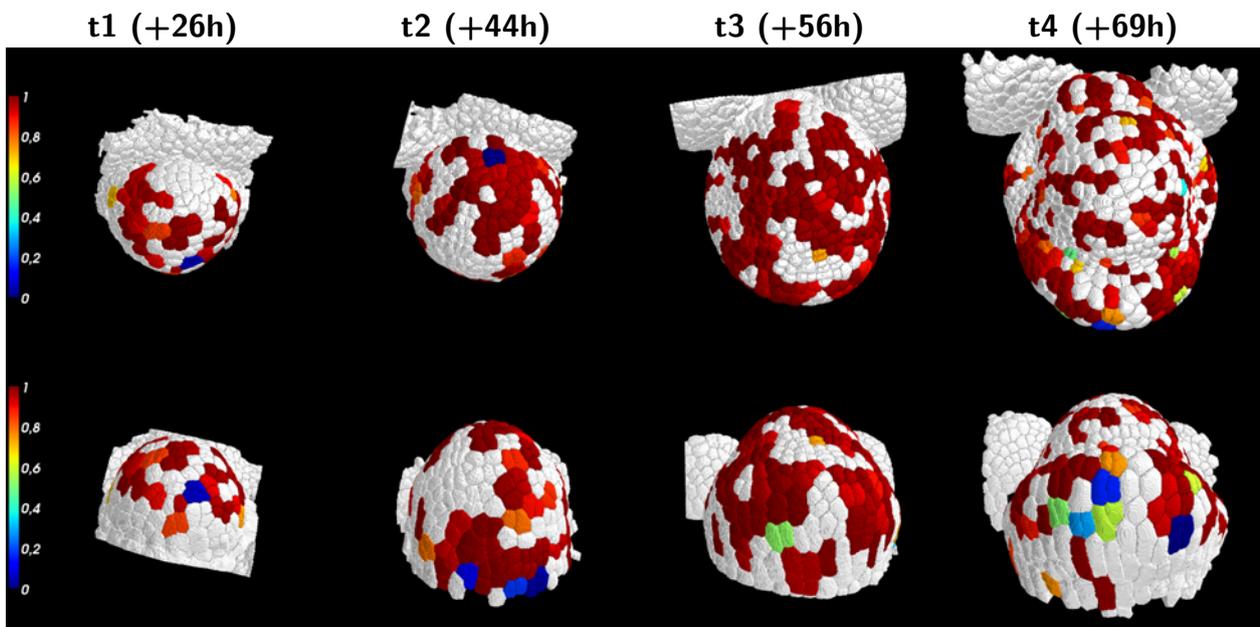


Figure 3.24: Main axis of inertia and division wall normal vectors correlations at t_{n+1} . Collinear vectors have a value close to 1 and orthogonal vectors close to 0. We use the inertia axis obtained with the fusion of the daughters since the division wall is defined at t_{n+1} . Only the cells presenting a binary division (i.e. only for two siblings) are presented.

Finally, the main axis of inertia and division plane collinearity does not seem to be influenced by neither the volume of the cell, nor its shape (see figure 3.26). This property seems independent of both size and shape of the cell, but may be spatially influenced as seen on the abaxial views (bottom row of figure 3.24), although the limited number of observations prevents us to draw general conclusions.

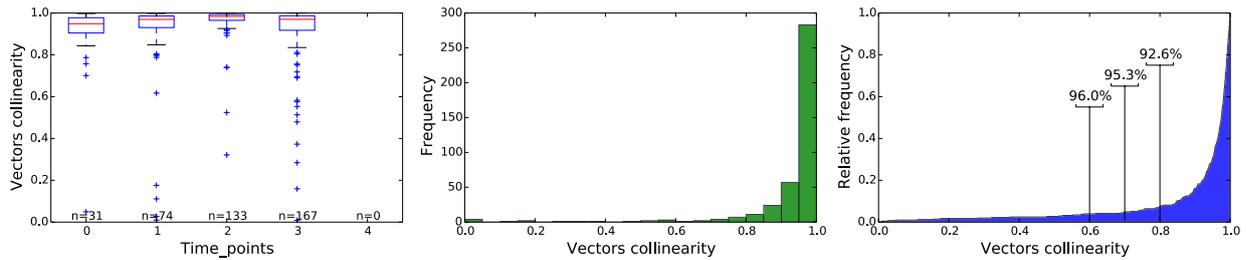
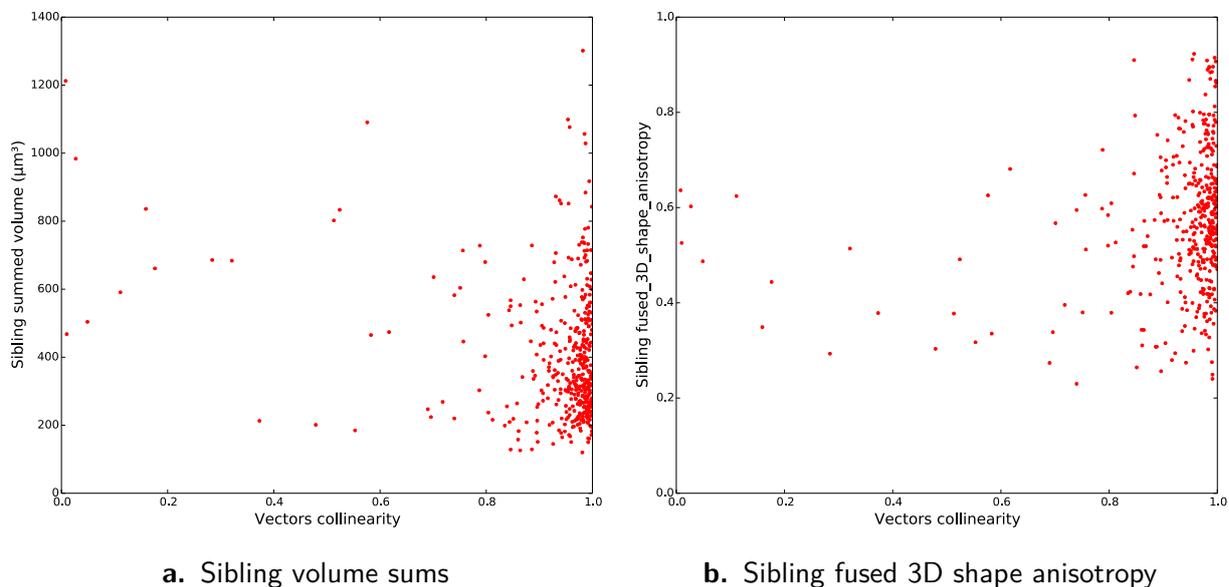


Figure 3.25: Distribution of main axis of inertia and division plane normal vector collinearity computed over the whole time-series and attributed at t_n . Boxplot per time-point (left), distribution (middle) and cumulative distribution (right) of the sibling volume ratios over the whole time-series; the vertical lines indicate the percentage of sibling under the curve starting from 1 (the perfect ratio) to the line defined ratio (e.g. 92.6% of the ratios are superior to 0.8). The box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data as follow: $Q3 + 1.5 * IQR$, where $IQR = \text{interquartile range}$ (i.e. $Q3 - Q1$). Flier points are those past the end of the whiskers.



a. Sibling volume sums

b. Sibling fused 3D shape anisotropy

Figure 3.26: Representation of the vectors collinearity in correlation to **(a.)** the sibling volume sums and **(b.)** the 3D shape anisotropy obtained from the inertia tensor computed with the sibling fused.

3.1.2 Mathematical relevance of user defined regions.

One of the most obvious ways to explore morphogenesis is by exploring the behaviour of known, predicted or putative groups of cells within a tissue. It might be of importance indeed to assess previous biological studies quantitatively, meaning to test the features distribution in groups of cells predicted by gene expression or morphological studies. For example, the $t4$ introduced in figure 3.3, is a stage 3 with its attendants morphological features such as the emerging sepals on the sides, and a central dome in the middle, separated by a boundary region it between. Manual annotation of each apparent morphological region by an expert biologist define what we call an “expert defined clustering” as presented in figure 3.27. Since we have defined a way to compute a

standardised pairwise distance matrix (see equation 3.13 section 2.3.1), we can assess this ‘expert defined’ clustering regarding selected features.

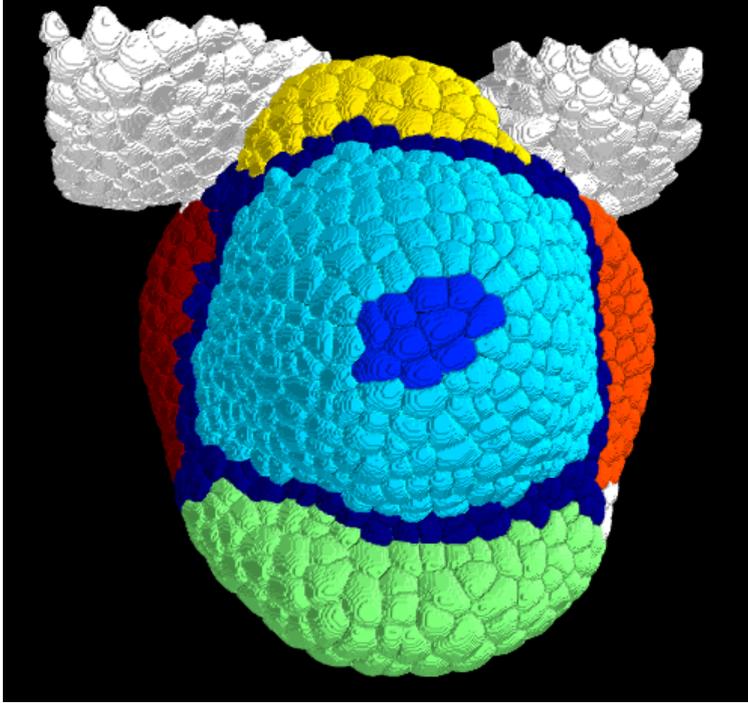


Figure 3.27: Example of morphology related regions defined on the last time point (t4). We manually defined the boundary (dark blue) region between the central dome and the sepals. The central zone (blue) is surrounded by a peripheral zone (light blue). Each sepals are reported in different colors (abaxial: light green; adaxial: yellow; left: dark red; right: orange).

Before building a pairwise adjacency matrix, we need to select, among the available features, those displaying differences by regions. To help select regions with a differential behaviour in terms of feature, I first represented the distribution of the features by cluster as boxplots in figure 3.28. Such representation provides preliminary insights into the cell feature distributions by regions.

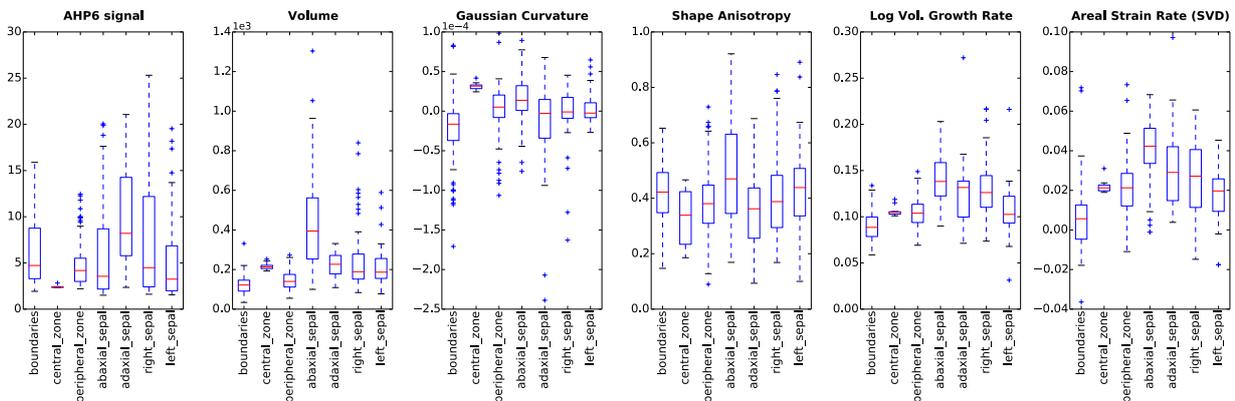


Figure 3.28: Example of features distributions by morphologically defined regions (at t4).

As it was expected, using manually defined regions of the basis of morphological features, we are far from regions defined by mixture models (i.e. with non-overlapping distribution for a cell

feature). Although, it is possible to obtain some information that might be crucial when using unsupervised partitioning. Indeed, the boundaries present the lowest median and overall values for the following features: volume, Gaussian curvature, log volumetric growth rate and areal strain rate. Especially for the two temporally differentiated feature (log volumetric growth rate and areal strain rate) it is quite clear that using them within an unsupervised framework would reveal the boundary by grouping cells presenting low values.

As presented in figure 3.2 (p. 136), the expression domain of a gene called *ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN 6* (*AHP6*) has been acquired. *AHP6* is a negative regulator of cytokinin signalling, that is expressed very early during flower primordia initiation and during sepal initiation at morphological stage 3 [Bartrina *et al.*, 2011; Besnard *et al.*, 2014]. Considering this acquired genetical signal, it should be noted that only the central and peripheral zone display a rather concentrated distribution of values, when other region seems more spread out (figure 3.28). With the notable exception of the adaxial sepals with a high median but a small number of individual defining the region, the medians of the other regions seems comparable. This seems contradictory with the previous observation, but this is in fact due to the large number of individual composing the sepals regions, thus grouping rather different *AHP6* levels that lead to a high dispersion, but not apparent differences in the medians.

The shape anisotropy feature seems incapable of differentiating any regions, except maybe for the abaxial sepal which display a large distribution due to highly anisotropic cells at the base of this sepal. This abaxial sepal also present high values for the temporally differentiated feature (log volumetric growth rate and areal strain rate). In a less obvious fashion, the adaxial and right sepal also present higher values for those temporal features. The left sepals might be late in its formation since it display globally lower values for those features even presenting a median close to the central and peripheral zones. In the perspective of defining cellular identities on the base of growth related features, using the temporally differentiated features, log volumetric growth rate and areal strain rate, would therefore be a good idea to separate biologically relevant regions such as the sepals, the boundary and the central dome.

Finally, I would stress that the regions can be defined by morphology or genetic identity. In that respect, using a gene atlas (maps of gene expression domain) or an embedded fluorescent marker will also enable a statistical comparison between genetic and feature patterns. Although

in the case of the atlas, which does not exist yet in a numerical form at cell resolution for *A. thaliana*, it would also be necessary to have it at a developmental stages as close as possible to our acquisitions. In addition, we have seen that the genetic signal, AHP6, was not well defined by the manual annotation and since using such important features as developmental genes might be crucial in understanding the link between gene and shape, I propose to explore it further in the next section using unsupervised methods.

3.2 Cellular pattern identification on the basis of geometrical features clustering

The main aim and interest when doing clustering analyses is to simplify complex data by grouping individuals presenting similar features, thus summarising and highlighting their similarities and differences. It is therefore an helpful tool for researchers to understand the main mechanism at work and their relations within complex data. It prove particularly useful to simplify quantitative data in a modelling perspective where only a limited amount of structure can be modelled (e.g. ordinary differential equations) and a quantified variability can be introduced.

The major challenge having all those different features and many ways to combine them with different weightages is that the resulting number of possible analyses is very large. Therefore I begin by explaining univariate clustering below, to illustrate the possibilities and outcomes of such analyses. I will first illustrate the capacity of the spatio-temporal graph framework to test the potential relation between an acquired gene pattern against geometrical features in order to determine their morphological function. We here want to test if a direct relation can be done within the 3D+t geometrical features. The capacity of the clustering methods to groups individuals of comparable genetic signal will provide groups for which we will analyse the features distributions.

3.2.1 Relating genetic spatio-temporal patterns to quantitative feature changes by clustering AHP6 signal.

Acquiring genetic spatio-temporal patterns is indeed a key possibility offered by live-imaging techniques, allowing us to follow the expression and distribution patterns of a tagged protein through time. Tagging proteins with fluorescent markers is becoming a common tool in many model organism like *A. thaliana*. It is also possible to follow several proteins by tagging them with fluorescent

marker emitting at different wavelengths. Indirect tracking of other biological mechanisms like hormone concentration or activity is also possible [Brunoud *et al.*, 2012]. However these “simple” constructions usually do not provide a quantitative marker, meaning that changing acquisition parameters, between acquisition for instance, will artificially alter the amount of signal in the image. Nevertheless, for a single image, the relative levels are correlated to fluorescent protein concentration, although not necessarily in a linear manner. Therefore, considering as identical the signal levels from successive acquisitions of the same object (or between different objects) is wrong, but they are comparable. Many artefacts can indeed change the relative levels of expression:

- repetitive acquisition in a short time frame can induce photo-bleaching, i.e. quench the fluorochrome molecules;
- large or unequal intervals of acquisition can result in large variations of fluorochrome-tagged protein levels subsequently to changes in the metabolic activity (daily and long term regulated);
- change in confocal acquisition parameters.

To begin to correlate genetic patterns to morphogenetic features, I have made use of the clustering method to group individuals by features proximity to discretize the acquired genetic signal. There are several ways of discretising the genetic signal, but binarisation (presence/absence) and 3-4 clusters (None, Low, Medium and High) are the most interesting cases biologically speaking. Binarisation should be indeed useful to create an atlas of the tracked genetic patterns, and the categorisation into 3 or 4 signal levels should enable interesting correlations with other features.

Figure 3.29 (top row) shows the AHP6 signal projected onto the cell-segmented reconstruction. Using only the AHP6 signal spatial feature to build the pairwise distance matrix and applying a 4-group hierarchical Ward clustering method, we obtained a clustering given in figure 3.29 (bottom row). Comparing the two rows of figure 3.29, it is easy to understand the meaning of each cluster holds even without looking at the feature distribution values of every clusters (figure 3.30). Furthermore, representing the distribution of the signal per cluster in figure 3.30 (AHP6 signal), it is clear that the clustering acts as a mixture model, separating individuals into non-overlapping groups of signal levels. We can further explore this by now considering the other features (Gaussian curvature, volume, log volumetric growth rate and division rate) and observing their relation to the discretised signal levels (figure 3.30. In that respect, cell volumes does not seem to be significantly influenced by the activity of AHP6. The shape anisotropy might be related to the signal levels,

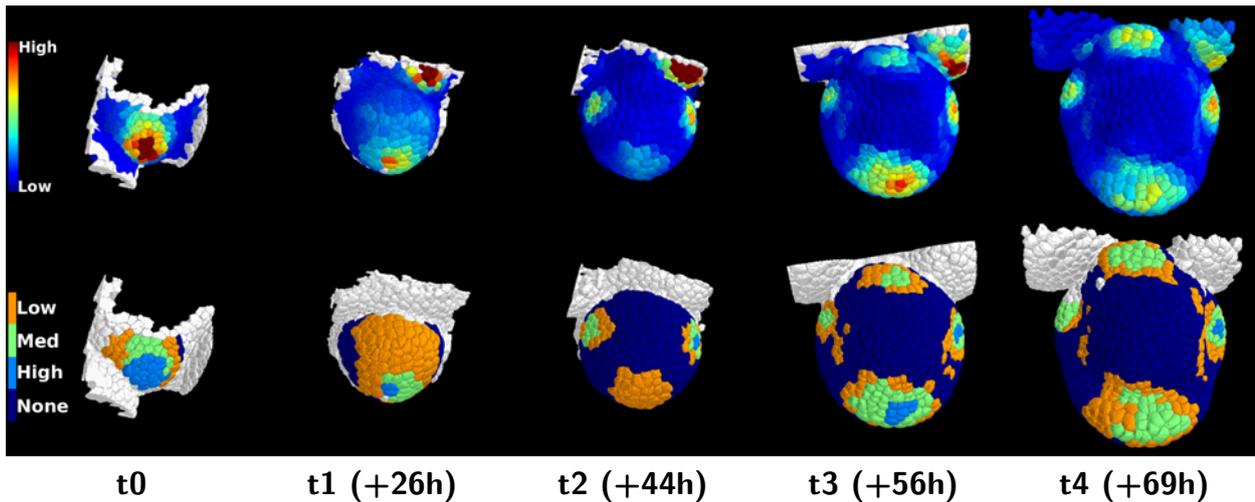


Figure 3.29: Example of an acquired genetical marker: AHP6 signal (top) and the associated 4-groups clustering by ward method (bottom). In the top row the colormap range from low-intensity signal in dark blue to high-intensity signal in dark red. The highest intensities have not been reproduce for lower signal readability reasons. The threshold used is presented in supplementary figure S3.1. In the bottom row the colormap is made as follow: no-signal = dark blue; low-intensity signal = orange; medium-intensity signal = light green; high-intensity signal = light blue.

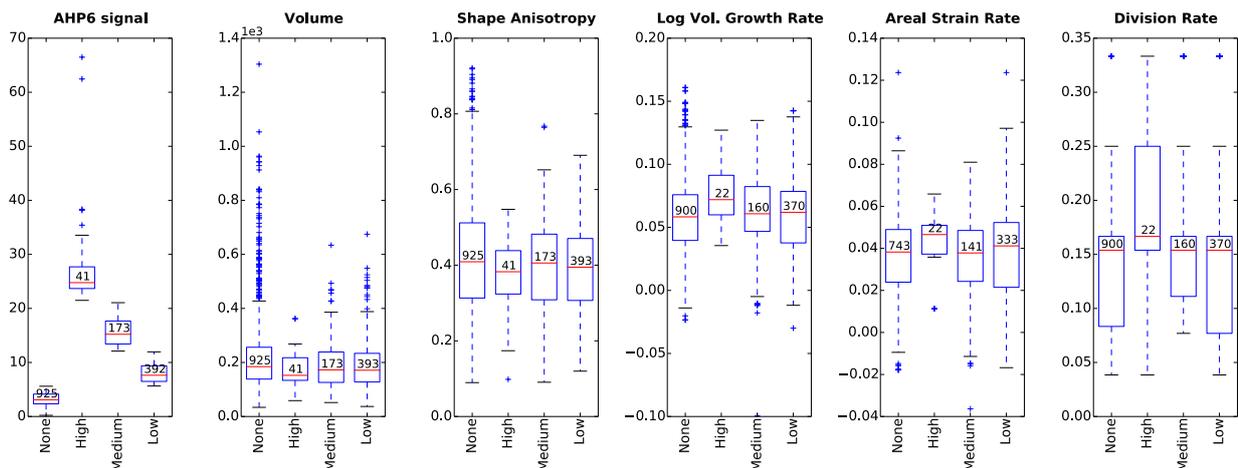


Figure 3.30: Cell features boxplot per groups obtained from 4-clusters Ward clustering of AHP6 signal. We give the distribution of AHP6 signal, 3D shape anisotropy, volume, log volumetric growth rate and division rate. It seems that the log volumetric growth rate and division rate are related to AHP6 signal level since both show an increase relate to the increase of signal. Volume property seems inversely proportional to AHP6 signal levels, but it might be a secondary effect of increased growth and division rate.

since a slight reduction of the data dispersion around the median placed below those of the other level-related groups. Although I may point out that this could be circumstantial since the number of individual is significantly smaller and for this feature the data dispersion might be biased. Finally, it seems that volumetric growth rate, areal strain rates and division rate are positively related to AHP6 signal level since they all display an increase related to the increase of signal. At least high levels of AHP6 could induce an increase of the cell growth rates, and a more graduated

response for the division rates can be observed.

Finally, I would like to stress that it is better to search for relations between signal-discretised groups and geometrical features rather than a full linear or more complex relation, since:

- these are noisy data and the obtained correlation would be poor;
- there are uncertainties coming from the fact that the signal is not quantitative, and especially that relative levels between acquisitions might change slightly.

3.2.2 Univariate clusterings on the basis of geometrical features.

In the following paragraphs I present two univariate clusterings on different features to illustrate the kind of insight one can get from such methods. We here focus on finding biologically relevant clusters using the Ward hierarchical agglomerative method, for the two reasons that follows:

- its capacity to detect non-convex clusters;
- to obtain hierarchically-nested cluster (clustering tree).

This second property, together with a bipartite graph matching of cluster composition, makes the analysis of clustering outputs for successive numbers of clusters much easier. Indeed, going up (fewer clusters) or down (more clusters), will result in the merging of two clusters or the splitting of one cluster without the possibility of redistributing of the individuals to other clusters. This make the successive clusterings highly coherent and therefore facilitates their analysis.

Before starting the analysis of the unsupervised clustering outcomes, I would like to stress that I do not expect the methods to explain all relations between all measured geometrical features for every time point with a high biological coherence! This is because I use algorithmic clustering, as opposed to model-based clustering, and thus the absence of a spatio-temporal model will limit the capacity of the methods to explain several identities at once. Indeed there are many complex (multi-scale, inter-dependent, ...) relations between all these features, making it impossible to unravel at once!

In the case of univariate clustering I expect to find at least one strongly biologically relevant cluster per feature, and using a reasonably small range of clusters numbers (e.g. from 4- to 6-cluster clusterings, which might ‘disappear’ by being split if the number of cluster is too high). This cluster, then, tells us which cells behave the same way regarding the used features and how (and to what extent) they are different from the other clusters. These others clusters are retained

however, since they do hold a meaning regarding the feature, but being less obviously related to known biological patterns it might be more difficult to draw conclusions about them. Ultimately, we expect univariate clustering to act as a mixture model, separating the used feature values into non-overlapping clusters.

Regarding multi-variate clustering, the separation of the used feature into non-overlapping clusters is not always expected since the pairwise distance matrix will be a weighted linear combination of those features. However I would expect to enhance or strengthen the definition of known patterns and also to obtain new ones (not obtainable from univariate clustering) since I will try to relate biologically linked features a biological meaning.

Finally, we define high spatial coherence as spatially well separated clusters, and inversely low spatial coherence as a random distribution of the cluster across the flower meristem. In the same way, high temporal coherence means that the cells follow coherent states changes through their lineage, as opposed to random and mixed state changes.

Ward 4-cluster analysis on ‘epidermis area’ feature starts with the visualisation of the obtained clusters onto the segmented time series given in figure 3.31. This figure shows that it is possible to detect a cluster (3, orange) that is close to the boundary cluster on t_4 , although some cells belonging to this cluster also appear on earlier time points, making it non-exclusive to the boundary representation. Nevertheless, as shown by the cluster distance heatmap (t_0 bottom series) this cluster is well defined since presenting the lowest within-cluster distance. Moreover, looking at the feature distribution boxplot (figure 3.32b) we get back the expected property that boundary have smaller cells and therefore smaller epidermis surface, if under a slower growth rate as observed when manually defining regions (figure 3.28).

However, one could note a somewhat random disposition of the clusters that are not cleanly separated leading to the idea of an overall poor spatial coherence of this clustering, including the “boundary-like” cluster. Even without a precise knowledge of the lineage we also note a low temporal coherence, mostly for cluster 0 (dark blue) and 1 (light blue) where cells seems to alternate between these two states. This low temporal coherence could be linked to the low spatial coherence detected earlier. Although noting that these two clusters cover the average values of the feature (figure 3.32 b) and represent many cells (figure 3.32a) one could argue that both clusters are in fact a single undifferentiated state. Keeping in mind that cell epidermis area is dependant on the cell cycle, we may hypothesise that this cyclic state changes of cluster 0 into 1 is the reflection of

the cell cycle, where a small cell (stage 0) grow (up to stage 1) then divide into small cells again (going back to stage 0). Therefore the apparent low temporal coherence might be a false impression since the other two clusters, 2 and 3, are quite coherent (apart from the earlier time points where their presence is limited) despite the total lack of temporal information. It would therefore be interesting in the case of a multi-factorial analysis if topological information or other features could help enhance the spatial and temporal coherence.

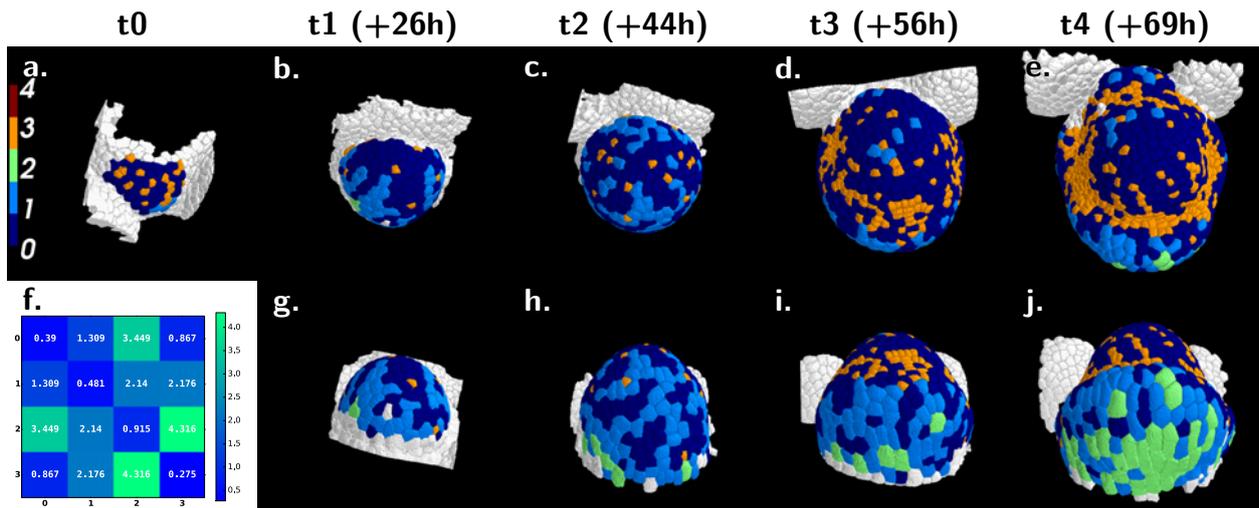


Figure 3.31: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 'epidermis area' feature (a-e,g-j), and the cluster distances associated heatmap (f).

All illustrations similar to those presented in figures 3.31 and 3.32 are part of automatically generated features and clusterings reports made with the temporal property graph structure from FM time-series segmented data.

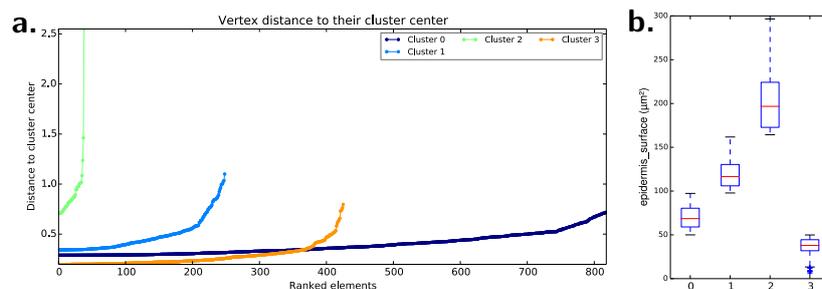


Figure 3.32: Ordered vertex distance to their cluster center (a) and clusters feature distributions (b) obtained using a 4-cluster Ward model based on 'epidermis area' features.

These previous biological conclusions are ascertained by the good quality of the clustering. Indeed in the clustering heatmap, we observe a low within-cluster distance (diagonal) and a high between-cluster distances (figure 3.32f). One caveat here could be that the within-cluster distance of cluster 2 is higher than the distance between clusters 0 and 3. Although observing figures 3.32 a and b, I note that cluster 2 present the largest distribution (fig. 3.32 b), even larger than the sum of both clusters 0 and 3 (ranging on 100 units against 150 units for cluster 3). As figure 3.32 a

highlights, cluster 2 contains three outliers represented by the major increase of the distance for its last three points (around 2.5 fold), artificially enlarging the within-cluster distance as represented. These considerations lead us to conclude to a valid clustering.

Ward 4-cluster analysis on ‘volumetric log temporal rate’ feature at t_{n+1} displays a more spatially coherent (well separated groups) clustering than the one previously made in the same conditions with the feature ‘epidermis area’. This clustering reveals the boundary (cluster 3 orange) with the lowest volumetric log temporal rate (fig. 3.34 b.) and the emerging abaxial sepal (cluster 2 green) with the highest rate (fig. 3.34 b.). This boundary clustering is better defined than with the use of the epidermis area feature, suggesting a higher control of local growth rate than the epidermis area, the later being subject to fluctuations linked to cell growth cycles. Again, two clusters are detected in the intermediary growth rate values. Cluster 1 (light blue) has the higher rate of those two and is detected at t1 and might be linked to the emergence of the dome shaped aspect of the FM. Later this same cluster is present at the emergence positions of abaxial and adaxial sepals (t2) and then lateral ones (t3). Finally, as suggested in figure 3.34a, by the presence of cell with a null distance to the cluster center, the 5-cluster model separates cells without temporally differentiated values (t0 cells and others without ancestors).

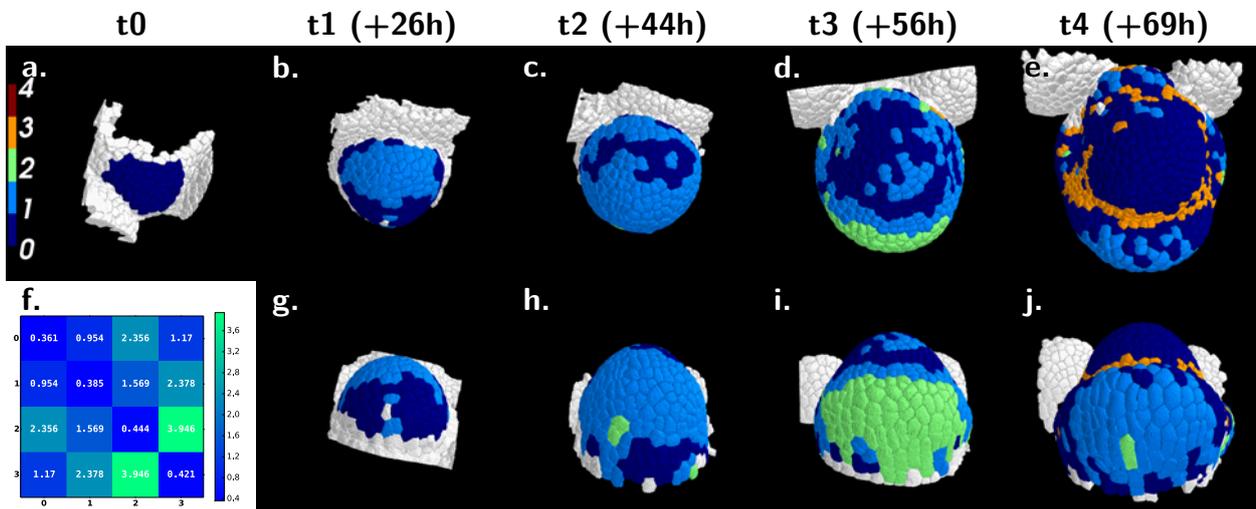


Figure 3.33: Spatial projection of the clustering obtained using a 4-cluster Ward model based on ‘volumetric log temporal rate’ feature at t_{n+1} (a-e,g-j), and the cluster distances associated heatmap (f).

Having demonstrated the capacity of univariate clustering to identify and define biologically coherent patterns on the basis of the extracted geometrical features, we now address the same questions for a multi-variate clustering, with the aim of shedding light on the relations between the

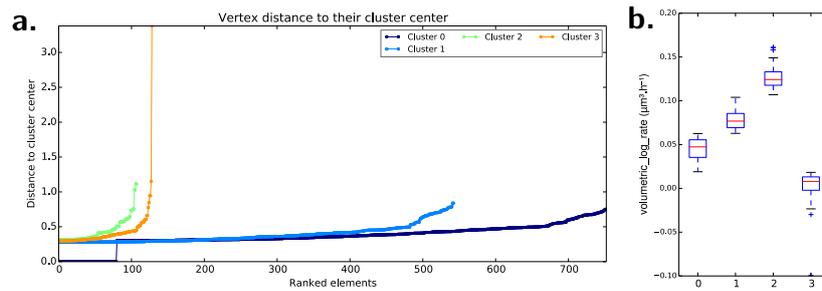


Figure 3.34: Ordered vertex distance to their cluster center (a) and clusters feature distributions (b) obtained using a 4-cluster Ward model based on ‘volumetric log rate’ features.

combined features.

3.2.3 Multi-variate clusterings analysis.

Ward 4-cluster analysis on combined ‘epidermis area’ and ‘topology’ features were realised with increasing weights of topological information (10%, 20%, 30% and 40%); see supplementary figures S3.7, S3.8, S3.9 and S3.10.

Surprisingly, adding 10% of topological information improves neither spatial nor temporal coherence compared to univariate clustering. Almost no effect of adding 10% of topology is detected. However, from 20% and above, the topological information modify dramatically modifies the obtained clustering, leading to a high spatial coherence and a virtual temporal coherence. We can hypothesise that the spatial coherence is somewhat artificially imposed by the low number of clusters and the absence of topology between successive time points, constraining the definition of a very limited number of cluster per time point. This effect can be seen when increasing the percentage of topological information (supplementary figures S3.8, S3.9 and S3.10).

Finally, increasing the number of cluster did not solved the problem (data not shownn). Indeed as hypothesised before, the topological constrain is to important and it result in over-definition of clusters since clusters with similar epidermis area distribution are found repetitively at different times points.

Ward 4-cluster analysis on combined ‘volumetric log temporal rate’ and ‘topology’ features were also realised with increasing weights of topological information (10%, 20%, 30% and 40%); see supplementary figures S3.11, S3.12 (same as fig. 3.35), S3.13 and S3.14. Contrary to what was observed when adding a limited amount (up to 20%) of topological information to epidermis area, here we observe a structuring effect of the topology without disrupting the structure observed with the univariate clustering, brought about by the volumetric growth rate

feature. Using the weighted combination of 80% volumetric rate and 20% topological information, we enhance the previously detected effect with only volumetric growth rate information (figure 3.35): a sepal-to-dome boundary (cluster 1, light blue) with a low growth rate at t4, a high growth rate for the emerging abaxial sepal (cluster 2, green) at t3 and a higher rate for one of the two undifferentiated states (cluster 3, orange) preceding dome and sepal emergence.

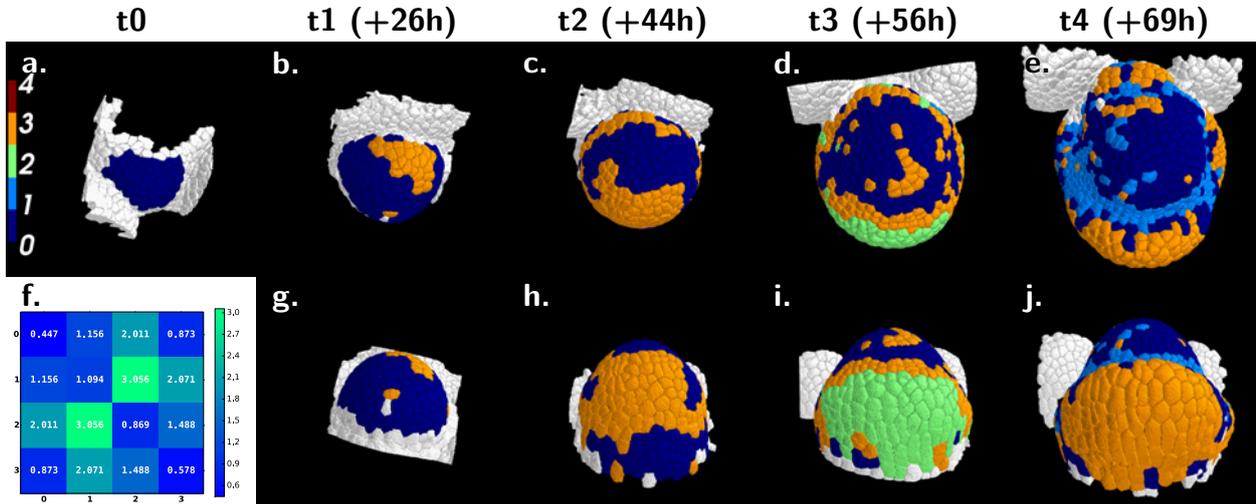


Figure 3.35: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 80% 'volumetric log temporal rate' + 20% 'topology' features (a-e,g-j), and the cluster distances associated heatmap (f).

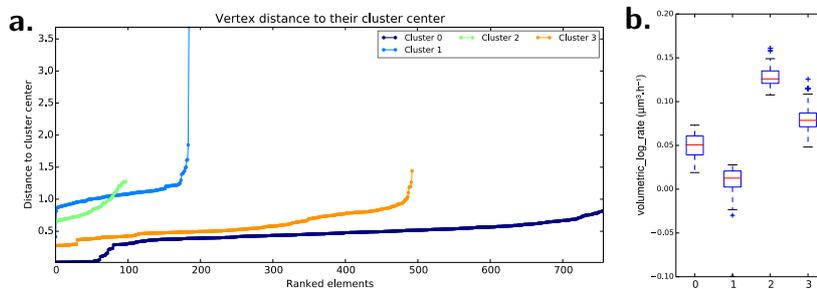


Figure 3.36: Ordered vertex distance to their cluster center (a) and clusters feature distributions (b) obtained using a 4-cluster Ward model based 80% 'volumetric log temporal rate' + 20% 'topology' features.

Again increasing the number of clusters to 5 separates the low growth rate and no-growth rate (cells without ancestors) as may be predicted from figure 3.36a, since cluster 0 shows a gap between cells having low temporal rate values and those that do not have temporally differentiated values. Figure 3.36b illustrate the definition of good clusters whose distributions are well separated.

Finally, as shown in supplementary figures S3.13 and S3.14, from 30% of topological information and above, we lose the structure of the growth rate information, switching again to a highly constrained spatial distribution of the cells among clusters.

Ward 4-cluster analysis on combined ‘volumetric log temporal rate’ and ‘Gaussian curvature’ features was made to compare a growth scalar with a shape descriptor. The Gaussian curvature of the epidermis, is computed using a radius (here of 70 voxels) which make this descriptor an interesting one to describe how the growth have impacted the shape of the tissue, both locally and globally. Indeed, the curvature is defined at cellular scale, but it is the integrated result of tensions, themselves imposed by local changes of growth rates, that act at a much larger scale. I thus wondered if it was possible to find common groups for both volumetric growth rates and Gaussian curvature, and what properties they could have. To that end, we present in figure 3.37, a balanced (50/50) 4-cluster Ward clustering made using volumetric log growth rate and a Gaussian curvature estimated with a radius of 70 voxels (around the geometric medians of the epidermal wall of each cell).

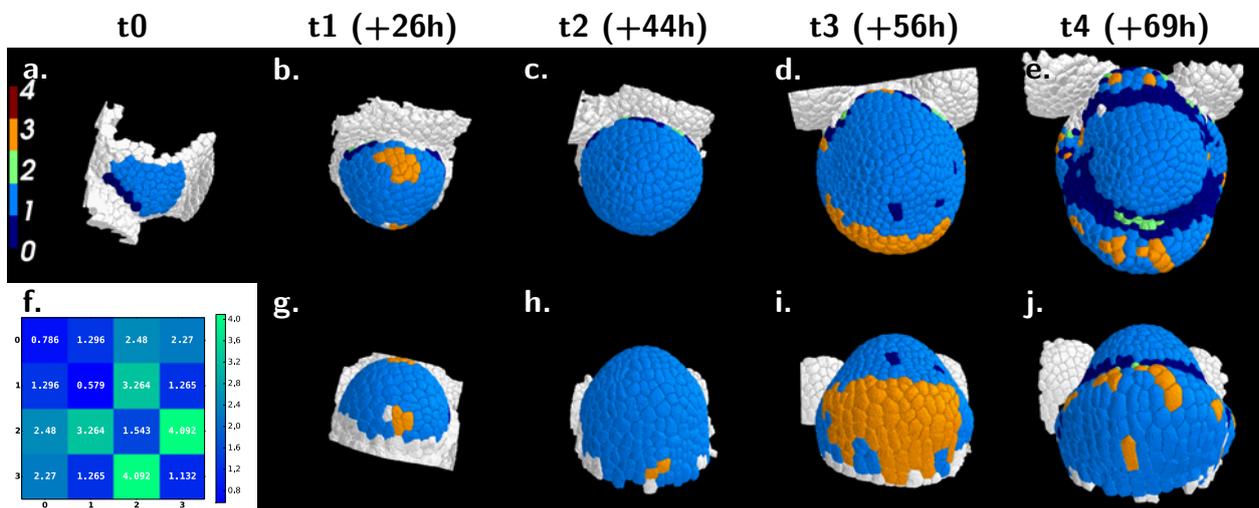


Figure 3.37: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 50% ‘volumetric log temporal rate’ + 50% ‘Gaussian curvature’ features (a-e,g-j), and the cluster distances associated heatmap (f).

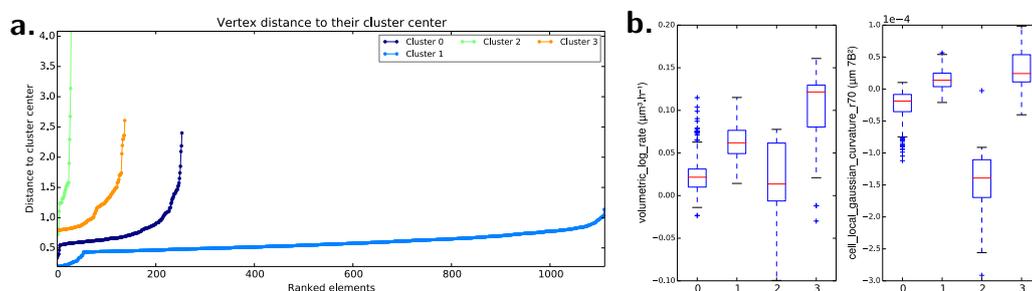


Figure 3.38: Ordered vertex distance to their cluster center (a) and clusters feature distributions (b) obtained using a 4-cluster Ward model based on 50% ‘volumetric log temporal rate’ + 50% ‘Gaussian curvature’ features.

To validate this clustering, let’s first look at the distance matrix heatmap in figure 3.37f. As

we can see, cluster 2 have a large within cluster distance, but that can be explained by the large dispersion of the two features for this group. In addition, this group is relatively small compared to the others (see figure 3.38a). Cluster 3 also present a rather large within-cluster distance, but it is inferior to all other between-clusters distances (off diagonal). I thus declare the clustering as valid.

Now, looking at the spatial projection of the clustering (see figure 3.37 a-e,g-j), we can observe that cluster 0 define the boundaries, at t0 with the other FM growing on its right side (left of the image), at t1 to t3, with the SAM, and at t4 between the central dome and the sepals (mainly abaxial and adaxial). Figure 3.38b confirm that this group present quasi-exclusively cells with a negative curvature, together with a rather low to null growth rate, typical of boundary regions. Cluster 1 seem to define the undifferentiated state, since spatial projections show this zone at every time point associated to no apparent morphological structure. Figure 3.38b also present intermediary distributions for the two selected feature, enforcing the idea of an undifferentiated state. Cluster 2 meaning is not clear while looking up at the spatial projection, but the boxplots (figure 3.38b) highlight that these region is made on the basis of the most negative curvature values, associated to low to null growth values. This cluster might be considered as secondary regarding the displayed property, even if strongly defined by the clustering and present in a 3-cluster clustering (data not shown), since we have reasons to believe that some of the values are outliers. Cluster 3 seems to define emerging dome-like regions when looking at the spatial projections, while looking at the boxplots we confirm this impression noting that this cluster posses the highest distributions for the two features.

Finally, after validating and analysing the results of the clustering, I would like to stress the high coherence displayed by the two features clustered distributions, where low growth values are associated to null or low Gaussian curvature, and high growth values are associated with highly positive curvature.

Multivariate clustering analysis using shape, growth, genetic signal and spatial topology features was finally possible to try to include all the players previously identified. Also, since we are interested in linking genes, growth and shape in a spatially coherent manner, such clustering may prove crucial in validating the method. In that respect, we have combined the following features: ‘volumetric log temporal rate’, ‘Gaussian curvature’, ‘AHP6 signal’ and ‘spatial topology’. Except for the spatial topology which, as we noted before, be used with a low weight and

was thus limited to 10% of the combined information, the other 3 features have the same weight, hence each representing 30% of the combined information. This, time, I have chosen to present the 5-cluster model (figure 3.39), since I find it to be more relevant than the 4-cluster (presented in supplementary figure S3.15).

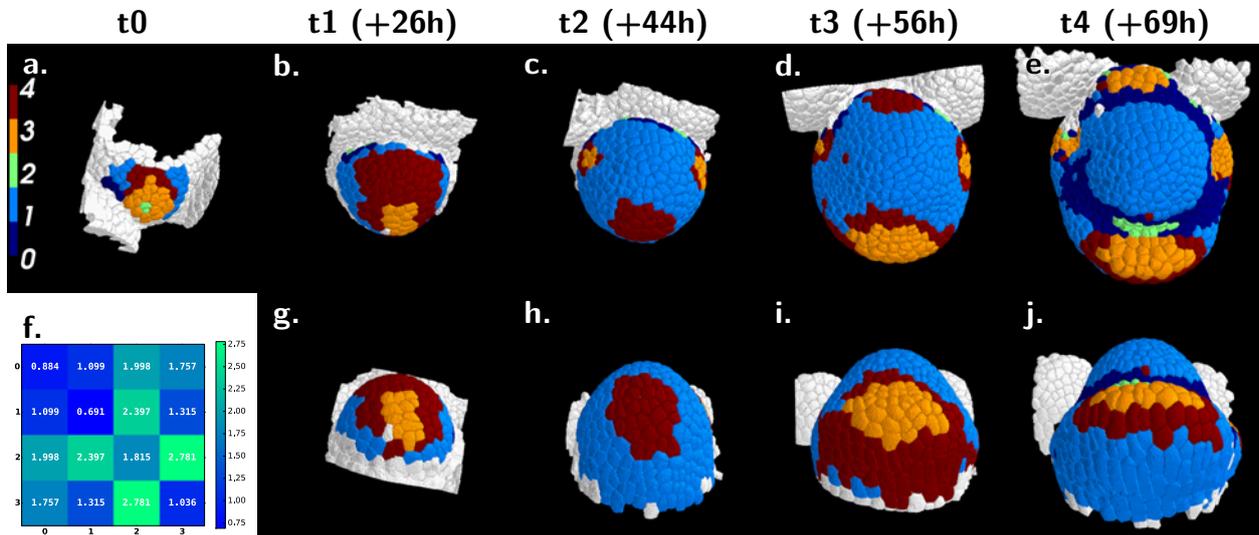


Figure 3.39: Spatial projection of the clustering obtained using a 5-cluster Ward model based on 30% ‘volumetric log temporal rate’ + 30% ‘Gaussian curvature’ + 30% ‘AHP6’ + 10% ‘spatial topology’ features (a-e,g-j), and the cluster distances associated heatmap (f).

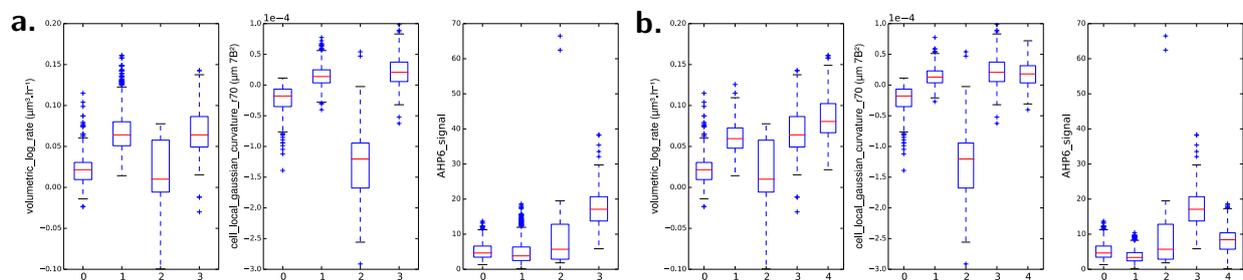


Figure 3.40: Clusters feature distributions comparison for the 4- (a) and 5-cluster (b) Ward model based on 30% ‘volumetric log temporal rate’ + 30% ‘Gaussian curvature’ + 30% ‘AHP6’ + 10% ‘spatial topology’ features. In the 5-cluster model, cluster 1 and 4 have been obtained splitting cluster 1 from the 4-cluster model.

As when using ‘volumetric log temporal rate’ and ‘Gaussian curvature’, we observe that cluster 0 define the boundaries, and for the same reasons as before. For the same reasons than in the previous paragraph, I will consider cluster 1 to remains the undifferentiated state. I will again consider cluster 2 as marginal, since presenting a high intra-cluster distance and a low number of individual, its spatial projections does not seems informative. Also, its spatial proximity with cluster 3 in the spatial projections at t0 and t4 is not coherent with the extreme difference this two clusters displays in terms of features distributions (figure 3.40b). Cluster 3 represent the emerging

dome-like structure, such as the early dome and the sepals, and poses the highest distribution for AHP6 feature. I would like to stress that compared to the previously presented clustering, cluster 1 and 3 now poses a much higher cluster distance related to the addition of AHP6 signal feature. Finally, cluster 4 is often found surrounding cluster 3, and as highlighted by the cluster distance matrix (figure 3.39f) it seems close to cluster 1. Comparing the 5-cluster (figure 3.39) to the 4-cluster (supplementary figure S3.15) configuration, we observe that adding a fifth cluster result in the separation of cluster 1 into cluster 1 and 4. Looking at the cluster distance matrix (figure 3.39f) it seems that this cluster is artificial since the obtain distance between clusters 1 and 4 is low ($D_{1,4} = 0.87$), and lower than cluster 3 within-cluster distance (also much lower than cluster 2 within-cluster distance but which is considered as marginal). However looking at the boxplots esenting the feature distributions obtained for the 4- (figure 3.40a) and 5-cluster (figure 3.40b) configurations, adding the fifth cluster create correctly separated distributions for the volumetric growth and AHP6 features. I thus consider this fifth cluster meaningful.

Focussing on the features distribution (figure 3.40b) it is possible to, mathematically and biologically, characterise the identified growth patterns as follows:

- cluster 0 can be identified as the boundary, characterised by a low to null volumetric growth, associated with negative curvature and a rather low AHP6 activity;
- cluster 1 can be identified with an undifferentiated region, characterised by a quasi-absence of AHP- activity and an average growth rate and low positive curvature;
- cluster 2 cannot be biologically identified since regroups mostly outliers for the shape and growth features;
- cluster 3 can be identified as emerging dome-like structures, characterised by high AHP6 activity, relatively high growth rates and positive curvature;
- cluster 4 can be identified as the fast-growing zone around the emerging structure, surprisingly characterised by slightly higher growth rate than cluster 3 but a much lower activity of AHP6;

3.3 Cellular pattern identification on the basis of Markov tree analysis of the lineage and geometrical features

The algorithmic clustering methods we applied focus mainly on the spatial structuring, and even if we accounted for temporal information using temporally-differentiated variables such as volumetric

growth, these methods do not explicitly model temporal relationships. A shortcoming of algorithmic clustering compared to model-based clustering is the lack of inferential procedures with a firm mathematical basis to select the number of clusters and to characterize the latent space (corresponding in our context to the possible assignments of cells to clusters). Model-based clustering relies on mixture models and these models have been generalized to various structures (sequences, tree-graphs or images) leading in particular to diverse families of hidden Markov models. We chose to investigate a specific family of hidden Markov models for tree data in order to characterize the cell lineages. These models that focus on the cell division process are thus complementary from the algorithmic clustering previously presented that focus on cell neighbourhood. We thus expected in this way to gain more information about the temporal processes taking place during early stages of floral development.

3.3.1 Definition of hidden multi-type branching processes

In order to investigate cell lineage, we built a hidden multi-type branching process. This is a two-level model where at the first level, the cell division process is represented by a multi-type branching process. A multi-type branching process is defined by two subsets of parameters:

1. Initial distribution to model which is the cell identity at the root of a lineage tree;
2. Generation distributions to model the cell identities of the children of a cell of a given identity along a lineage tree. Each generation probability is the joint probability of having n_0 children in state 0, n_1 children in state 1 ... given the parent state $P(N_0 = n_0, N_1 = n_1, \dots | S_u = j)$, simplified in $\Gamma_j(n_0, n_1, \dots)$, such that $\sum_i n_i = 2$ in our case of a binary tree

Hence the types or states of the branching process represent the cell identities. The cell identities are assumed to be not directly observable but only indirectly through different features (e.g. dimension such as the cell volume, shape descriptor such as the anisotropy, hormone signal...). A hidden multi-type branching process adds a third subset of parameters to the two subsets of parameters of the non-observable branching process:

3. Observations distributions to model a descriptor for a given cell identity.

An observation distribution was thus estimated for each state and each descriptor. We chose as possible observation distributions Gaussian and Gamma distributions. In our case where several features were observed for each cell, the observed variables were assumed to be conditionally independent (given the state).

A multi-type branching process can also be viewed as a Markov unordered out-tree model (a finite state process defined on a tree without children ordering) that models the parent/children local dependencies including direct dependencies between children. Hidden Markov unordered out-tree models are thus direct generalizations of the hidden Markov out-tree models with conditionally independent children discussed in Durand *et al.* [2004]. The maximum likelihood estimation of the parameters of a hidden multi-type branching process requires an iterative optimization technique, which is an application of the expectation-maximization (EM) algorithm. Once a hidden multi-type branching process has been estimated, the most probable lineage tree labelled in terms of cell identities with its associated posterior probability can be computed for each observed multivariate lineage tree using a restoration algorithm. The EM and the restoration algorithms are direct generalizations of the similar algorithms proposed by Durand *et al.* [2004] in the context of hidden Markov out-tree models with conditionally independent children.

Although the selection of the number of states and the initialisation of the iterative estimation algorithm of such model require some expertise, we have been able to use them to characterise the temporal relations of the 3D+t data. Contrarily to clustering methods which are spatial by nature, we here consider the lineage tree as the observed structure, and use the cell features as multivariate data to restore unobservable cell identities. Obviously, using temporally differentiated features would not be a relevant idea, since the model relies on temporal dependencies. Finally, even if the only spatial structuring taken into account corresponds to siblings with respect to a given parent cell, this minimal spatial information is particularly relevant for plants where cells cannot migrate.

3.3.2 Multivariate hidden multi-type branching process model outputs.

Figure 3.41 represents the spatial projection of the states restored for each cell using the estimated hidden multi-type branching process.

This figure indicates that:

- the central dome has been assigned to state 2 which presents a high spatio-temporal coherence from t_0 to t_4 ;
- the sepals are have been split in 2 states (0 and 3);
- the boundary zone has been assigned to state 1.

State 2 is the main state of the first time point. Despite an early stage of meristem differentiation at t_0 , few cells are already identified to putative sepals. At subsequent time points, the multiplication

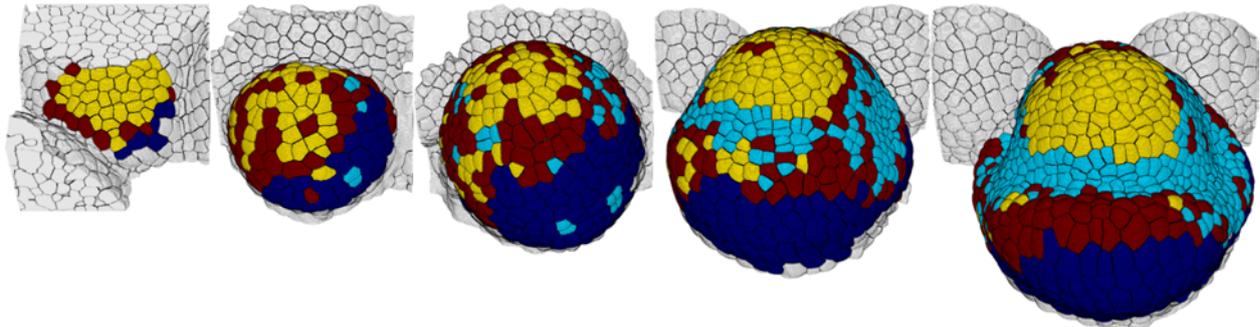


Figure 3.41: Spatial projection of the four states obtained using a multivariate hidden multi-type branching process computed using the epidermis area, the internal area (i.e. between the L1 and L2), the volumes, the Gaussian curvature, the shape anisotropy and the AHP6 signal. State 0 is in dark blue, state 1 in light blue, state 2 in yellow and state 3 in dark red. Sepals are mostly identified by considering state 0 and 3, the dome by state 2 and boundary cells by state 1.

of sepal and dome cells with the apparition of boundary cells, delimiting the frontier between sepals and the dome zone, is observed. The apparition of boundary cells is unobtrusive in t_1 and t_2 but significant as the continuous border is clearly identified starting from t_3 . In fact, at this time point and the next, state 2 clearly correspond to the dome, state 0 and 3 to the sepals and state 1 to the boundary.

The model has been estimated using the following geometrical features: the epidermis area, the internal area (i.e. between the L1 and L2), the volumes, the principal and secondary curvatures, the shape anisotropies and the AHP6 signal. As for the clustering, we expect non-structuring features not to alter the result since the method is robust to them. Using figure 3.42, it is quite clear that the epidermal area, internal area, the volume and the curvatures of the cells are structuring features in the model, since the estimated observation distributions for the different states are well separated for this five features. These observation distributions allow us to characterize the different states:

- state 1 correspond to small cells with both curvatures almost of the same norm and mostly negatives (saddle form) which is coherent with boundary cells;
- states 0 and 3 correspond to big cells and are mostly differentiated with their curvatures (both positive for state 3 and negative for state 0);
- state 2 is in-between considering size but with clearly positive curvatures corresponding to a dome area.

States do not have marked differences with respect to anisotropies and AHP6 signal (data not shown).

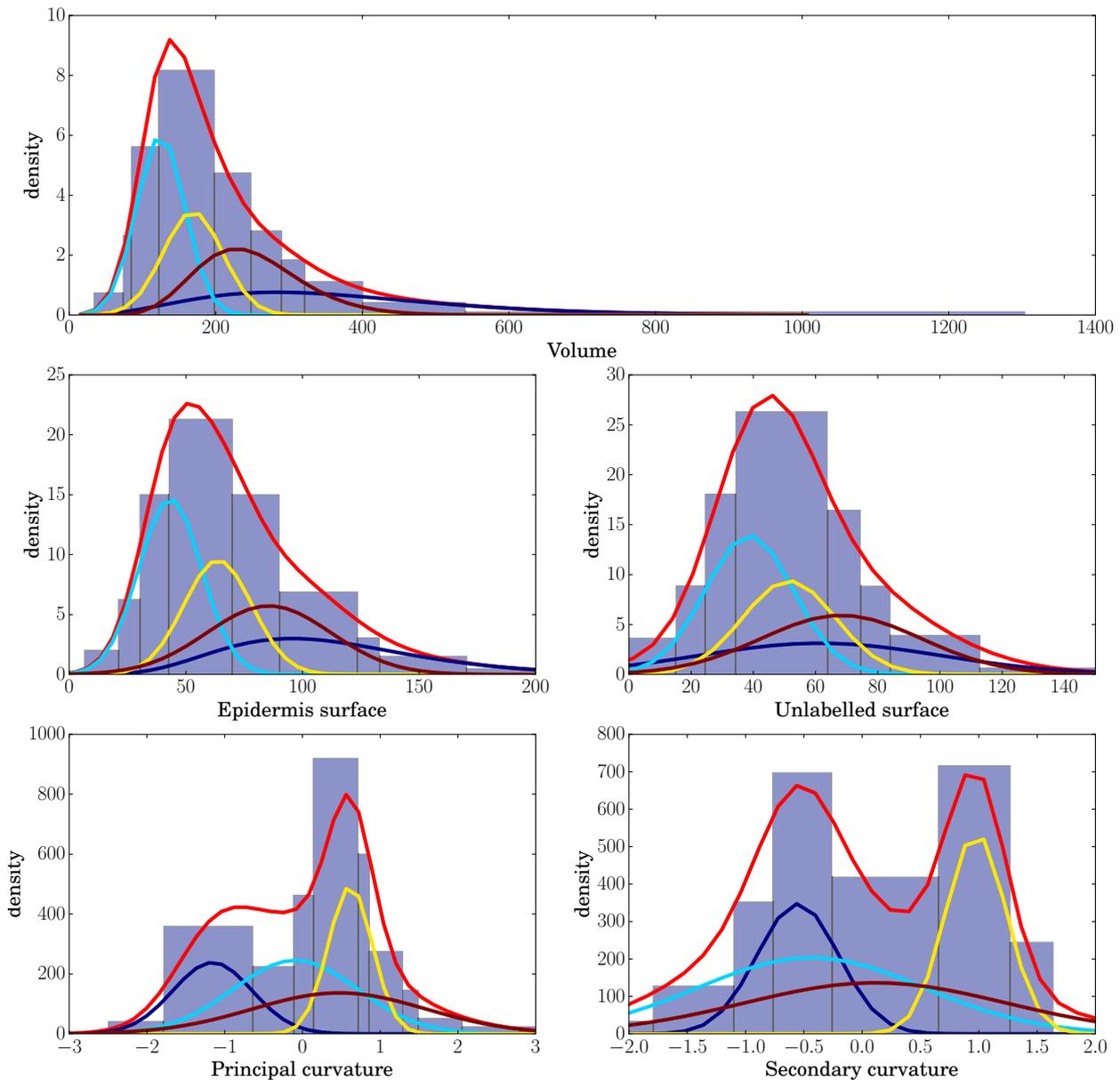


Figure 3.42: Observed histogram and mixture of observed distributions for each structuring feature. State 0 is in dark blue, state 1 in light blue, state 2 in yellow and state 3 in dark red (same than in the spatial projections). Areas and volumes have been modelled by Gamma distributions and curvatures by Gaussian distribution. Combining separations induced by areas and volume in one hand and curvatures in the other hand indicates that states are well separated using only these features.

Recall that $\Gamma_j(n_0, n_1, n_2, n_3)$ denotes the probability of having jointly n_k children in state k for $k \in \{0, 1, 2, 3\}$, considering a parent cell in state j . Temporal changes in cell identity are highlighted by the highest estimated probabilities in each generation distribution:

A graphical representation of the generation laws is provided in figure 3.43. The reproduction and emergence of cell identities underlined by generation distributions are consistent with biological beliefs. The state 3 is a hub as transition from the state 2 at t_0 to other states at subsequent time points. Transition from state 0 to state 3 corresponds to transition from early cells to late ones

State 0	State 1	State 2	State 3
$\Gamma_0(0, 0, 0, 2)=0.05$	$\Gamma_1(0, 0, 0, 1)=0.12$	$\Gamma_2(0, 0, 0, 1)=0.17$	$\Gamma_3(0, 0, 0, 2)=0.17$
$\Gamma_0(1, 0, 0, 0)=0.33$	$\Gamma_1(0, 1, 0, 0)=0.52$	$\Gamma_2(0, 0, 0, 2)=0.08$	$\Gamma_3(0, 0, 1, 1)=0.13$
$\Gamma_0(1, 0, 0, 1)=0.06$	$\Gamma_1(0, 2, 0, 0)=0.35$	$\Gamma_2(0, 0, 1, 0)=0.2$	$\Gamma_3(0, 0, 2, 0)=0.05$
$\Gamma_0(2, 0, 0, 0)=0.56$		$\Gamma_2(0, 0, 1, 1)=0.17$	$\Gamma_3(0, 1, 0, 1)=0.22$
		$\Gamma_2(0, 0, 2, 0)=0.33$	$\Gamma_3(0, 2, 0, 0)=0.32$
			$\Gamma_3(2, 0, 0, 0)=0.03$

in sepals. Transition from state 3 to state 1 corresponds to emergence of border cells induced by sepal formation, it is a passive phenomena more than an active one.

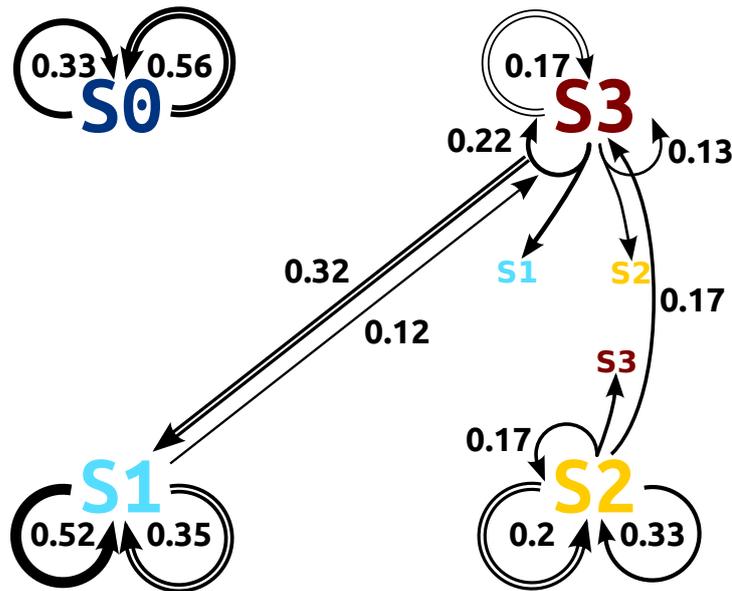


Figure 3.43: Representation of the generation distributions by states. Generations distribution relates to the evolution to the cell state in time, meaning the state of its descendants (either binary or no division). For graph readability reasons, only generation laws with a probability higher than 0.1 were represented. A single line indicates generation law of undivided cell when double lines and two parts arrows indicate a binary division. The thickness of the line is proportional to its associated generation probability, which are also given as plain numbers. The state colors are the same than in the spatial projection and the distribution histograms.

4 Discussion and perspectives

The attributed spatio-temporal graph.

The first aim of this paper was to introduce an effective representation of developing tissue that has led to the cell-based spatio-temporal graph. Extracting the cells neighbourhood information, first permits the definition of a spatial graph (vertices and spatial edges), then using the lineage (temporal edges) we obtain a spatio-temporal graph. This formal definition of the graph thus corresponds to the natural decomposition of a developing tissue into its basic constituents. It was thus an important step as it lead us to an adequate *in silico* representation of the tissue.

The subsequent task was then to search for adequate geometric descriptors for the cells, as well as spatial and temporal differentiation functions, robust landmark definition and tensor transformation functions. We have proposed a deformation estimator robust to segmentation errors using cell wall medians as landmarks, instead of the traditional use of cell vertices. Introducing the spatial and temporal differentiation functions allow us to expand the number of features and to compute spatial variations (Laplacian) or temporal variations (growth rates). Again the graph structure allows an efficient programming of these differentiation functions. We stress that the framework is independent of the segmentation and lineage algorithm, but is currently limited to the analysis of voxel-based segmented images.

Finally, this generic formalisation and implementation of the attributed spatio-temporal graph constitutes a powerful database, which architecture is based on the natural decomposition of the segmented tissue, allows to efficiently organise the extracted features. This is indeed another key aspect of the presented work, to propose a large number of 2D and 3D geometrical cell features. Overall it is an important contribution to morphogenesis quantification, since previous works in the field were based on *ad hoc* methods made either in 2D, or surfacic 3D, therefore proposing 2D description of the shape (e.g. shape and strain anisotropy).

Regarding that aspect of the work, the spatio-temporal graph idea in itself, I do not see much space for improvements, although its python implementation and the proposed functions can still be improved. New differentiation functions could be added as well as cellular features. Ultimately, I think that a graphical interface allowing to generate the spatio-temporal graph from the cell-segmented images and to display the computed information would be important. Although an intermediary solution could be to compute automatically (as it is already possible) all fea-

tures, and then to make use of existing 3D visualisation software such as BioImageXD. Using the spatio-temporal as a database within such software, it should then be possible to display and navigate seamlessly among the extracted cell features, clustering or other information. Talking about database, I argue here that it should be of the uttermost importance to make use of one to organise, version control and save the whole process, from image acquisition to segmentation and analysis. Indeed, the whole pipeline starts with the images acquisition, followed by their segmentation, two step presenting an incredible number of parameters that should be kept. In case of successive corrections of the lineage or use of various parameters to segment, these information would then be available, and for instance would help preventing mixed-up analysis. Finally, facing many caveats and limits using the MARS-ALT software, and its development currently being on hold, it would be interesting to test the segmentation tools proposed by ITK and integrated in BioImageXD.

Exploratory analysis to uncover links between genes, growth and shape.

In a second phase I demonstrate on the basis of live-3D segmented data of developing floral meristem, the advantages of relying on attributed spatio-temporal graph structures. Using the introduced framework, it is possible to compare and differentiates many features, thus proposing an efficient exploratory analysis of the data. This *a priori* approach of the data is facilitated by 3D representation of the segmented tissue as well as diverse dedicated graphical representation of the data given their nature (property plots). In addition, unsupervised methods allows to define cellular patterns and identities on the basis of geometric and genetic features. Altogether this provide a clear picture of the data for a biologist to confront his knowledge and searches for hypothesised behaviours of the tissue.

The data I presented here, are cell-segmented images of a particular morphological event, that is the organogenesis of the flower of *A. thaliana*. The floral development see a growing tissue taking shape and we know that mutations in major developmental genes, such as *LFY* or *AG*, can alter dramatically the aspect of the flower. In that respect, I have tried to demonstrate links between genes, growth and shape.

I first started by using the attributed spatio-temporal graph to test the “universal rule for symmetric division” as introduced in Besson and Dumais [2011] with 3D+t data and highlighted a first relation between shape and growth. Since the main vector is oriented along the maximum length of the cell, the two others thus form an orthogonal plane close to the barycentre and the

minimal area, which is consistent with the division plane alignment. Comparing the division plane collinearity with the main axis of inertia has allowed to observe that the large majority of the cells respect Errera's rule (figure 3.25) . Using the inertia tensor obtained from the fused-siblings, we could observe a possible difference of the abaxial sepal that seems to display more exceptions to the rule. In addition, relating the vector collinearity to the sibling volume sums and sibling fused 3D shape anisotropy we could not observe any correlations (figure 3.26). Therefore we can conclude that the orientation of the division plane is made independently of its size or degree of anisotropy.

According to the authors observations and models, the cell nucleus is positioned close to the barycenter of the cell subsequently to the tensional forces within the cytoskeleton reaching an equilibrium, thus dividing a parent cell along a plane that minimizes its area. The shape of the cell, by modifying the tensional force and the barycenter position, thus yield two descendants of identical size. Visualising the spatial projection of the sibling volumetric ratio for all known siblings did not revealed a spatial pattern, however, a temporal pattern could be present (figure 3.21b top row) Again, comparing the sibling volumetric ratio to the sibling volume sums and sibling fused 3D shape anisotropy we could not observe any correlations (figure 3.23), thus concluding to the independence of the volumetric ratio regarding size and shape.

Going further in my *a priori* analyses, I have manually defined cell regions corresponding to observable and known morphological patterns at the last segmented time point (t_4). The four sepals, as well as the boundary separating them from the central and peripheral zones were defined. This yielded geometrical features distributions that were not clearly separated (figure 3.28), but it was to be expected since we defined the regions without considering them. Nevertheless, it was an interesting preliminary analysis since it revealed that the growth related features (volumetric growth rates and areal strain rates) could help discriminate regions potentially close to the manually defined sepals, boundary and central zone. Ultimately, the genetic feature distributions were also not well separated (figure 3.28), but since the manually defined abaxial sepal display high levels of genetic signal and it coincide with the region presenting the maximal growth rate, I thought at correlating a discretised version of the genetic feature to the geometrical ones.

Many other exploratory analysis are possible with the framework and command-line accessible methods. For example, to go further in our understanding of the division wall orientation, it could be interesting to use microtubule impairing-function genes and tissue shape or division plane

orientation. It could also be possible to inspect growth impact on genes by monitoring genes activity while altering the formation of the division wall like in the case of endoreduplication, where the genetic material is duplicated but no division occurs. This phenomenon occurs frequently in the epidermal layer of the sepals during their formation [Roeder *et al.*, 2010] (observable at latter developmental stages than those presented). In that respect, the first presented use of the framework for exploratory analysis purpose, is to be taken as an example of the multiples of the possibilities it offers.

The second use however, when testing the manual definition of the cluster, was a necessary step to gain knowledge about the potential structure among the data, prior to the unsupervised analysis. It also opened perspectives in proving the feasibility of a 3D+t atlas of genes expression domains for the developing flower, or other tissue for that matter. Building this atlas would probably require, or at least push toward, a generic growing floral meristem, that could be achieved in 3D using growing mechanistic models. Later on, using such growing framework with known genes spatio-temporal patterns it would be possible to test their relation to known features or even recreate their interaction networks and make predictions about their functions before doing the experiment.

Unsupervised analysis to identify cellular patterns linking genes, growth and shape.

In a first time, I used univariate clustering methods to explore the behaviour of some cellular features. Using unsupervised methods to groups individuals based on their similarity can indeed be a way to discretise a distribution. Applying a 4-cluster Ward clustering on the AHP6 signal yielded regions that were later characterised as none, low, medium and high signal levels. Using the obtained regions, I have presented the features distribution for each regions and we could see a significant link between increasing levels of AHP6 and increase in volumetric growth rates, areal strain rates and division rate. This confirmed the presence of links between genes activity and local changes in growth related parameters within the 3D+t data, which have been successfully identified and characterised using the proposed framework.

In addition, I have proposed to define cellular patterns using clustering technique based on the geometrical features. Again, we designed dedicated methods to transform the attributed spatio-temporal graph structure, into data suitable for multivariate clustering analysis using standardised distance. However, analysing a single time-series of the early stages of a wild-type developing flower, I could only propose preliminary analyses of the detected biological processes.

Nevertheless, I demonstrated the capacity of the proposed unsupervised method to identify and characterise cellular patterns on the basis of univariate clustering. I was indeed capable of identifying biologically relevant regions as the boundary between the central dome and the sepals of the flowers using both static (epidermis area; see figure 3.31) and dynamic (volumetric log temporal rate; see figure 3.33) features. This was suggested by the explanatory analysis, where a manually defined boundary displayed lower growth rate than epidermis area. In addition the emerging sepals and central dome were also identified as separate regions on the basis of the volumetric log temporal rate. The emerging sepals indeed displayed the highest growth rate, when the central zone an intermediate one. These clusters were both spatially and temporally coherent, and without the explicit use of temporal information.

In a second time, using multivariate clustering, I showed that the spatial topological information could be used to increase spatial coherence of the clusters. Indeed using topological information in addition to the volumetric log rate, I could obtain more coherent clusters (compare figures 3.33 and 3.35). However, this information is mathematically defined only within each time point, and thus giving too much weight to that information could lead to a radical transformation of the clusters found without spatial topology. Then, I presented a balanced (50/50) 4-cluster clustering made on the basis of growth and shape descriptors (figure 3.37). I was able to characterise the clusters as the boundary, emerging dome-like structure, undifferentiated region and an outliers group. I there demonstrated the spatio-temporal coherence of growth and shape feature, as illustrated by the good relation between ‘volumetric log temporal rate’ and ‘Gaussian curvature’ in figure 3.38b. Biologically we have shown that the shape (and thus dome and sepals) emergence is not only subsequent to an accelerated growth rate for the sepals but also present a dramatic reduction for the boundary.

The last multivariate clustering I hereby present, successfully combined four distinct types of features: shape, growth, genetic and spatial topology. Using a 5-cluster Ward clustering on the basis of 30% ‘volumetric log temporal rate’ + 30% ‘Gaussian curvature’ + 30% ‘AHP6’ + 10% ‘spatial topology’, I was indeed capable of characterise 5 mathematically and 4 were biologically relevant. Using the features distribution (figure 3.40b) I could demonstrate that the boundary are characterised by a low volumetric growth, associated with negative curvature and a rather low AHP6 activity. Moreover, the sepals are likely to come from emerging dome-like structures and

they are characterised by high AHP6 activity, relatively high growth rates and positive curvature. The same phenomenon might also explain the transition from morphological stages 1 to 2, where we observe the bulging of the floral primordia. I was also able to identify an undifferentiated region, coherently characterised by a quasi-absence of AHP6 activity and an average growth rate and low positive curvature. A cluster (c2 fig. 3.39) could not be biologically identified, but since it regroups mostly outliers for the shape and growth features I define it as marginal. Lastly I could observe a fast-growing zone around the emerging structure, characterised by slightly higher growth rate than cluster 3 but a much lower activity of AHP6. A broad observation of the spatial representation could have led us to think that the emerging structures (c3 fig. 3.39) have a higher growth rate than its surrounding regions (c4 fig. 3.39). However, observing the growth feature distributions, it's not the case. One way to explain this would be the possible existence of a delay between AHP6 expression and its effect, thus I hypothesise a potential indirect action of AHP6 on growth.

Having demonstrated the capacity of the unsupervised approach to identify and characterise biologically relevant geometric patterns at cellular resolution, we believe that using contrasted situations, e.g. comparing a wild-type flower with a mutant presenting a different phenotype, will make biological conclusions easier. For example the *ettin* mutant displays a fifth sepal usually appearing as a double abaxial sepal. ETTIN is a transcription factor, also known as ARF3, implicated in the signalling pathway of auxin. We have tried to obtain those *ettin* data, however due to difficulties to obtain confocal acquisition suitable for a proper segmentation, I could not realise this analysis. The fluorochrome used to stain the cells' membrane is the FM4-64 vital-dye, however the fluorescent signal is not detected in the same amounts as for the WT tissue. Since the dye is lethal at high dosages it could not be raised, and the watershed segmentation algorithm failed. Although a study by Cutler *et al.* [2000] has identified (among others) a membrane protein called LOW TEMPERATURE INDUCIBLE 6b (LTI6b), which added to a fluorochrome should enable us to acquire those data. Possessing a powerful framework capable to extract numerous features and to identify changes in cellular patterns, we are confident that it will help explaining the origin of this fifth sepal.

Another mutant, *perianthia* (*pan*), also presents a five sepal phenotype. PERIANTHIA is a transcriptional factor necessary in short days conditions to activate AGAMOUS [Maier *et al.*, 2009], a major developmental gene for the flower. It could thus be interesting to compare the features and

patterns display by this mutant to those obtained using *ettin* flowers. There might then be several ways of doing a fifth sepal.

Finally, those data and their analysis will be strengthened by the increasing capacity in acquisition rate and resolution for the experimental aspect and by the development of robust full 3D segmentation and lineaging algorithm in image processing. Moreover, possessing several overlapping time-series will increase the reliability of the conclusions. Although some statistical methods might require the temporal alignment of the different time-series, the clustering will be able to analyse several at once thanks to the standardisation step. For time-series with the same genetical background, it should be possible to obtain a relative ordering of the static images. This should also be robust to changes in the floral meristem growth dynamics, since relying on those feature at a cellular level.

Using hidden multi-type branching processes to characterise temporal evolution of cell states. The previously presented clustering analysis have focused on identifying spatial structure within the pairwise distance matrix to groups cells on this basis. However, except from what is contained within the temporally differentiated variables, no temporal information is used. To overcome this issue, we have used hidden Markov models for tree data in order to characterize the cell lineages. Using the temporal topology as structure and the cell features as hidden variables, the model estimated observation distribution for each state and each descriptor, as well as the generation distributions.

We found that the first observable state (S2) at t_0 is coherently found over time, up to the last time point where it defines the central dome (figure 3.41). Consistently with the clustering analysis, the boundary (S1) is defined by a low volume, areas and negative curvature. Surprisingly however, the generation laws (figure 3.43) indicate that it is established from a state (S3) defined by large volumes and areas, and with a high primary curvature. State 3 thus seems to act as a hub generating the boundary (S1) and central dome (late S2). Finally we observed the sepal state (S0) to be very conservative.

Since the roots of lineage trees are not systematically at time t_0 , it would be relevant to orient the lineage trees from the leaf vertices at the last time point to the roots. This leads to the hidden Markov in-tree model, which is parametrized by transition probabilities from the children

states to the parent state. The transition probability matrix is thus similar to that of a high-order or variable-order Markov chain where, in the context of tree structures, the variable number of child vertices plays the role of the order. This model can be viewed as a hidden coalescence process (branching viewed backward is coalescence). We expect this model to give complementary information, with respect to the hidden multi-type branching process, regarding in particular the cell identities identified at the first time points on the basis of cell identities propagated from the last time points.

Author contributions

The initial floral meristem 3D+t dataset was generated by Pradeep DAS and the segmentation has been done by Vincent MIRABET. I carried out the manual lineage and sub-lineage definition. Together with Yann GUÉDON, I defined the attributed spatio-temporal framework that I subsequently implemented with the help of Frédéric BOUDON. I carried out the search for adapted features formulations, as well as their spatial and temporal differentiation and transformation functions. The standardised multivariate pairwise matrix formulation was done by Yann GUÉDON and I implemented it. I carried out the exploratory analyses as well as the clustering analyses presented here. Pierre FERNIQUE implemented the multi-variate hidden Markov Tree estimation algorithm and carried out the presented analysis with inputs from Yann GUÉDON and myself. I subsequently analysed. Frédéric BOUDON, Pierre FERNIQUE and Yann GUÉDON participated to the writing and I benefited of Arezki BOUDAUD and Pradeep DAS corrections and remarks.

Supplementary Materials

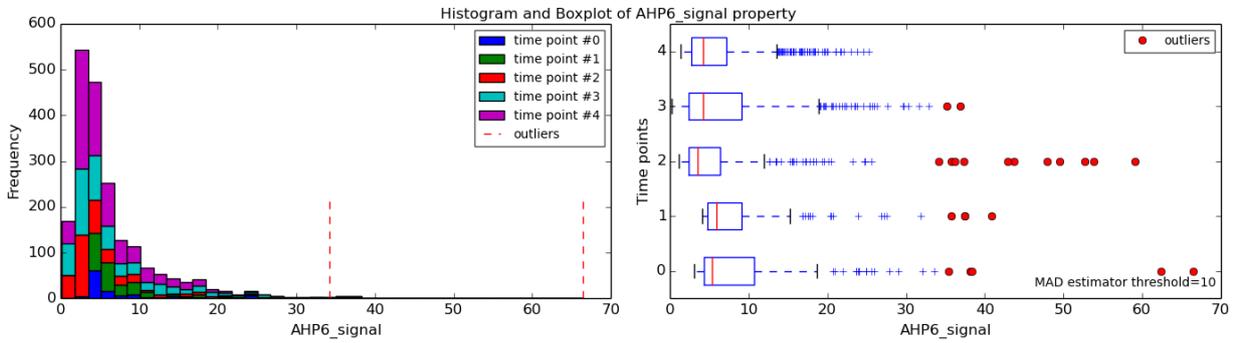


Figure S3.1: Distribution of AHP6 signal by time point. The red dots on the boxplot (right) and upper limits on the histogram (left) give the outliers positions estimated from a *mean absolute deviation* estimator.

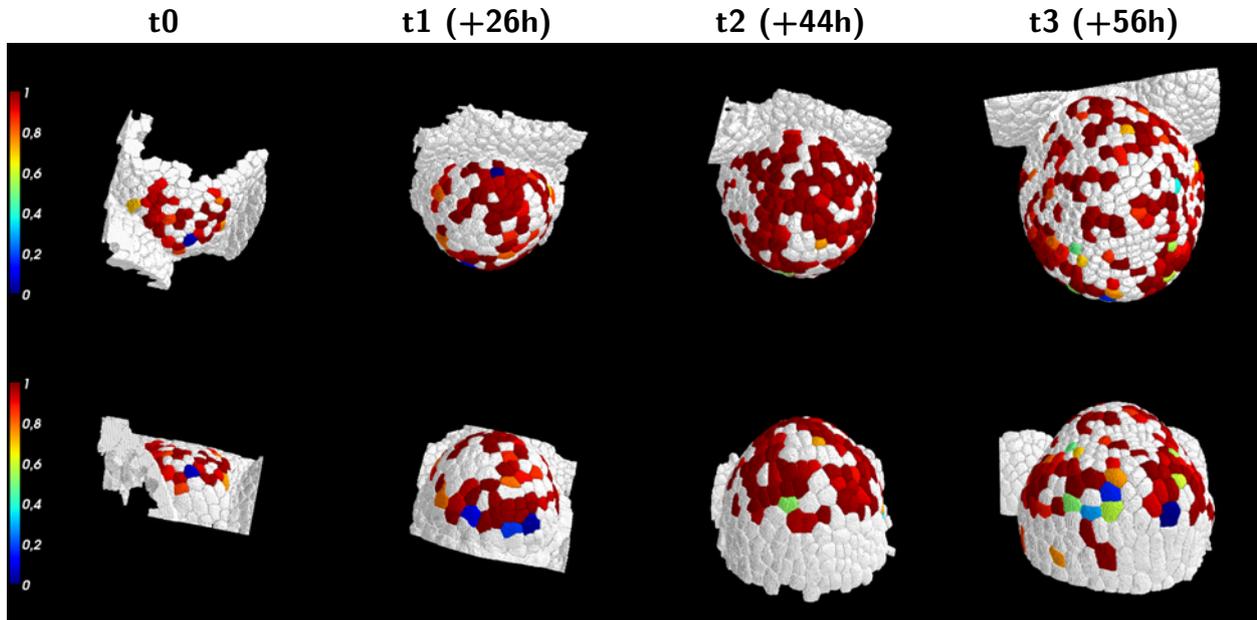


Figure S3.2: Main axis of inertia and division wall normal vectors correlations projected at t_n . Collinear vectors have a value close to 1 and orthogonal vectors close to 0. We use the inertia axis obtained with the fusion of the daughters since the division wall is defined at t_{n+1} . To be coherent with the assumption of a 'small deformation' between successive time-points, we only represented cells presenting a binary division (i.e. only for two siblings).

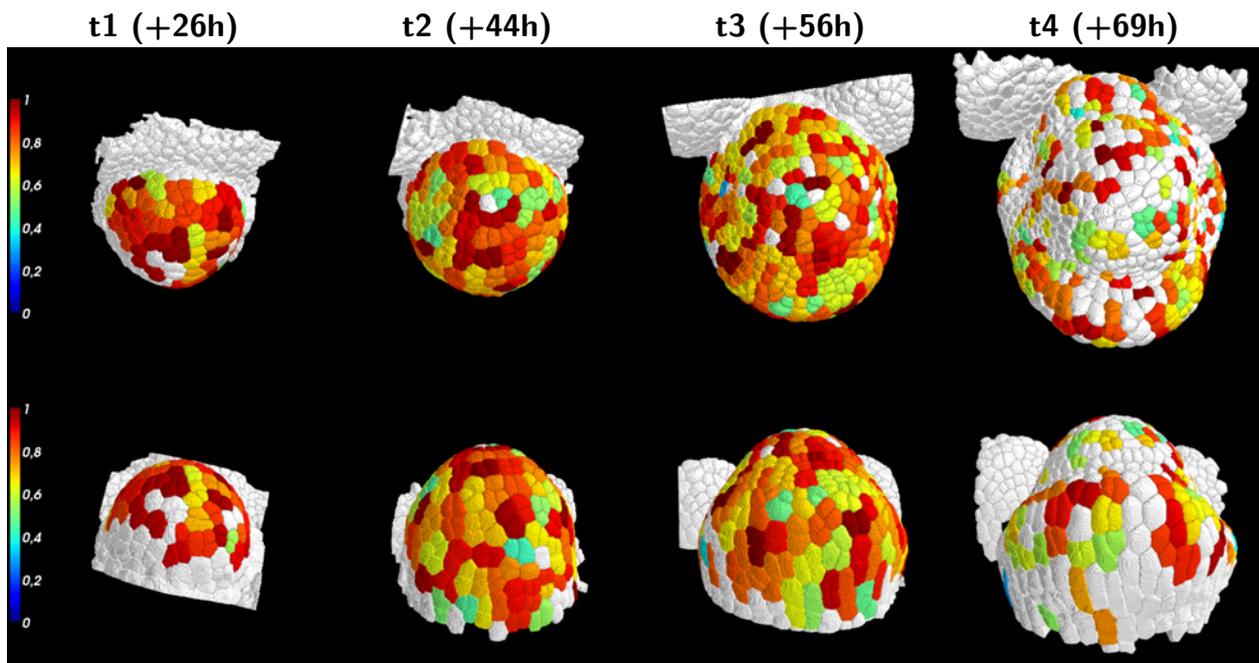


Figure S3.3: Spatial projection (top and abaxial views) of the siblings volume ratio at $t_n + 1$. In order to be thorough we use the sublineage (hand-made) to compute volumetric ratios for all cells presenting a division. The t4 presents a limited amount of data, but this is consistent with the low division number observed between t3 and t4 (see supplementary figure S3.5).

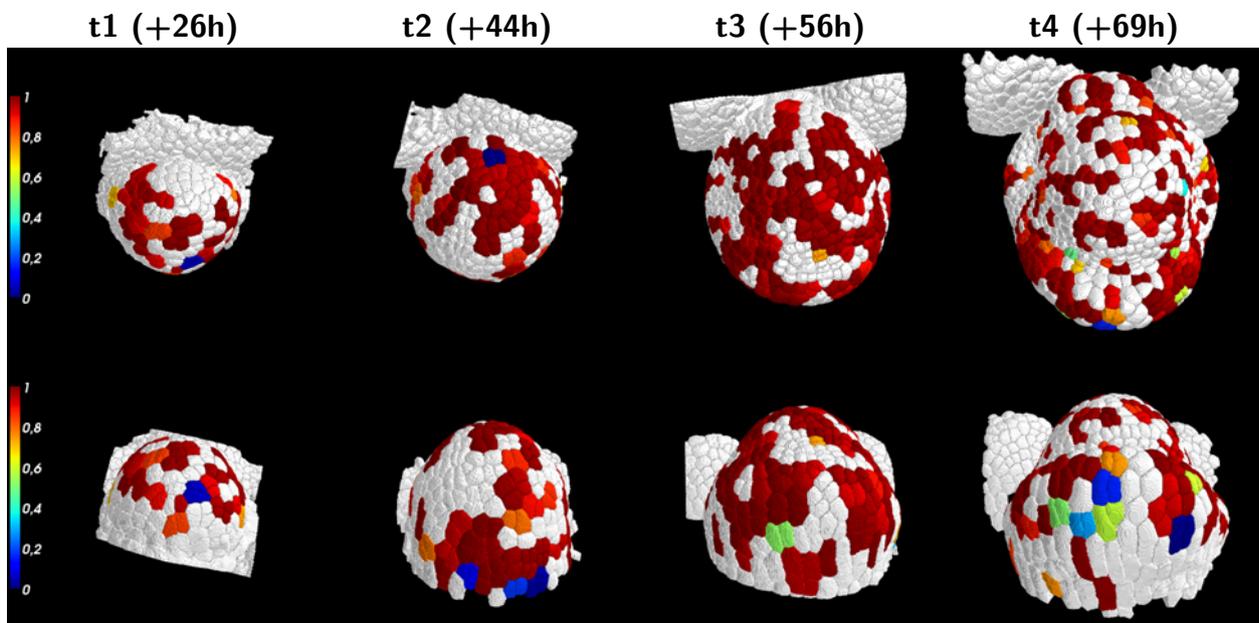


Figure S3.4: Main axis of inertia and division wall normal vectors correlations projected at t_{n+1} . Collinear vectors have a value close to 1 and orthogonal vectors close to 0. We use the inertia axis obtained with the fusion of the daughters since the division wall is defined at t_{n+1} . Only the cells presenting a binary division (i.e. only for two siblings) are presented.

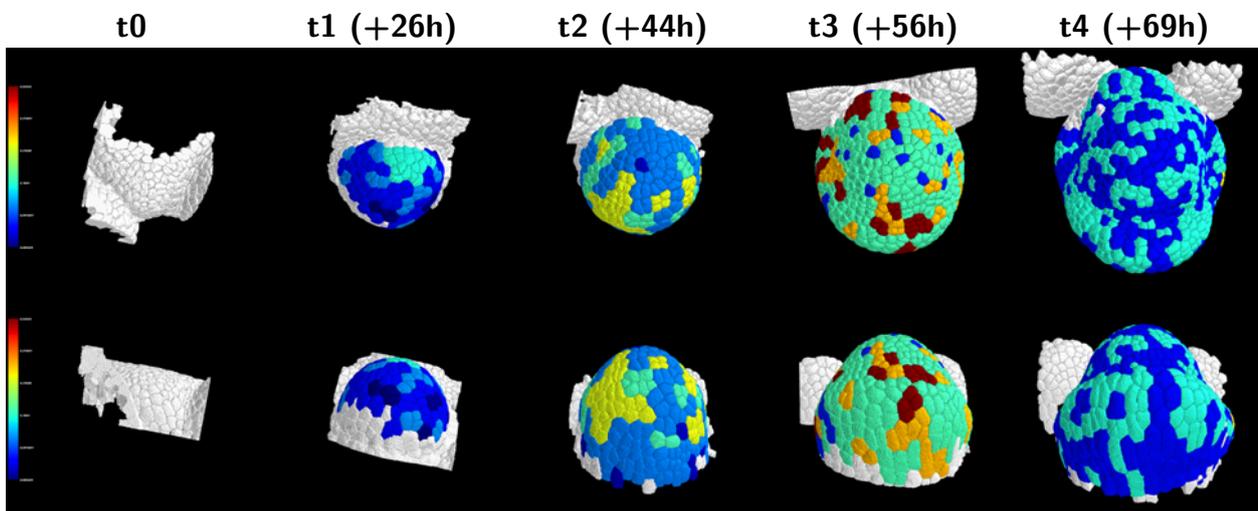
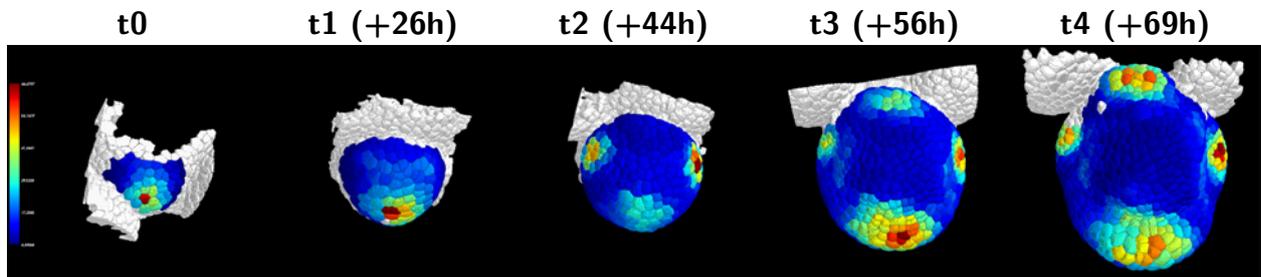
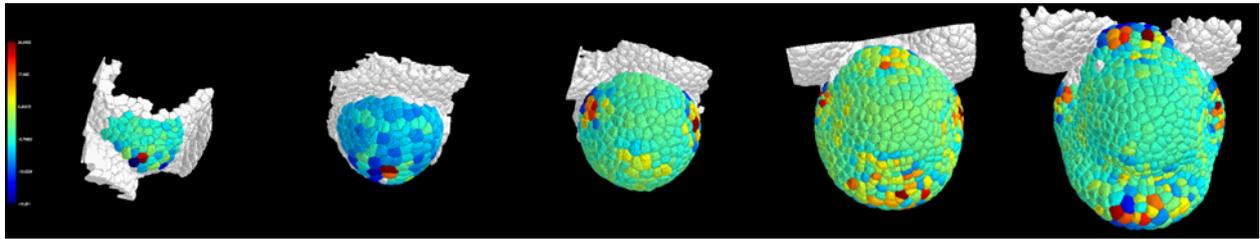


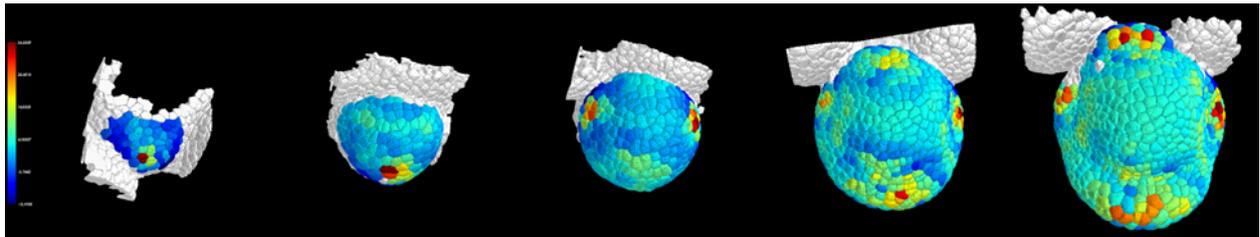
Figure S3.5: Top and abaxial view of the division rates projected at t_{n+1} .



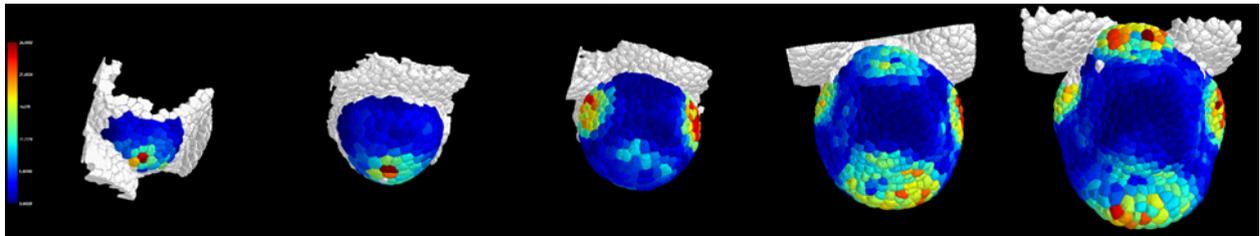
a. Top view of AHP6 signal feature.



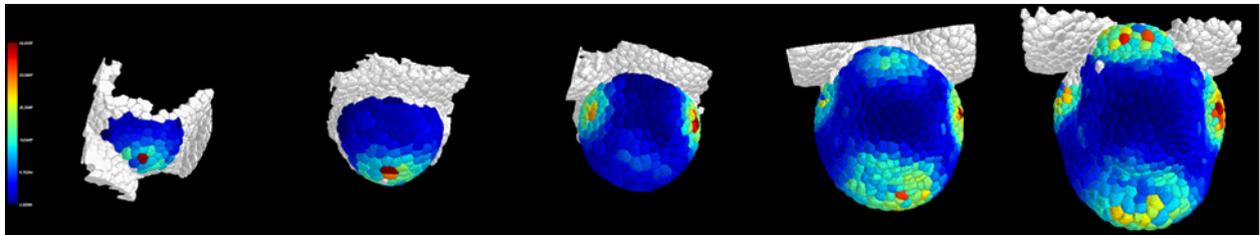
b. Top view of AHP6 signal feature rank-1 laplacian.



c. Top view of AHP6 signal feature rank-2 laplacian.



d. Top view of AHP6 signal feature rank-1 mean absolute deviation.



e. Top view of AHP6 signal feature rank-2 mean absolute deviation.

Figure S3.6: Examples of spatial differentiation functions *Laplacian* and *mean absolute deviation* applied on AHP6 signal (genetic marker) for rank-1 and 2 (topological distance).

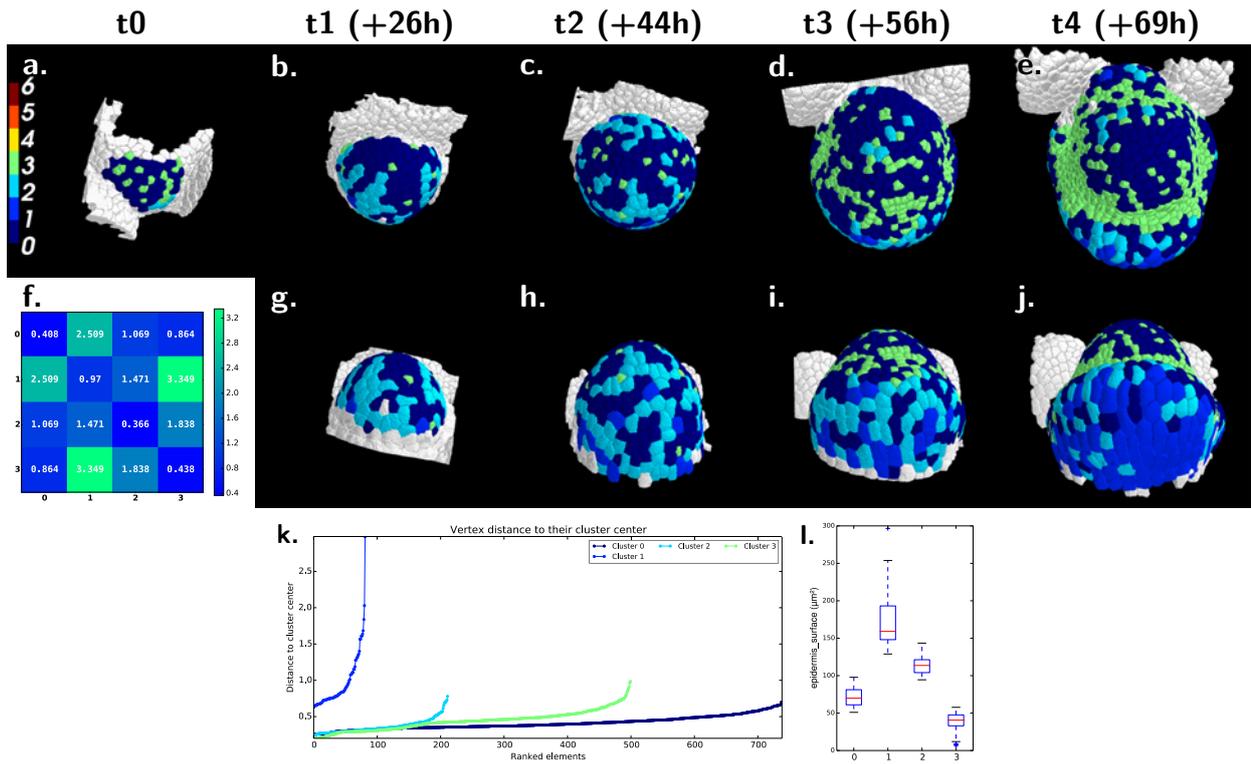


Figure S3.7: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 90% 'epidermis area' + 10% 'spatial topology' features, (a-e,g-j), and the cluster distances associated heatmap (f), ordered vertex distance to their cluster center (k) and clusters feature distributions (l).

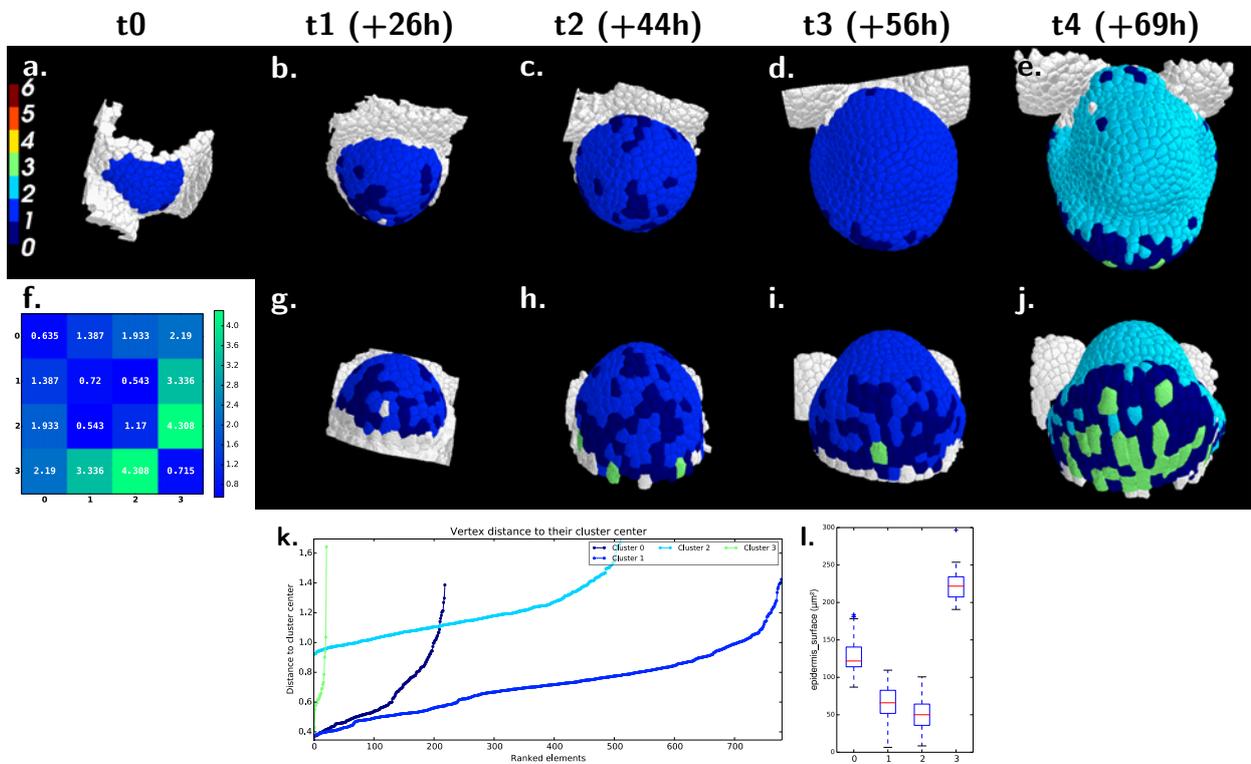


Figure S3.8: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 80% 'epidermis area' + 20% 'spatial topology' features, (a-e,g-j), and the cluster distances associated heatmap (f), ordered vertex distance to their cluster center (k) and clusters feature distributions (l).

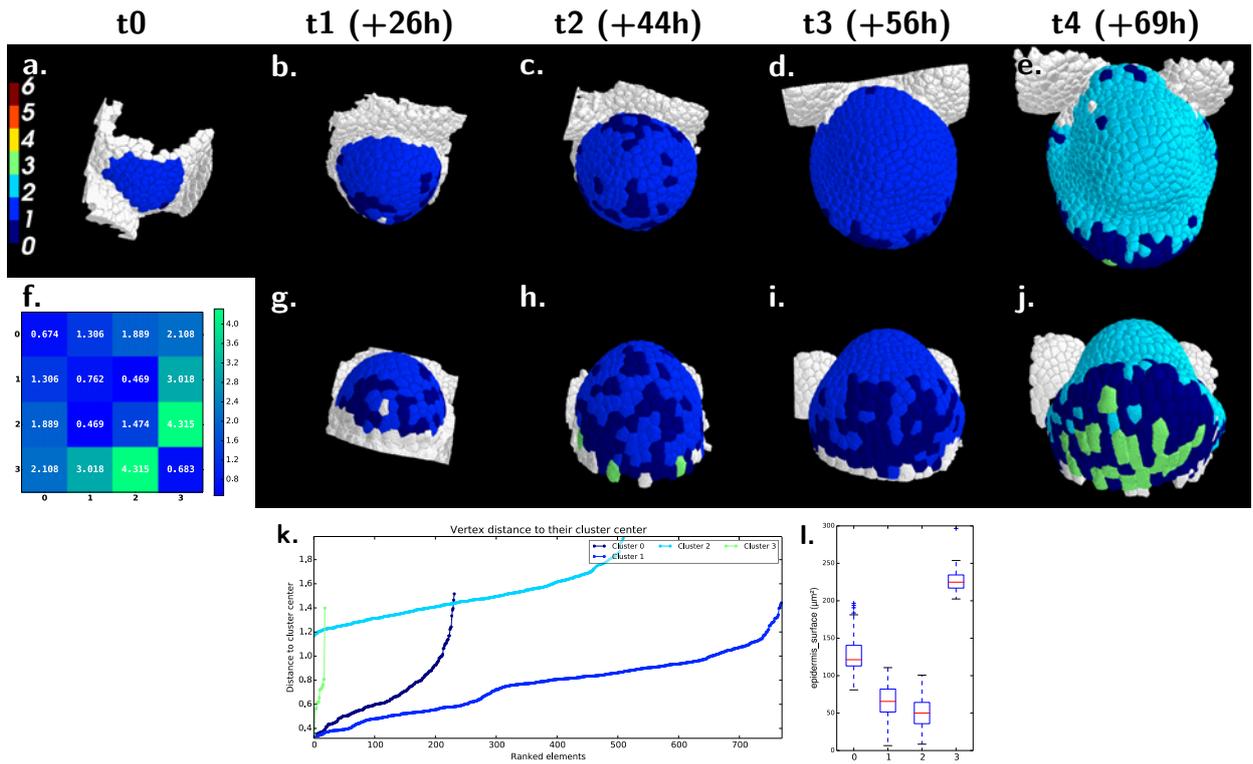


Figure S3.9: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 70% 'epidermis area' + 30% 'spatial topology' features, (a-e,g-j), and the cluster distances associated heatmap (f), ordered vertex distance to their cluster center (k) and clusters feature distributions (l).

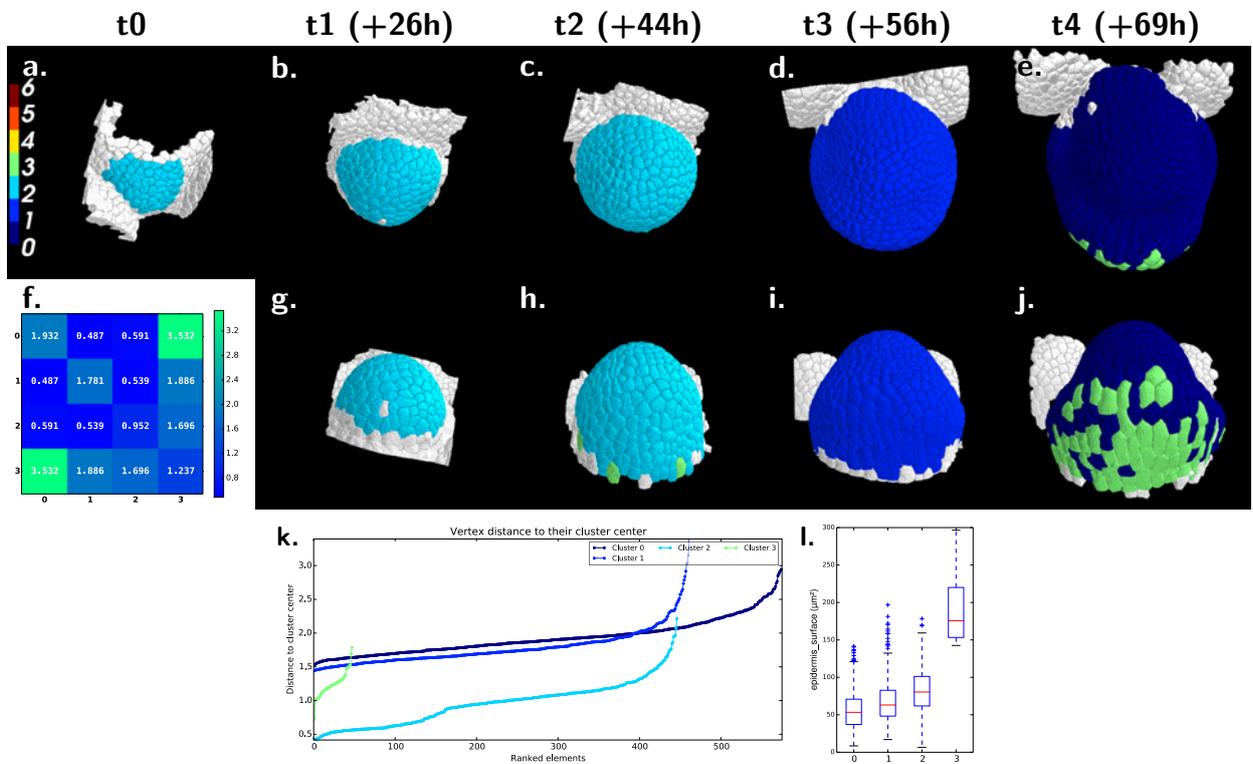


Figure S3.10: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 60% 'epidermis area' + 40% 'spatial topology' features, (a-e,g-j), and the cluster distances associated heatmap (f), ordered vertex distance to their cluster center (k) and clusters feature distributions (l).

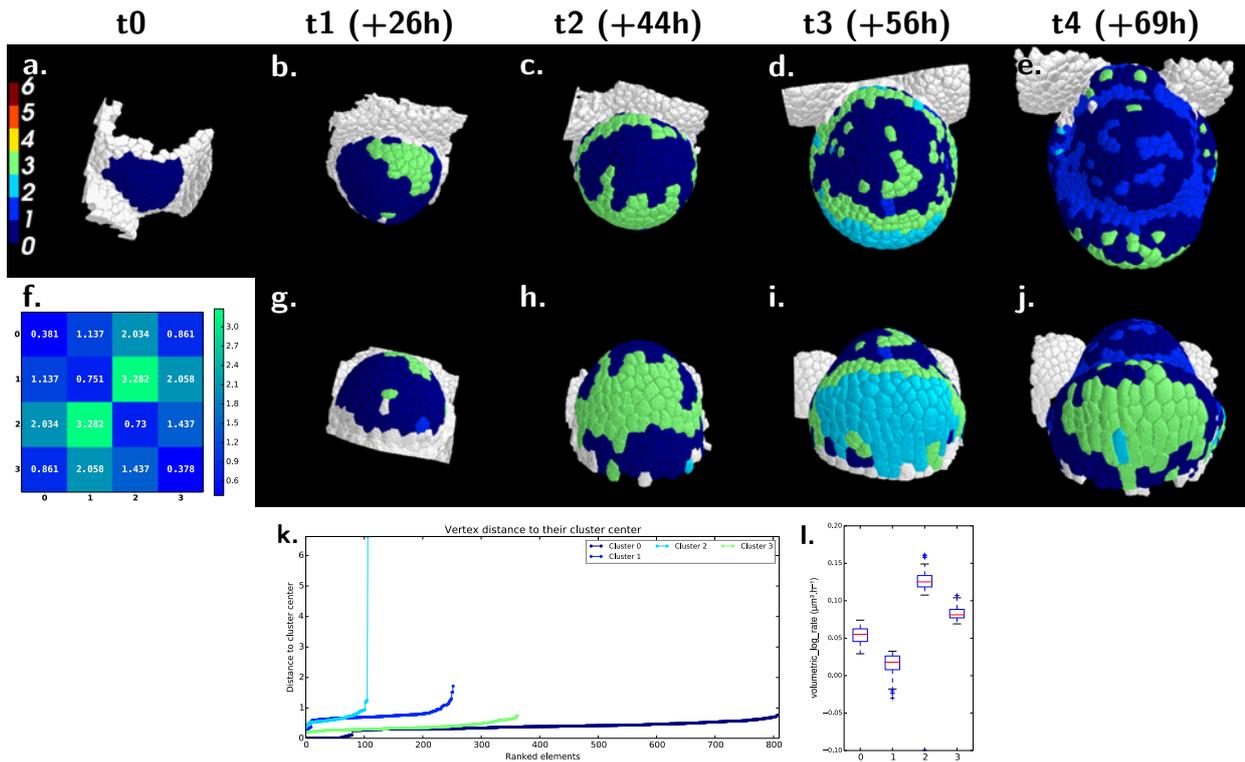


Figure S3.11: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 90% 'volumetric log rate' + 10% 'spatial topology' features, (a-e.g-j), and the cluster distances associated heatmap (f), ordered vertex distance to their cluster center (k) and clusters feature distributions (l).

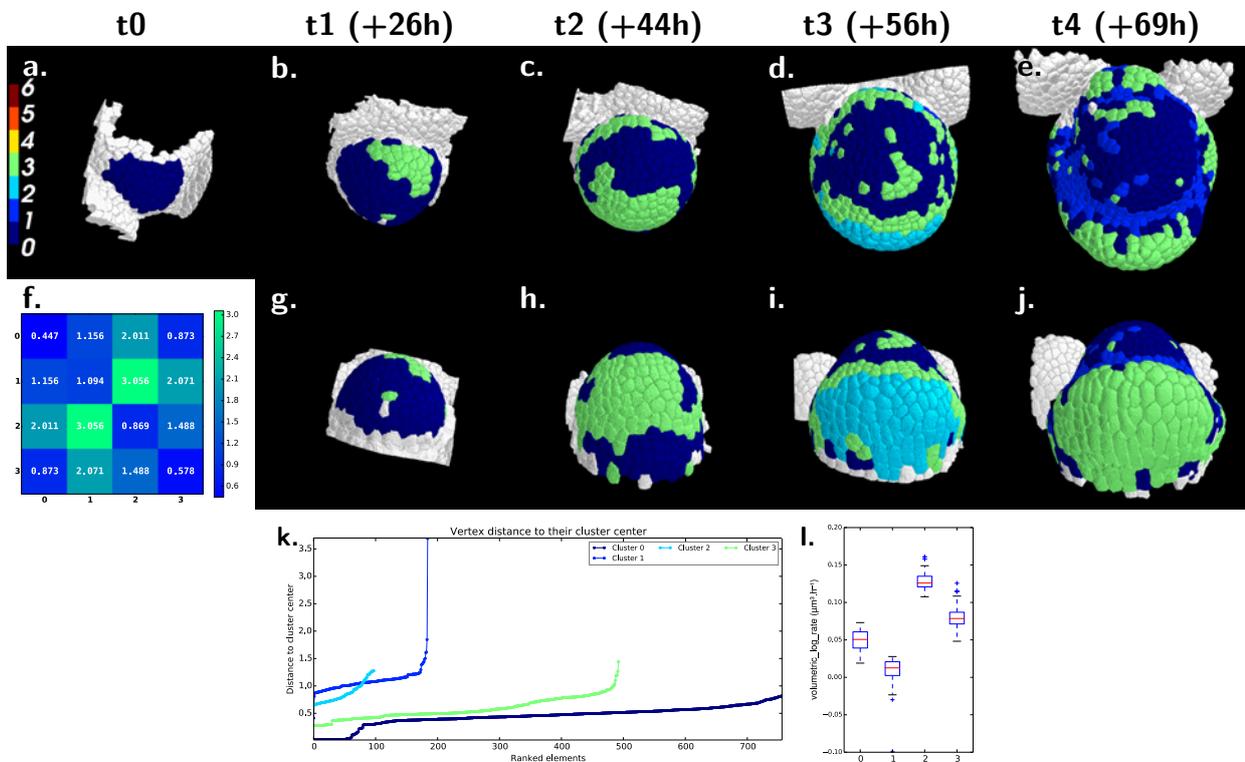


Figure S3.12: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 80% 'volumetric log rate' + 20% 'spatial topology' features, (a-e.g-j), and the cluster distances associated heatmap (f), ordered vertex distance to their cluster center (k) and clusters feature distributions (l).

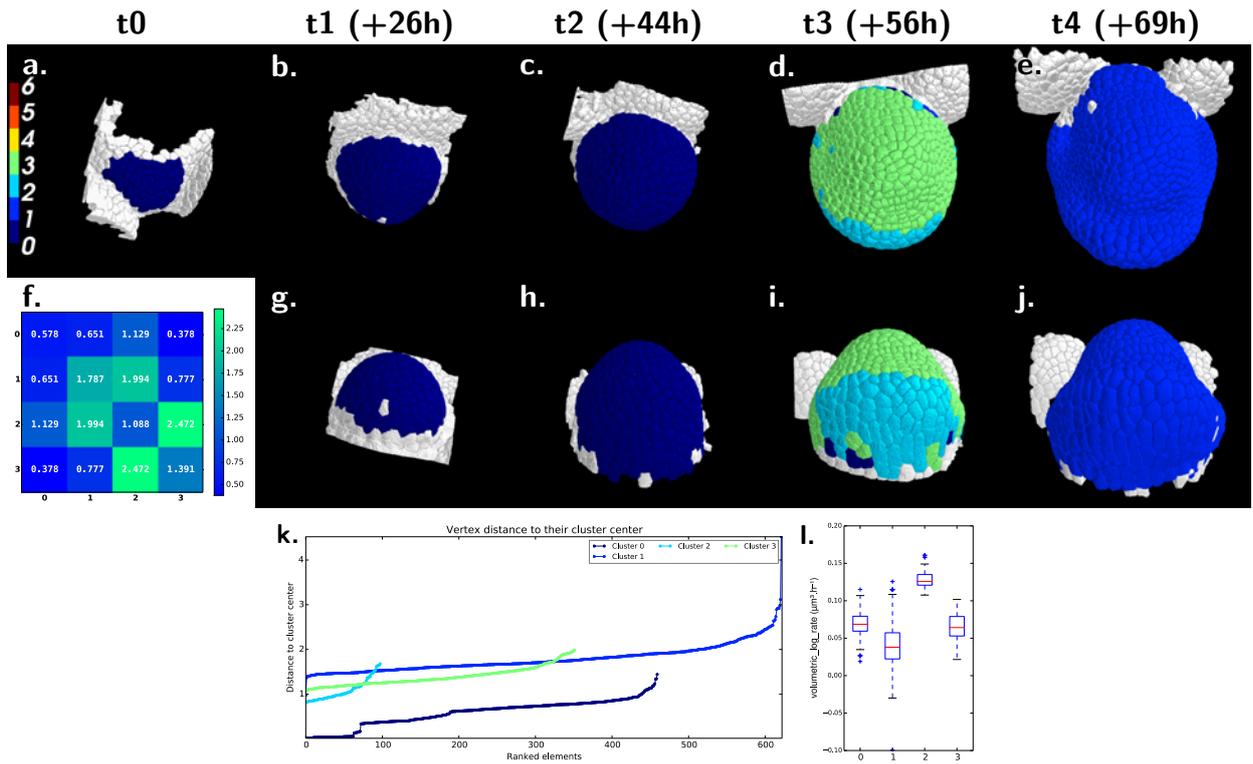


Figure S3.13: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 70% 'volumetric log rate' + 30% 'spatial topology' features, (a-e,g-j), and the cluster distances associated heatmap (f), ordered vertex distance to their cluster center (k) and clusters feature distributions (l).

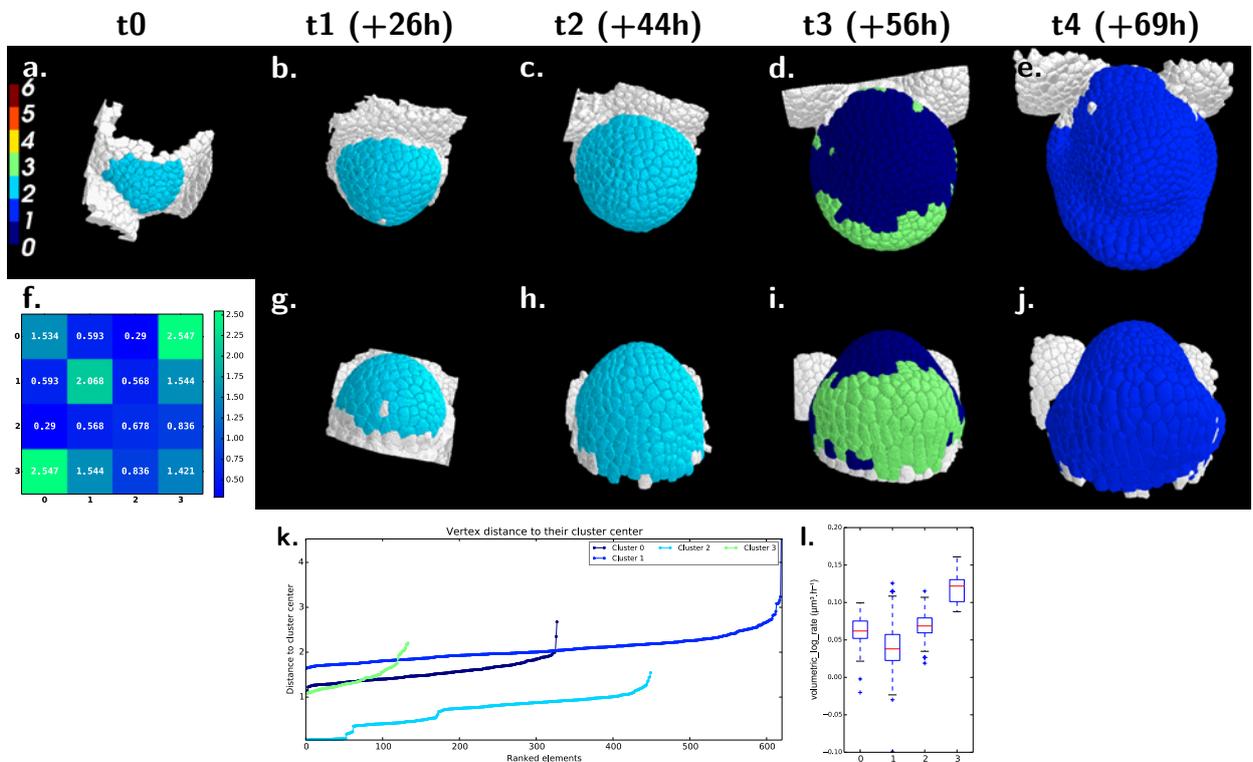


Figure S3.14: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 60% 'volumetric log rate' + 40% topological features, (a-e,g-j), and the cluster distances associated heatmap (f), ordered vertex distance to their cluster center (k) and clusters feature distributions (l).

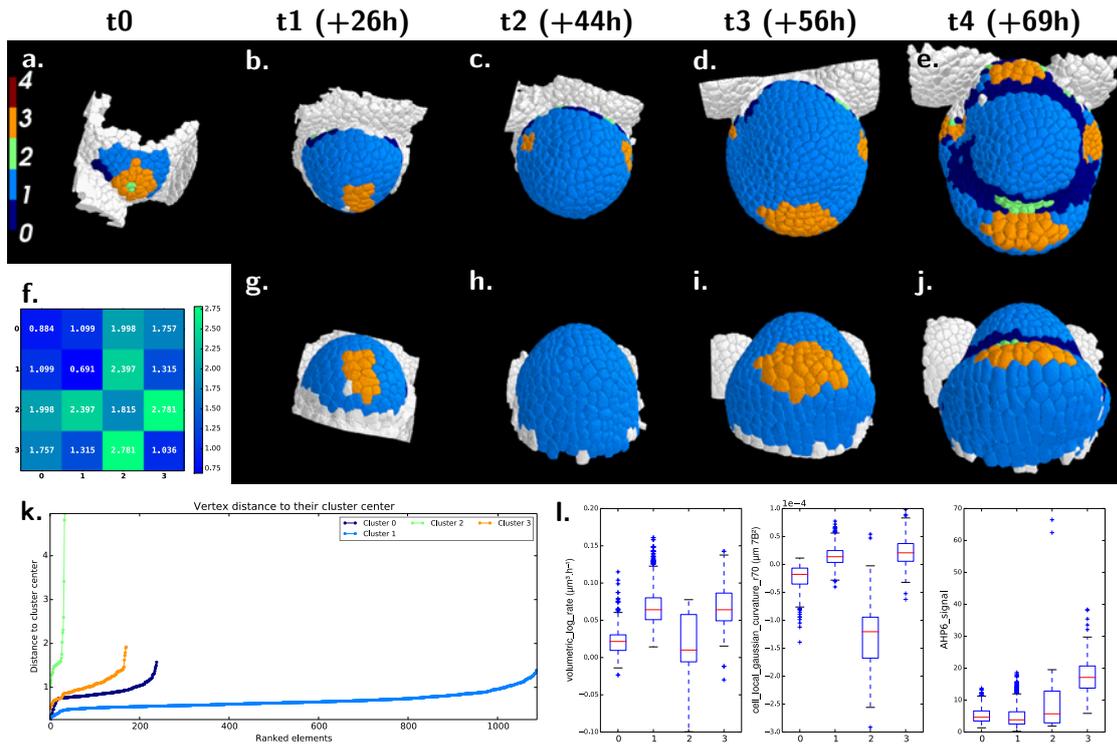


Figure S3.15: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 30% 'volumetric log temporal rate' + 30% 'Gaussian curvature' + 30% 'AHP6' + 10% 'spatial topology' features, (a-e,g-j), and the cluster distances associated heatmap (f), ordered vertex distance to their cluster center (k) and clusters feature distributions (l).

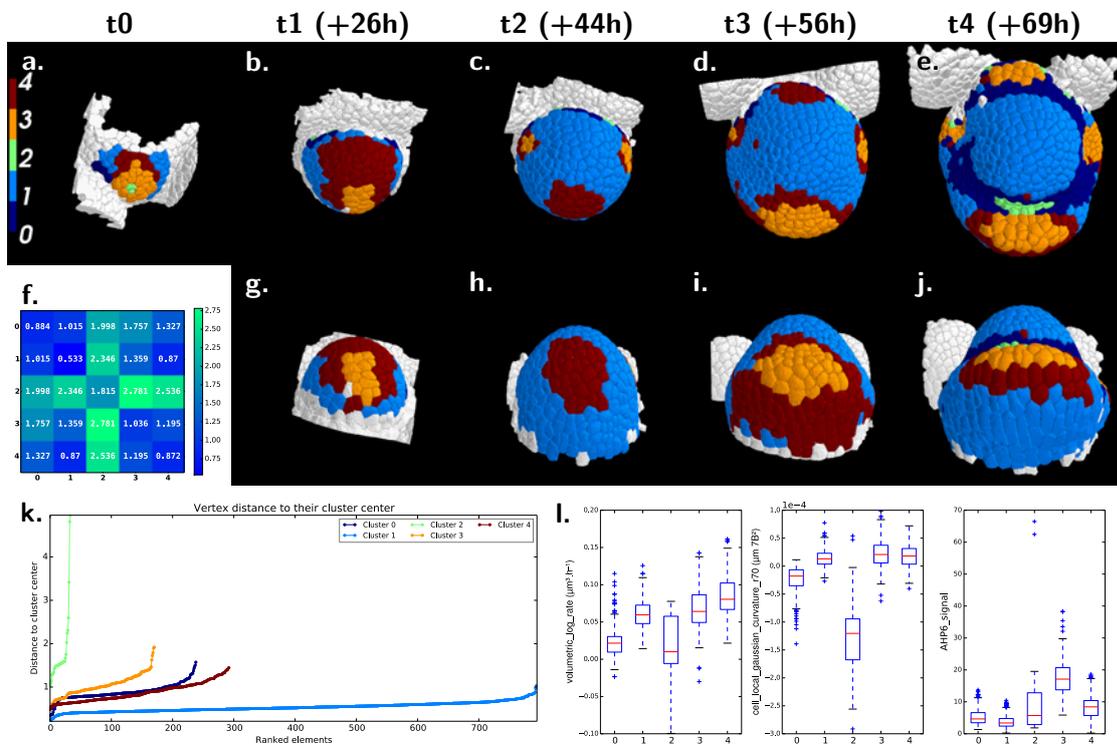


Figure S3.16: Spatial projection of the clustering obtained using a 4-cluster Ward model based on 30% 'volumetric log temporal rate' + 30% 'Gaussian curvature' + 30% 'AHP6' + 10% 'spatial topology' features, (a-e,g-j), and the cluster distances associated heatmap (f), ordered vertex distance to their cluster center (k) and clusters feature distributions (l).

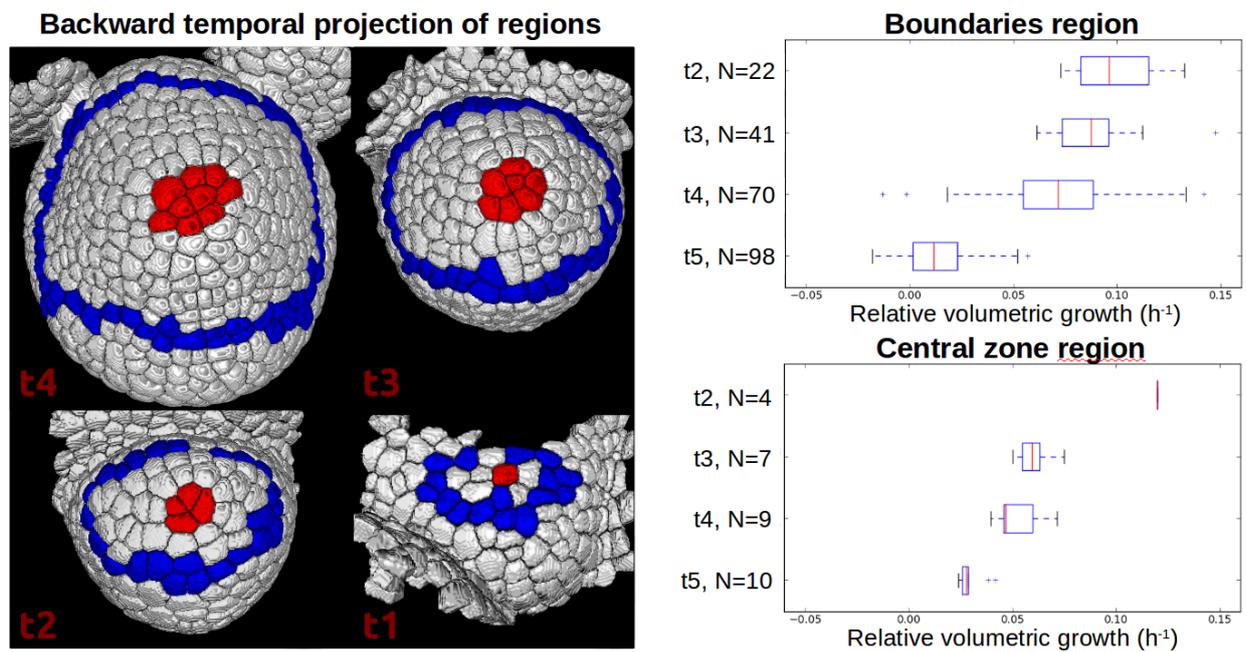


Figure S3.17: Backward temporal projections of two manually defined regions: boundary and central zone. On the left we can observe the backward projection of the regions onto the 4 time points preceding t4 presented in figure 3.27. on the right are represented the distributions of the volumetric growth rate for those two regions. In both case, we can observe a dramatic reduction of the growth rate over time, with the most significant reduction between t4 and t5. The time points are here numbered from t1 to t5 unfortunately.

General perspectives

As illustrated along the three chapters of this thesis, making use of network or spatio-temporal graph representations has allowed us to summarise, organise and relate the –often complex– information within the data. In the same time, unsupervised methods such as clustering methods (algorithmic or model-based) and modelling are powerful methods to analyse complex data, as they possess the capacity to identify and characterise emerging properties and patterns of the system under study.

Considering the analysis we have conducted about the auxin signalling network, it could be argued that the used network is only the “potential” network as we used every (testable) transcription factors of the network, however it is known that the TFs expression patterns are spatio-temporally regulated. Using the methods to analyse live-images, it could be interesting to create this spatio-temporal map at the scale of the SAM of the FM. We would indeed be able to identify TFs combinations by cells, allowing to create groups of commonly expressed auxin-signalling proteins, which combined to the quantified geometrical features, would permit to establish putative role of the sub-network regarding growth and shape processes. In addition, the identification of the auxin-responsive genes downstream-regulated by each TFs would thus allow to predict expression patterns of these genes.

In a second time, using the phylogenetic information we could hypothesise a probable domain-specific interaction capacity. Although, the relatively small size of the network may have limited the regression model and thus the biological conclusions. It would therefore be possible to enlarge the modelled network by using mutated versions of the transcription factors dimerisation domains and experimentally test their interacting capacities. Noting that the key amino-acids and their positions have been identified [Nanao *et al.*, 2014], mutating the TF randomly could result in the alteration of these precise 3D conformations. The simple proteic distance might thus be not the more adequate metric for computing the co-variables to use in the regression model. However, rice and tomato also present the Aux/IAA-ARF family, with different versions of the proteins present in *A. thaliana*. Re-conducting the analysis with those plants would thus allow to compare if they

display similar mode of dimerisation and the same buffering capacity.

Regarding the development of alternative strategies to analyse the growing tissue, the growth analysis by means of tensors fields on *D. melanogaster* presented in Bosveld *et al.* [2012] is very interesting. Although it require to posses a reference system common to every images, I believe that, by means of registrations systems and the increase of temporal resolution offered by the new in-depth fluorescent imaging microscopes, we can achieve decent interpolation of the growth dynamic. Another solution would be to limit the exposure of the sample by only imaging the surface, but at higher frequencies, in a framework allowing to measure and model of the surface shapes changes and movements. The registration of the in-depth images could then be made on the basis of the previously built model. In that respect, using the work of Boyer and Lipman [2011] would certainly be of a great help, since they propose algorithms to automatically quantify the geometric similarity of anatomical surfaces.

A good strategy to increase the temporal coverage, seems to successively acquire partially overlapping times frames (stage 1 to 3, then S2-S4, S3-S5 and so on) while repeating the coverage. The quality of the model surface will indeed be directly linked to the number and quality of time-series. In turn we would also be able to registrate any static data to the model thus opening the door to new data otherwise impossible to integrate.

This would most probably push towards the creation of an *in sillico* 3D growing tissue, using mechanical modelling of the cell walls [Smith *et al.*, 2006], incorporating cell division process [Smith *et al.*, 2006] and the possibility to add equations including fundamental biological processes such as diffusion or transport processes [Jönsson *et al.*, 2006; de Reuille *et al.*, 2006] and genetic spatio-temporal patterns [Jönsson *et al.*, 2005; Robinson *et al.*, 2011; Yadav *et al.*, 2013]. The mechanical model could in turn be refined by modelling microtubules behaviour (for review see Deinum and Mulder [2013]).

Using backwards temporal projection of manually defined regions, boundary and the central zone, we have been able to observe dramatic decrease in volumetric growth rates (see supplementary figure S3.17 from chapter 3). This was expected from the putative biological functions of these regions. However, this could be use as temporal landmarks such as the activation of the repression of majors developmental genes can be. Referencing those features changes could thus give a morphogenetic time-line, that could serve to compare time-series between them and align or compare with other ecotypes or species. This analysis could be made in combination of the hidden

Markov in-tree model, which is parametrized by transition probabilities from the children states to the parent state.

When looking at figure 3.3 in section 2.1, I get puzzled by the external shape similarity between the initial lineage-coloured cells and the shape displayed by their descendants. That could be explained by the symmetric rule of division, but yet for a tissue layer undergoing a lot of mechanical stress and growing rapidly, this is striking. In addition, we have observe that this tissue present local differences in growth rates, that should have distorted the layer. Although such behaviour would fit in the hypothesis of a protective layer made quasi-exclusively by periclinal divisions, which role is to restrain the anisotropic growth of the inner corpus thus yielding coherent shapes.

This observation of the shape steadiness bring another question regarding the origin of the cells genetical identity. Indeed, how is it possible, if we only consider the lineage to be responsible, that the boundary cross all those patches of different cells (as observable at t4 fig. 3.3). There must be a spatial signal that activate sepal emergence only when the floral bulb is big enough. To test whether an acquired genetic or hormonal patterns are lineage dependant or not, the introduced hidden Markov tree models could prove helpful by computing generation law probabilities on the basis of such signals. Quantitative versions of the signal would be advised, but it would be possible to work with other type of data such as ordered interval-scaled or binary variables.

On larger perspectives about live-imaging techniques, technological advances will continues to provides better spatio-temporal resolutions by proposing methods to acquire several angles at once for instance. Super resolution fluorescence microscopy Huang *et al.* [2009] present the new frontier to fluorescence imaging since they achieved a spatial resolution that is not limited by the diffraction of light. This technology thus allow to measure interactions and record dynamic processes in living cells at the nanometer scale. Overall, these perspectives suggest that in-depth live-imaging will soon become a preponderant method in developmental biology.

Bibliography

- ABRÀMOFF, M. D.; MAGALHÃES, P. J. and RAM, S. J. “Image Processing with ImageJ”. *Biophotonics International*, vol. 11: pp. 36–41, **2004**. doi:10.1117/1.3589100.
- AKEMANN, W.; MUTOH, H.; PERRON, A.; PARK, Y. K.; IWAMOTO, Y. and KNOPFEL, T. “Imaging neural circuit dynamics with a voltage-sensitive fluorescent protein”, **2012**. doi:10.1152/jn.00452.2012.
- BAINBRIDGE, K.; GUYOMARC’H, S.; BAYER, E.; SWARUP, R.; BENNETT, M.; MANDEL, T. and KUHLEMEIER, C. “Auxin influx carriers stabilize phyllotactic patterning.” *Genes & development*, vol. 22: pp. 810–823, **2008**. doi:10.1101/gad.462608.
- BAO, Z.; MURRAY, J. I. J. I.; BOYLE, T.; OOI, S. L.; SANDEL, M. J. and WATERSTON, R. H. “Automated cell lineage tracing in *Caenorhabditis elegans*.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103(8): pp. 2707–2712, **2006**. doi:10.1073/pnas.0511111103.
- BATRINA, I.; OTTO, E.; STRNAD, M.; WERNER, T. and SCHMÜLLING, T. “Cytokinin regulates the activity of reproductive meristems, flower organ size, ovule formation, and thus seed yield in *Arabidopsis thaliana*.” *The Plant cell*, vol. 23(1): pp. 69–80, **2011**. doi:10.1105/tpc.110.079079.
- BASSER, P. J. and PIERPAOLI, C. “Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI.” *Journal of magnetic resonance. Series B*, vol. 111: pp. 209–219, **1996**. doi:10.1016/j.jmr.2011.09.022.
- BAYER, E. M.; SMITH, R. S.; MANDEL, T.; NAKAYAMA, N.; SAUER, M.; PRUSINKIEWICZ, P. and KUHLEMEIER, C. “Integration of transport-based models for phyllotaxis and midvein formation.” *Genes & development*, vol. 23: pp. 373–384, **2009**. doi:10.1101/gad.497009.
- BENKOVÁ, E.; MICHNIEWICZ, M.; SAUER, M.; TEICHMANN, T.; SEIFERTOVÁ, D.; JÜRGENS, G. and FRIML, J. “Local, Efflux-Dependent Auxin Gradients as a Common Module for Plant Organ Formation”. *Cell*, vol. 115: pp. 591–602, **2003**. doi:10.1016/S0092-8674(03)00924-3.
- BENNETT, M. J.; MARCHANT, A.; GREEN, H. G.; MAY, S. T.; WARD, S. P.; MILLNER, P. A.; WALKER, A. R.; SCHULZ, B. and FELDMANN, K. A. “*Arabidopsis* AUX1 gene: a permease-like regulator of root gravitropism.” *Science (New York, N.Y.)*, vol. 273: pp. 948–950, **1996**. doi:10.1126/science.273.5277.948.
- VAN BERKEL, K.; DE BOER, R. J.; SCHERES, B. and TEN TUSSCHER, K. “Polar auxin transport: models and mechanisms.” *Development (Cambridge, England)*, vol. 140(11): pp. 2253–68, **2013**. doi:10.1242/dev.079111.
- BESNARD, F.; REFAHI, Y.; MORIN, V.; MARTEAUX, B.; BRUNOUD, G.; CHAMBRIER, P.; ROZIER, F.; MIRABET, V.; LEGRAND, J.; LAINÉ, S.; THÉVENON, E.; FARCOT, E.; CELLIER, C.; DAS, P.; BISHOPP, A.; DUMAS, R.; PARCY, F.; HELARIUTTA, Y.; BOUDAUD, A.; GODIN, C.; TRAAS, J.; GUÉDON, Y. and VERNOUX, T. “Cytokinin signalling inhibitory fields provide robustness to phyllotaxis.” *Nature*, vol. 505(7483): pp. 417–21, **2014**. doi:10.1038/nature12791.

- BESSON, S. and DUMAIS, J. “Universal rule for the symmetric division of plant cells.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108(15): pp. 6294–9, **2011**. doi:10.1073/pnas.1011866108.
- BINTU, L.; BUCHLER, N. E.; GARCIA, H. G.; GERLAND, U.; HWA, T.; KONDEV, J. and PHILLIPS, R. “Transcriptional regulation by the numbers: models.” *Current opinion in genetics & development*, vol. 15(2): pp. 116–24, **2005**. doi:10.1016/j.gde.2005.02.007.
- BIZZARRI, R.; SERRESI, M.; LUIN, S. and BELTRAM, F. “Green fluorescent protein based pH indicators for in vivo use: A review”, **2009**. doi:10.1007/s00216-008-2515-9.
- BLAZQUEZ, M.; GREEN, R.; NILSSON, O.; SUSSMAN, M. and WEIGEL, D. “Gibberellins promote flowering of arabidopsis by activating the LEAFY promoter”. *The Plant cell*, vol. 10(5): pp. 791–800, **1998**. doi:10.1105/tpc.10.5.791.
- BORUC, J.; VAN DEN DAELE, H.; HOLLUNDER, J.; ROMBAUTS, S.; MYLLE, E.; HILSON, P.; INZÉ, D.; DE VEYLDER, L. and RUSSINOVA, E. “Functional modules in the Arabidopsis core cell cycle binary protein-protein interaction network.” *The Plant cell*, vol. 22(4): pp. 1264–80, **2010**. doi:10.1105/tpc.109.073635.
- BOSSINGER, G. and SMYTH, D. R. “Initiation patterns of flower and floral organ development in Arabidopsis thaliana.” *Development (Cambridge, England)*, vol. 122(4): pp. 1093–102, **1996**.
- BOSVELD, F.; BONNET, I.; GUIRAO, B.; TLILI, S.; WANG, Z.; PETITALOT, A.; MARCHAND, R.; BARDET, P.-L. P.-L.; MARCQ, P.; GRANER, F.; BELLAÏCHE, Y. and BELLAÏCHE, Y. “Mechanical control of morphogenesis by Fat/Dachsous/Four-jointed planar cell polarity pathway.” *Science*, vol. 336(6082): pp. 724–7, **2012**. doi:10.1126/science.1221071.
- BOYER, D. M. M. and LIPMAN, Y. “Algorithms to automatically quantify the geometric similarity of anatomical surfaces”. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108(45): pp. 18221–18226, **2011**. doi:10.1073/pnas.1112822108.
- BRAND, U.; FLETCHER, J. C.; HOBE, M.; MEYEROWITZ, E. M. and SIMON, R. “Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity.” *Science (New York, N.Y.)*, vol. 289: pp. 617–619, **2000**. doi:10.1126/science.289.5479.617.
- BRAND, U.; GRÜNEWALD, M.; HOBE, M. and SIMON, R. “Regulation of CLV3 expression by two homeobox genes in Arabidopsis.” *Plant physiology*, vol. 129: pp. 565–575, **2002**. doi:10.1104/pp.001867.
- BRUNOUD, G.; WELLS, D. M.; OLIVA, M.; LARRIEU, A.; MIRABET, V.; BURROW, A. H.; BEECKMAN, T.; KEPINSKI, S.; TRAAS, J.; BENNETT, M. J. and VERNOUX, T. “A novel sensor to map auxin response and distribution at high spatio-temporal resolution.” *Nature*, vol. 482(7383): pp. 103–6, **2012**. doi:10.1038/nature10791.
- CASTRESANA, J. “Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis”. *Moleculr Biology and Evolution*, vol. 17(4): pp. 540–552, **2000**.
- CHAPMAN, E. J. and ESTELLE, M. “Mechanism of auxin-regulated gene expression in plants.” *Annual review of genetics*, vol. 43: pp. 265–285, **2009**. doi:10.1146/annurev-genet-102108-134148.
- CHERRY, J. L. and ADLER, F. R. “How to make a biological switch.” *Journal of theoretical biology*, vol. 203: pp. 117–133, **2000**. doi:10.1006/jtbi.2000.1068.

- CHIEN, C. T.; BARTEL, P. L.; STERNGLANZ, R. and FIELDS, S. “The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88: pp. 9578–9582, **1991**. doi:10.1073/pnas.88.21.9578.
- CLOUGH, S. J. and BENT, A. F. “Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*”. *Plant Journal*, vol. 16: pp. 735–743, **1998**. doi:10.1046/j.1365-313X.1998.00343.x.
- COEN, E.; ROLLAND-LAGAN, A.-G.; MATTHEWS, M.; BANGHAM, J. A. and PRUSINKIEWICZ, P. “The genetics of geometry.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101(14): pp. 4728–35, **2004**. doi:10.1073/pnas.0306308101.
- COEN, E. S. and MEYEROWITZ, E. M. “The war of the whorls: genetic interactions controlling flower development.” *Nature*, vol. 353(6339): pp. 31–7, **1991**. doi:10.1038/353031a0.
- CRONE, W. and LORD, E. M. “Floral organ initiation and development in wild-type *Arabidopsis thaliana* (Brassicaceae) and in the organ identity mutants *apetala2-1* and *agamous-1*”. *Canadian Journal of Botany*, vol. 72(3): pp. 384–401, **1994**. doi:10.1139/b94-052.
- CUTLER, S. R.; EHRHARDT, D. W.; GRIFFITTS, J. S. and 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*, vol. 97(7): pp. 3718–3723, **2000**. doi:10.1073/pnas.97.7.3718.
- CZECHOWSKI, T.; BARI, R. P.; STITT, M.; SCHEIBLE, W. R. and UDVARDI, M. K. “Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: Unprecedented sensitivity reveals novel root- and shoot-specific genes”. *Plant Journal*, vol. 38: pp. 366–379, **2004**. doi:10.1111/j.1365-313X.2004.02051.x.
- DAUDIN, J.; PICARD, F. and ROBIN, S. “A mixture model for random graphs”. *Statistics and Computing*, vol. 18(2): pp. 173–183, **2007**. doi:10.1007/s11222-007-9046-7.
- DE RYBEL, B.; VASSILEVA, V.; PARIZOT, B.; DEMEULENAERE, M.; GRUNEWALD, W.; AUDE-NAERT, D.; VAN CAMPENHOUT, J.; OVERVOORDE, P.; JANSEN, L.; VANNESTE, S.; MÖLLER, B.; WILSON, M.; HOLMAN, T.; VAN ISTERDAEL, G.; BRUNOUD, G.; VUYLSTEKE, M.; VERNOUX, T.; DE VEYLDER, L.; INZÉ, D.; WEIJERS, D.; BENNETT, M. J. and BEECKMAN, T. “A novel Aux/IAA28 signaling cascade activates GATA23-dependent specification of lateral root founder cell identity”. *Current Biology*, vol. 20: pp. 1697–1706, **2010**. doi:10.1016/j.cub.2010.09.007.
- DE SMET, I.; LAU, S.; VOSS, U.; VANNESTE, S.; BENJAMINS, R.; RADEMACHER, E. H.; SCHLERETH, A.; DE RYBEL, B.; VASSILEVA, V.; GRUNEWALD, W.; NAUDTS, M.; LEVESQUE, M. P.; EHRISMANN, J. S.; INZÉ, D.; LUSCHNIG, C.; BENFEY, P. N.; WEIJERS, D.; VAN MONTAGU, M. C. E.; BENNETT, M. J.; JÜRGENS, G. and BEECKMAN, T. “Bimodular auxin response controls organogenesis in *Arabidopsis*.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107: pp. 2705–2710, **2010**. doi:10.1073/pnas.0915001107.
- DEAN, K. M. and PALMER, A. E. “Advances in fluorescence labeling strategies for dynamic cellular imaging.” *Nature chemical biology*, vol. 10: pp. 512–23, **2014**. doi:10.1038/nchembio.1556.
- DEINUM, E. E. and MULDER, B. M. “Modelling the role of microtubules in plant cell morphology.” *Current opinion in plant biology*, vol. 16(6): pp. 688–92, **2013**. doi:10.1016/j.pbi.2013.10.001.
-

- DEL BIANCO, M. and KEPINSKI, S. “Context, specificity, and self-organization in auxin response”. *Cold Spring Harbor Perspectives in Biology*, vol. 3: pp. 10–20, **2011**. doi:10.1101/cshperspect.a001578.
- DESPREZ, T.; JURANIEC, M.; CROWELL, E. F.; JOUY, H.; POCHYLOVA, Z.; PARCY, F.; HÖFTE, H.; GONNEAU, M. and VERNHETTES, S. “Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104(39): pp. 15572–7, **2007**. doi:10.1073/pnas.0706569104.
- DHARMASIRI, N.; DHARMASIRI, S. and ESTELLE, M. “The F-box protein TIR1 is an auxin receptor.” *Nature*, vol. 435(7041): pp. 441–5, **2005a**. doi:10.1038/nature03543.
- DHARMASIRI, N.; DHARMASIRI, S.; WEIJERS, D.; LECHNER, E.; YAMADA, M.; HOBBIE, L.; EHRISMANN, J. S.; JÜRGENS, G. and ESTELLE, M. “Plant development is regulated by a family of auxin receptor F box proteins”. *Developmental Cell*, vol. 9: pp. 109–119, **2005b**. doi:10.1016/j.devcel.2005.05.014.
- DITTA, G.; PINYOPICH, A.; ROBLES, P.; PELAZ, S.; YANOFSKY, M. F.; DIEGO, S. and JOLLA, L. “The SEP4 gene of *Arabidopsis thaliana* functions in floral organ and meristem identity.” *Current biology : CB*, vol. 14(21): pp. 1935–40, **2004**. doi:10.1016/j.cub.2004.10.028.
- DREHER, K. A.; BROWN, J.; SAW, R. E. and CALLIS, J. “The *Arabidopsis* Aux/IAA protein family has diversified in degradation and auxin responsiveness.” *The Plant cell*, vol. 18: pp. 699–714, **2006**. doi:10.1105/tpc.105.039172.
- DUFOUR-KOWALSKI, S.; PRADAL, C.; DONES, N.; BARBIER; BOUDON, F.; CHOPARD, J.; DA SILVA, D.; DURAND, J. B.; FERRARO, P.; FOURNIER, C.; GUÉDON, Y.; OUANGRAOUA, A.; SMITH, C.; STOMA, S.; THÉVENY, T.; SINOQUET, H. and GODIN, C. “OpenAlea: An open-source platform for the integration of heterogeneous FSPM components”. In 5th International Workshop on Functional-Structural Plant Models, pp. 36+ (**2007**).
- DUMAIS, J. and KWIATKOWSKA, D. “Analysis of surface growth in shoot apices”. *The Plant Journal*, vol. 31(2): pp. 229–241, **2002**. doi:10.1046/j.1365-313X.2001.01350.x.
- DURAND, J. B.; GONÇALVÈS, P. and GUÉDON, Y. “Computational methods for hidden Markov tree models-An application to wavelet trees”. *IEEE Transactions on Signal Processing*, vol. 52: pp. 2551–2560, **2004**. doi:10.1109/TSP.2004.832006.
- ELICEIRI, K. W.; BERTHOLD, M. R.; GOLDBERG, I. G.; IBÁÑEZ, L.; MANJUNATH, B. S.; MARTONE, M. E.; MURPHY, R. F.; PENG, H.; PLANT, A. L.; ROYSAM, B.; STUURMAN, N.; STUURMANN, N.; SWEDLOW, J. R.; TOMANCAK, P. and CARPENTER, A. E. “Biological imaging software tools.” *Nature methods*, vol. 9(7): pp. 697–710, **2012**. doi:10.1038/nmeth.2084.
- ELLIOTT, R. C.; BETZNER, A. S.; HUTTNER, E.; OAKES, M. P.; TUCKER, W. Q.; GERENTES, D.; PEREZ, P. and SMYTH, D. R. “AINTEGUMENTA, an APETALA2-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth.” *The Plant cell*, vol. 8(2): pp. 155–168, **1996**.
- ERICKSON, R. O. “Modeling of Plant Growth”. *Annual Review of Plant Physiology*, vol. 27(1): pp. 407–434, **1976**. doi:10.1146/annurev.pp.27.060176.002203.
- ESTER, M.; KRIEGEL, H. P.; SANDER, J. and XU, X. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. *Second International Conference on Knowledge Discovery and Data Mining*, pp. 226–231, **1996**. doi:10.1.1.71.1980.

- FABRI, A. and PION, S. “CGAL – the Computational Geometry Algorithms Library”, **2009**. doi:10.1145/1653771.1653865.
- FERNANDEZ, R.; DAS, P.; MIRABET, V.; MOSCARDI, E.; TRAAS, J.; VERDEIL, J.-L.; MALLANDAIN, G. and GODIN, C. “Imaging plant growth in 4D: robust tissue reconstruction and lineage at cell resolution.” *Nature methods*, vol. 7(7): pp. 547–53, **2010**. doi:10.1038/nmeth.1472.
- FLETCHER, J. C.; BRAND, U.; RUNNING, M. P.; SIMON, R. and MEYEROWITZ, E. M. “Signaling of cell fate decisions by CLAVATA3 in Arabidopsis shoot meristems.” *Science (New York, N.Y.)*, vol. 283: pp. 1911–1914, **1999**. doi:10.1126/science.283.5409.1911.
- FORTUNATO, S. “Community detection in graphs”, **2010**. doi:10.1016/j.physrep.2009.11.002.
- FRALEY, C. and RAFTERY, A. E. “{MCLUST} Version 3 for {R}: {N}ormal Mixture Modeling and Model-Based Clustering”. Technical Report 504, University of Washington, Department of Statistics, **2006**.
- GÄLWEILER, L.; GUAN, C.; MÜLLER, A.; WISMAN, E.; MENDGEN, K.; YEPHREMOV, A. and PALME, K. “Regulation of polar auxin transport by AtPIN1 in Arabidopsis vascular tissue.” *Science (New York, N.Y.)*, vol. 282(5397): pp. 2226–30, **1998**. doi:10.1126/science.282.5397.2226.
- GODA, H.; SAWA, S.; ASAMI, T.; FUJIOKA, S.; SHIMADA, Y. and YOSHIDA, S. “Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in Arabidopsis.” *Plant physiology*, vol. 134: pp. 1555–1573, **2004**. doi:10.1104/pp.103.034736.
- GORDON, A. D. Classification, vol. 82 (**1999**).
- GRANDJEAN, O.; VERNOUX, T.; LAUFS, P.; BELCRAM, K.; MIZUKAMI, Y. and TRAAS, J. “In vivo analysis of cell division, cell growth, and differentiation at the shoot apical meristem in Arabidopsis.” *The Plant cell*, vol. 16(1): pp. 74–87, **2004**. doi:10.1105/tpc.017962.
- GRAY, W. M.; OSTIN, A.; SANDBERG, G.; ROMANO, C. P. and ESTELLE, M. “High temperature promotes auxin-mediated hypocotyl elongation in Arabidopsis.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95: pp. 7197–7202, **1998**. doi:10.1073/pnas.95.12.7197.
- GREEN, A. A. A.; KENNAWAY, J. R. R. and HANNA, A. I. I. “Genetic control of organ shape and tissue polarity”. *PLoS biology*, vol. 8(11): p. e1000537, **2010**. doi:10.1371/journal.pbio.1000537.
- GREENBAUM, D.; COLANGELO, C.; WILLIAMS, K. and GERSTEIN, M. “Comparing protein abundance and mRNA expression levels on a genomic scale.” *Genome biology*, vol. 4(9): p. 117, **2003**. doi:10.1186/gb-2003-4-9-117.
- GREENHAM, K.; SANTNER, A.; CASTILLEJO, C.; MOONEY, S.; SAIRANEN, I.; LJUNG, K. and ESTELLE, M. “The AFB4 auxin receptor is a negative regulator of auxin signaling in seedlings.” *Current biology : CB*, vol. 21(6): pp. 520–5, **2011**. doi:10.1016/j.cub.2011.02.029.
- GUÉDON, Y.; HEURET, P. and COSTES, E. “Comparison methods for branching and axillary flowering sequences”. *Journal of Theoretical Biology*, vol. 225(3): pp. 301–325, **2003**. doi:10.1016/S0022-5193(03)00249-2.
- GUILFOYLE, T. J. and HAGEN, G. “Auxin response factors.” *Current opinion in plant biology*, vol. 10(5): pp. 453–60, **2007**. doi:10.1016/j.pbi.2007.08.014.

- GYGI, S. P.; ROCHON, Y.; FRANZA, B. R. and AEBERSOLD, R. “Correlation between protein and mRNA abundance in yeast.” *Molecular and cellular biology*, vol. 19: pp. 1720–1730, **1999**.
- HAGEN, G. and GUILFOYLE, T. “Auxin-responsive gene expression: genes, promoters and regulatory factors.” *Plant molecular biology*, vol. 49(3-4): pp. 373–385, **2002**. doi:10.1023/A:1015207114117.
- HAMANN, T.; BENKOVA, E.; BÄURLE, I.; KIENZ, M. and JÜRGENS, G. “The Arabidopsis *BODENLOS* gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning”. *Genes and Development*, vol. 16: pp. 1610–1615, **2002**. doi:10.1101/gad.229402.
- HAMANT, O.; HEISLER, M. G.; JÖNSSON, H.; KRUPINSKI, P.; UYTTEWAAL, M.; BOKOV, P.; CORSON, F.; SAHLIN, P.; BOUDAUD, A.; MEYEROWITZ, E. M.; COUDER, Y. and TRAAS, J. “Developmental patterning by mechanical signals in Arabidopsis.” *Science (New York, N.Y.)*, vol. 322(5908): pp. 1650–5, **2008**. doi:10.1126/science.1165594.
- HARDTKE, C. S. and BERLETH, T. “The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development”. *EMBO Journal*, vol. 17: pp. 1405–1411, **1998**. doi:10.1093/emboj/17.5.1405.
- HARDTKE, C. S.; CKURSHUMOVA, W.; VIDAURRE, D. P.; SINGH, S. A.; STAMATIOU, G.; TIWARI, S. B.; HAGEN, G.; GUILFOYLE, T. J. and BERLETH, T. “Overlapping and non-redundant functions of the Arabidopsis auxin response factors MONOPTEROS and NONPHOTOTROPIC HYPOCOTYL 4.” *Development*, vol. 131(5): pp. 1089–1100, **2004**. doi:10.1242/dev.00925.
- HASTIE, T.; TIBSHIRANI, R. and FRIEDMAN, J. The elements of statistical learning (**2009**).
- HEISLER, M. G.; OHNO, C.; DAS, P.; SIEBER, P.; REDDY, G. V.; LONG, J. A. and MEYEROWITZ, E. M. “Patterns of auxin transport and gene expression during primordium development revealed by live imaging of the Arabidopsis inflorescence meristem.” *Current biology : CB*, vol. 15(21): pp. 1899–911, **2005**. doi:10.1016/j.cub.2005.09.052.
- HILL, J. P. and LORD, E. M. “Floral development in Arabidopsis thaliana : a comparison of the wild type and the homeotic pistillata mutant”. *Canadian Journal of Botany*, vol. 67(10): pp. 2922–2936, **1989**. doi:10.1139/b89-375.
- HONMA, T. and GOTO, K. “Complexes of MADS-box proteins are sufficient to convert leaves into floral organs.” *Nature*, vol. 409(6819): pp. 525–9, **2001**. doi:10.1038/35054083.
- HU, Y.; XIE, Q. and CHUA, N.-H. “The Arabidopsis auxin-inducible gene ARGOS controls lateral organ size.” *The Plant cell*, vol. 15(9): pp. 1951–1961, **2003**. doi:10.1105/tpc.013557.
- HUANG, B.; BATES, M. and ZHUANG, X. “Super-resolution fluorescence microscopy.” *Annual review of biochemistry*, vol. 78: pp. 993–1016, **2009**. doi:10.1146/annurev.biochem.77.061906.092014.
- JENIK, P. D. and IRISH, V. F. “Regulation of cell proliferation patterns by homeotic genes during Arabidopsis floral development.” *Development (Cambridge, England)*, vol. 127: pp. 1267–1276, **2000**.
- JÖNSSON, H.; HEISLER, M.; REDDY, G. V.; AGRAWAL, V.; GOR, V.; SHAPIRO, B. E.; MJOLNESS, E. and MEYEROWITZ, E. M. “Modeling the organization of the WUSCHEL expression domain in the shoot apical meristem.” *Bioinformatics (Oxford, England)*, vol. 21 Suppl 1(suppl_1): pp. i232–40, **2005**. doi:10.1093/bioinformatics/bti1036.

- JÖNSSON, H.; HEISLER, M. G.; SHAPIRO, B. E.; MEYEROWITZ, E. M. and MJOLSNESS, E. “An auxin-driven polarized transport model for phyllotaxis.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103: pp. 1633–1638, **2006**. doi:10.1073/pnas.0509839103.
- JOUNG, J. K.; RAMM, E. I. and PABO, C. O. “A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97(13): pp. 7382–7387, **2000**. doi:10.1073/pnas.110149297.
- KANKAANPÄÄ, P.; PAAVOLAINEN, L.; TIITTA, S.; KARJALAINEN, M.; PÄIVÄRINNE, J.; NIEMINEN, J.; MARJOMÄKI, V.; HEINO, J. and WHITE, D. J. “BioImageXD: an open, general-purpose and high-throughput image-processing platform”, **2012**. doi:10.1038/nmeth.2047.
- KASAS, S.; THOMSON, N. H.; SMITH, B. L.; HANSMA, P. K.; MIKLOSSY, J. and HANSMA, H. G. “Biological applications of the AFM: From single molecules to organs”. *International Journal of Imaging Systems and Technology*, vol. 8(2): pp. 151–161, **1997**. doi:10.1002/(SICI)1098-1098(1997)8:2<151::AID-IMA2>3.0.CO;2-9.
- KASS, R. E. and RAFTERY, A. E. “Bayes Factors”. *Journal of the American Statistical Association*, vol. 90(430): pp. 773–795, **1995**.
- KAUFMAN, L. and ROUSSEEUW, P. J. Finding Groups in Data. Wiley Series in Probability and Statistics (John Wiley & Sons, Inc., Hoboken, NJ, USA, **1990**). doi:10.1002/9780470316801.
- KEPINSKI, S. and LEYSER, O. “The Arabidopsis F-box protein TIR1 is an auxin receptor.” *Nature*, vol. 435: pp. 446–451, **2005**. doi:10.1038/nature03542.
- KIERZKOWSKI, D.; NAKAYAMA, N.; ROUTIER-KIERZKOWSKA, A.-L.; WEBER, A.; BAYER, E.; SCHORDERET, M.; REINHARDT, D.; KUHLEMEIER, C. and SMITH, R. S. “Elastic domains regulate growth and organogenesis in the plant shoot apical meristem.” *Science (New York, N.Y.)*, vol. 335(6072): pp. 1096–9, **2012**. doi:10.1126/science.1213100.
- KIM, J.; HARTER, K. and THEOLOGIS, A. “Protein-protein interactions among the Aux/IAA proteins.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94(22): pp. 11786–91, **1997**.
- KNEEN, M.; FARINAS, J.; LI, Y. and VERKMAN, A. S. “Green fluorescent protein as a noninvasive intracellular pH indicator.” *Biophysical journal*, vol. 74: pp. 1591–1599, **1998**. doi:10.1016/S0006-3495(98)77870-1.
- KONCZ, C. and SCHELL, J. “The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector”. *MGG Molecular & General Genetics*, vol. 204: pp. 383–396, **1986**. doi:10.1007/BF00331014.
- KONDO, T.; SAWA, S.; KINOSHITA, A.; MIZUNO, S.; KAKIMOTO, T.; FUKUDA, H. and SAKAGAMI, Y. “A plant peptide encoded by CLV3 identified by in situ MALDI-TOF MS analysis.” *Science (New York, N.Y.)*, vol. 313: pp. 845–848, **2006**. doi:10.1126/science.1128439.
- KORASICK, D. A.; WESTFALL, C. S.; LEE, S. G.; NANAQ, M. H.; DUMAS, R.; HAGEN, G.; GUILFOYLE, T. J.; JEZ, J. M. and STRADER, L. C. “Molecular basis for AUXIN RESPONSE FACTOR protein interaction and the control of auxin response repression.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111(14): pp. 5427–32, **2014**. doi:10.1073/pnas.1400074111.

- KRIEGEL, H.-P.; KRÖGER, P.; SANDER, J. and ZIMEK, A. “Density-based clustering”. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1(3): pp. 231–240, **2011**. doi:10.1002/widm.30.
- KRIZEK, B. “AINTEGUMENTA and AINTEGUMENTA-LIKE6 act redundantly to regulate Arabidopsis floral growth and patterning.” *Plant physiology*, vol. 150: pp. 1916–1929, **2009**. doi:10.1104/pp.109.141119.
- KRIZEK, B. A. “Ectopic expression of AINTEGUMENTA in Arabidopsis plants results in increased growth of floral organs.” *Developmental genetics*, vol. 25(3): pp. 224–36, **1999**. doi:10.1002/(SICI)1520-6408(1999)25:3<224::AID-DVG5>3.0.CO;2-Y.
- KRIZEK, B. A. and FLETCHER, J. C. “Molecular mechanisms of flower development: an armchair guide.” *Nature reviews. Genetics*, vol. 6(9): pp. 688–98, **2005**. doi:10.1038/nrg1675.
- KWIATKOWSKA, D. “Surface growth at the reproductive shoot apex of Arabidopsis thaliana pin-formed 1 and wild type.” *Journal of experimental botany*, vol. 55(399): pp. 1021–32, **2004**. doi:10.1093/jxb/erh109.
- KWIATKOWSKA, D. “Flower primordium formation at the Arabidopsis shoot apex: quantitative analysis of surface geometry and growth”. *Journal of Experimental Botany*, vol. 57(3): pp. 571–80, **2006**. doi:10.1093/jxb/erj042.
- LAUX, T.; MAYER, K. F.; BERGER, J. and JÜRGENS, G. “The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis.” *Development (Cambridge, England)*, vol. 122: pp. 87–96, **1996**.
- LEE, H.; AKERS, W.; BHUSHAN, K.; BLOCH, S.; SUDLOW, G.; TANG, R. and ACHILEFU, S. “Near-infrared pH-activatable fluorescent probes for imaging primary and metastatic breast tumors”. *Bioconjugate Chemistry*, vol. 22: pp. 777–784, **2011**. doi:10.1021/bc100584d.
- LEYSER, O. “Dynamic integration of auxin transport and signalling.” *Current biology : CB*, vol. 16(11): pp. R424–33, **2006**. doi:10.1016/j.cub.2006.05.014.
- LJUNG, K.; BHALERAO, R. P. and SANDBERG, G. “Sites and homeostatic control of auxin biosynthesis in Arabidopsis during vegetative growth.” *The Plant journal : for cell and molecular biology*, vol. 28: pp. 465–474, **2001**. doi:10.1046/j.1365-313X.2001.01173.x.
- LJUNG, K.; HULL, A. K.; CELENZA, J.; YAMADA, M.; ESTELLE, M.; NORMANLY, J. and SANDBERG, G. “Sites and regulation of auxin biosynthesis in Arabidopsis roots.” *The Plant cell*, vol. 17: pp. 1090–1104, **2005**. doi:10.1105/tpc.104.029272.
- LONG, F.; PENG, H.; LIU, X.; KIM, S. K. K. S. K. and MYERS, E. “A 3D digital atlas of *C. elegans* and its application to single-cell analyses”. *Nature methods*, vol. 6(9): pp. 667–672, **2009**. doi:10.1038/nmeth.1366.
- LONG, J. A.; MOAN, E. I.; MEDFORD, J. I. and BARTON, M. K. “A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of Arabidopsis.” *Nature*, vol. 379: pp. 66–69, **1996**. doi:10.1038/379066a0.
- MAIER, A. T.; STEHLING-SUN, S.; WOLLMANN, H.; DEMAR, M.; HONG, R. L.; HAUBEISS, S.; WEIGEL, D. and LOHMANN, J. U. “Dual roles of the bZIP transcription factor PERIANTHIA in the control of floral architecture and homeotic gene expression.” *Development (Cambridge, England)*, vol. 136(10): pp. 1613–20, **2009**. doi:10.1242/dev.033647.

- MARIADASSOU, M.; ROBIN, S. and VACHER, C. “Uncovering latent structure in valued graphs: A variational approach”. *The Annals of Applied Statistics*, vol. 4(2): pp. 715–742, **2010**. doi:10.1214/10-AOAS361.
- MAYER, K. F. X.; SCHOOF, H.; HAECKER, A.; LENHARD, M.; JÜRGENS, G. and LAUX, T. “Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem”. *Cell*, vol. 95: pp. 805–815, **1998**. doi:10.1016/S0092-8674(00)81703-1.
- MEGASON, S. G. G. and FRASER, S. E. E. “Imaging in systems biology”. *Cell*, vol. 130(5): pp. 784–95, **2007**. doi:10.1016/j.cell.2007.08.031.
- MIESENBOCK, G.; DE ANGELIS, D. A. and ROTHMAN, J. E. “Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins.” *Nature*, vol. 394: pp. 192–195, **1998**. doi:10.1038/28190.
- MURRAY, J. I.; BAO, Z.; BOYLE, T. J.; BOECK, M. E.; MERICLE, B. L.; NICHOLAS, T. J.; ZHAO, Z.; SANDEL, M. J. and WATERSTON, R. H. “Automated analysis of embryonic gene expression with cellular resolution in *C. elegans*.” *Nature methods*, vol. 5(8): pp. 703–709, **2008**. doi:10.1038/nmeth.1228.
- MUTO, H.; WATAHIKI, M. K.; NAKAMOTO, D.; KINJO, M. and YAMAMOTO, K. T. “Specificity and similarity of functions of the Aux/IAA genes in auxin signaling of Arabidopsis revealed by promoter-exchange experiments among MSG2/IAA19, AXR2/IAA7, and SLR/IAA14.” *Plant physiology*, vol. 144: pp. 187–196, **2007**. doi:10.1104/pp.107.096628.
- NAGAI, T.; IBATA, K.; PARK, E. S.; KUBOTA, M.; MIKOSHIBA, K. and MIYAWAKI, A. “A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications.” *Nature biotechnology*, vol. 20: pp. 87–90, **2002**. doi:10.1038/nbt0102-87.
- NANAO, M. H.; VINOS-POYO, T.; BRUNOUD, G.; THÉVENON, E.; MAZZOLENI, M.; MAST, D.; LAINÉ, S.; WANG, S.; HAGEN, G.; LI, H.; GUILFOYLE, T. J.; PARCY, F.; VERNOUX, T. and DUMAS, R. “Structural basis for oligomerization of auxin transcriptional regulators.” *Nature communications*, vol. 5: p. 3617, **2014**. doi:10.1038/ncomms4617.
- NEMHAUSER, J. L.; MOCKLER, T. C. and CHORY, J. “Interdependency of brassinosteroid and auxin signaling in Arabidopsis”. *PLoS Biology*, vol. 2, **2004**. doi:10.1371/journal.pbio.0020258.
- NIMCHUK, Z. L.; TARR, P. T.; OHNO, C.; QU, X. and MEYEROWITZ, E. M. “Plant stem cell signaling involves ligand-dependent trafficking of the CLAVATA1 receptor kinase”. *Current Biology*, vol. 21: pp. 345–352, **2011**. doi:10.1016/j.cub.2011.01.039.
- NOWICKI, K. and SNIJDERS, T. A. B. A. B. “Estimation and prediction for stochastic block-structures”. *Journal of the American Statistical ...*, vol. 96(455): pp. 1077–1087, **2001**.
- OAKENFULL, E. A.; RIOU-KHAMLICHI, C. and MURRAY, J. A. H. “Plant D-type cyclins and the control of G1 progression.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 357: pp. 749–760, **2002**. doi:10.1098/rstb.2002.1085.
- OATES, A. C.; GORFINKIEL, N.; GONZÁLEZ-GAITÁN, M. and HEISENBERG, C.-P. “Quantitative approaches in developmental biology.” *Nature reviews. Genetics*, vol. 10(8): pp. 517–30, **2009**. doi:10.1038/nrg2548.
- OGAWA, M.; SHINOHARA, H.; SAKAGAMI, Y. and MATSUBAYASHI, Y. “Arabidopsis CLV3 peptide directly binds CLV1 ectodomain.” *Science (New York, N.Y.)*, vol. 319: p. 294, **2008**. doi:10.1126/science.1150083.

- OHARA-IMAIZUMI, M.; NAKAMICHI, Y.; TANAKA, T.; KATSUTA, H.; ISHIDA, H. and NAGAMATSU, S. “Monitoring of exocytosis and endocytosis of insulin secretory granules in the pancreatic beta-cell line MIN6 using pH-sensitive green fluorescent protein (pHluorin) and confocal laser microscopy.” *The Biochemical journal*, vol. 363: pp. 73–80, **2002**. doi:10.1042/0264-6021:3630073.
- OKADA, K.; UEDA, J.; KOMAKI, M. K.; BELL, C. J. and SHIMURA, Y. “Requirement of the Auxin Polar Transport System in Early Stages of Arabidopsis Floral Bud Formation.” *The Plant cell*, vol. 3(7): pp. 677–684, **1991**. doi:10.1105/tpc.3.7.677.
- OKUMOTO, S.; JONES, A. and FROMMER, W. B. “Quantitative imaging with fluorescent biosensors.” *Annual review of plant biology*, vol. 63: pp. 663–706, **2012**. doi:10.1146/annurev-arplant-042110-103745.
- OLIPHANT, T. E. “SciPy: Open source scientific tools for Python”. *Computing in Science and Engineering*, vol. 9: pp. 10–20, **2007**.
- PACIOREK, T. and FRIML, J. “Auxin signaling.” *Journal of cell science*, vol. 119(Pt 7): pp. 1199–1202, **2006**. doi:10.1242/jcs.02910.
- PACIOREK, T.; ZAZÍMALOVÁ, E.; RUTHARDT, N.; PETRÁSEK, J.; STIERHOF, Y.-D.; KLEINEVEHN, J.; MORRIS, D. A.; EMANS, N.; JÜRGENS, G.; GELDNER, N. and FRIML, J. “Auxin inhibits endocytosis and promotes its own efflux from cells.” *Nature*, vol. 435: pp. 1251–1256, **2005**. doi:10.1038/nature03633.
- PAPONOV, I. A.; PAPONOV, M.; TEALE, W.; MENGES, M.; CHAKRABORTEE, S.; MURRAY, J. A. H. and PALME, K. “Comprehensive transcriptome analysis of auxin responses in Arabidopsis”. *Molecular Plant*, vol. 1: pp. 321–337, **2008**. doi:10.1093/mp/ssm021.
- PARRY, G.; CALDERON-VILLALOBOS, L. I.; PRIGGE, M.; PERET, B.; DHARMASIRI, S.; ITOH, H.; LECHNER, E.; GRAY, W. M.; BENNETT, M. and ESTELLE, M. “Complex regulation of the TIR1/AFB family of auxin receptors.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106(52): pp. 22540–5, **2009**. doi:10.1073/pnas.0911967106.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M. and DUCHESNAY, E. “Scikit-learn: Machine Learning in Python”. *Journal of Machine Learning Research*, vol. 12: pp. 2825–2830, **2012**.
- PEKKER, I.; ALVAREZ, J. P. and ESHED, Y. “Auxin response factors mediate Arabidopsis organ asymmetry via modulation of KANADI activity.” *The Plant cell*, vol. 17: pp. 2899–2910, **2005**. doi:10.1105/tpc.105.034876.
- PELAZ, S.; DITTA, G. S.; BAUMANN, E.; WISMAN, E. and YANOFSKY, M. F. “B and C floral organ identity functions require SEPALLATA MADS-box genes.” *Nature*, vol. 405(6783): pp. 200–3, **2000**. doi:10.1038/35012103.
- PICARD, F.; MIELE, V.; DAUDIN, J.-J. J. J.; COTTRET, L. and ROBIN, S. “Deciphering the connectivity structure of biological networks using MixNet.” *BMC bioinformatics*, vol. 10 Suppl 6: p. S17, **2009**. doi:10.1186/1471-2105-10-S6-S17.
- PUFKY, J.; QIU, Y.; RAO, M. V.; HURBAN, P. and JONES, A. M. “The auxin-induced transcriptome for etiolated Arabidopsis seedlings using a structure/function approach.” *Functional & integrative genomics*, vol. 3(4): pp. 135–43, **2003**. doi:10.1007/s10142-003-0093-7.

- RAVEN, J. A. “Transport of indoleacetic acid in plant cells in relation to pH and electrical potential gradients, and its significance for polar IAA transport.” *New Phytologist*, vol. 74(2): pp. 163–172, **1975**. doi:10.1111/j.1469-8137.1975.tb02602.x.
- REDDY, G. V.; HEISLER, M. G.; EHRHARDT, D. W. and MEYEROWITZ, E. M. “Real-time lineage analysis reveals oriented cell divisions associated with morphogenesis at the shoot apex of *Arabidopsis thaliana*.” *Development (Cambridge, England)*, vol. 131(17): pp. 4225–4237, **2004**. doi:10.1242/dev.01261.
- REINHARDT, D.; PESCE, E.-R.; STIEGER, P.; MANDEL, T.; BALTENSBERGER, K.; BENNETT, M.; TRAAS, J.; FRIML, J. and KUHLEMEIER, C. “Regulation of phyllotaxis by polar auxin transport.” *Nature*, vol. 426(6964): pp. 255–60, **2003**. doi:10.1038/nature02081.
- REMINGTON, D. L.; VISION, T. J.; GUILFOYLE, T. J. and REED, J. W. “Contrasting modes of diversification in the Aux/IAA and ARF gene families.” *Plant physiology*, vol. 135(3): pp. 1738–52, **2004**. doi:10.1104/pp.104.039669.
- DE REUILLE, P. B.; BOHN-COURSEAU, I.; LJUNG, K.; MORIN, H.; CARRARO, N.; GODIN, C. and TRAAS, J. “Computer simulations reveal properties of the cell-cell signaling network at the shoot apex in *Arabidopsis*.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103: pp. 1627–1632, **2006**. doi:10.1073/pnas.0510130103.
- ROBINSON, S.; BARBIER DE REUILLE, P.; CHAN, J.; BERGMANN, D.; PRUSINKIEWICZ, P. and COEN, E. “Generation of spatial patterns through cell polarity switching.” *Science (New York, N.Y.)*, vol. 333(6048): pp. 1436–40, **2011**. doi:10.1126/science.1202185.
- ROEDER, A. H. K. H. K.; CHICKARMANE, V. and CUNHA, A. “Variability in the control of cell division underlies sepal epidermal patterning in *Arabidopsis thaliana*”. *PLoS biology*, vol. 8(5): p. e1000367, **2010**. doi:10.1371/journal.pbio.1000367.
- ROGG, L. E.; LASSWELL, J. and BARTEL, B. “A gain-of-function mutation in IAA28 suppresses lateral root development.” *The Plant cell*, vol. 13: pp. 465–480, **2001**. doi:10.1105/tpc.13.3.465.
- ROLLAND-LAGAN, A.-G.; BANGHAM, J. A. and COEN, E. “Growth dynamics underlying petal shape and asymmetry.” *Nature*, vol. 422(6928): pp. 161–3, **2003**. doi:10.1038/nature01443.
- ROUSE, D.; MACKAY, P.; STIRNBERG, P.; ESTELLE, M. and LEYSER, O. “Changes in auxin response from mutations in an AUX/IAA gene.” *Science (New York, N.Y.)*, vol. 279(5355): pp. 1371–3, **1998**. doi:10.1126/science.279.5355.1371.
- RUBERY, P. H. and SHELDRAKE, A. R. “Carrier-mediated auxin transport.” *Planta*, vol. 118(2): pp. 101–21, **1974**. doi:10.1007/BF00388387.
- SABATINI, S.; BEIS, D.; WOLKENFELT, H.; MURFETT, J.; GUILFOYLE, T.; MALAMY, J.; BENFEY, P.; LEYSER, O.; BECHTOLD, N.; WEISBEEK, P. and SCHERES, B. “An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root.” *Cell*, vol. 99: pp. 463–472, **1999**. doi:10.1016/S0092-8674(00)81535-4.
- SATO, A. and YAMAMOTO, K. T. “Overexpression of the non-canonical Aux/IAA genes causes auxin-related aberrant phenotypes in *Arabidopsis*”. *Physiologia Plantarum*, vol. 133: pp. 397–405, **2008**. doi:10.1111/j.1399-3054.2008.01055.x.
- SCHAEFFER, S. E. “Graph clustering”. *Computer Science Review*, vol. 1(1): pp. 27–64, **2007**. doi:10.1016/j.cosrev.2007.05.001.
-

- SCHENA, M.; SHALON, D.; DAVIS, R. W. and BROWN, P. O. "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." *Science (New York, N.Y.)*, vol. 270(5235): pp. 467–70, **1995**.
- SCHINDELIN, J.; ARGANDA-CARRERAS, I.; FRISE, E.; KAYNIG, V.; LONGAIR, M.; PIETZSCH, T.; PREIBISCH, S.; RUEDEN, C.; SAALFELD, S.; SCHMID, B.; TINEVEZ, J.-Y.; WHITE, D. J.; HARTENSTEIN, V.; ELICEIRI, K.; TOMANCAK, P. and CARDONA, A. "Fiji: an open-source platform for biological-image analysis", **2012**. doi:10.1038/nmeth.2019.
- SCHLEIFENBAUM, F.; ELGASS, K.; SACKROW, M.; CAESAR, K.; BERENDZEN, K.; MEIXNER, A. J. and HARTER, K. "Fluorescence intensity decay shape analysis microscopy (FIDSAM) for quantitative and sensitive live-cell imaging: a novel technique for fluorescence microscopy of endogenously expressed fusion-proteins." *Molecular plant*, vol. 3(3): pp. 555–62, **2010**. doi: 10.1093/mp/ssp110.
- SCHOOF, H.; LENHARD, M.; HAECKER, A.; MAYER, K. F.; JÜRGENS, G. and LAUX, T. "The stem cell population of Arabidopsis shoot meristems is maintained by a regulatory loop between the CLAVATA and WUSCHEL genes." *Cell*, vol. 100: pp. 635–644, **2000**. doi: 10.1016/S0092-8674(00)80700-X.
- SCHRUFF, M. C.; SPIELMAN, M.; TIWARI, S.; ADAMS, S.; FENBY, N. and SCOTT, R. J. "The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs." *Development (Cambridge, England)*, vol. 133(2): pp. 251–61, **2006**. doi:10.1242/dev.02194.
- SESSIONS, A.; NEMHAUSER, J. L.; MCCOLL, A.; ROE, J. L.; FELDMANN, K. A. and ZAMBRYSKI, P. C. "ETTIN patterns the Arabidopsis floral meristem and reproductive organs." *Development (Cambridge, England)*, vol. 124(22): pp. 4481–91, **1997**.
- SHANER, N. C.; CAMPBELL, R. E.; STEINBACH, P. A.; GIEPMANS, B. N. G.; PALMER, A. E. and TSIEN, R. Y. "Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein." *Nature biotechnology*, vol. 22(12): pp. 1567–72, **2004**. doi:10.1038/nbt1037.
- SHIMOMURA, O.; JOHNSON, F. H. and SAIGA, Y. "Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusa, *Aequorea*". *Journal of Cellular and Comparative Physiology*, vol. 59(3): pp. 223–239, **1962**. doi:10.1002/jcp.1030590302.
- SMITH, R. S.; GUYOMARC'H, S.; MANDEL, T.; REINHARDT, D.; KUHLEMEIER, C. and PRUSINKIEWICZ, P. "A plausible model of phyllotaxis." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103(5): pp. 1301–6, **2006**. doi:10.1073/pnas.0510457103.
- SMYTH, D. R.; BOWMAN, J. L. and MEYEROWITZ, E. M. "Early flower development in Arabidopsis." *The Plant cell*, vol. 2(8): pp. 755–67, **1990**. doi:10.1105/tpc.2.8.755.
- STEPANOVA, A. N.; ROBERTSON-HOYT, J.; YUN, J.; BENAVENTE, L. M.; XIE, D. Y.; DOLEŽAL, K.; SCHLERETH, A.; JÜRGENS, G. and ALONSO, J. M. "TAA1-Mediated Auxin Biosynthesis Is Essential for Hormone Crosstalk and Plant Development". *Cell*, vol. 133: pp. 177–191, **2008**. doi:10.1016/j.cell.2008.01.047.
- SZEMENYEI, H.; HANNON, M. and LONG, J. A. "TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis." *Science (New York, N.Y.)*, vol. 319: pp. 1384–1386, **2008**. doi:10.1126/science.1151461.

- TADA, M.; TAKEUCHI, A.; HASHIZUME, M.; KITAMURA, K. and KANO, M. “A highly sensitive fluorescent indicator dye for calcium imaging of neural activity in vitro and in vivo”. *European Journal of Neuroscience*, vol. 39: pp. 1720–1728, **2014**. doi:10.1111/ejn.12476.
- TAN, X.; CALDERON-VILLALOBOS, L. I. A.; SHARON, M.; ZHENG, C.; ROBINSON, C. V.; ESTELLE, M. and ZHENG, N. “Mechanism of auxin perception by the TIR1 ubiquitin ligase.” *Nature*, vol. 446: pp. 640–645, **2007**. doi:10.1038/nature05731.
- TAO, Y.; FERRER, J. L.; LJUNG, K.; POJER, F.; HONG, F.; LONG, J. A.; LI, L.; MORENO, J. E.; BOWMAN, M. E.; IVANS, L. J.; CHENG, Y.; LIM, J.; ZHAO, Y.; BALLARÉ, C. L.; SANDBERG, G.; NOEL, J. P. and CHORY, J. “Rapid Synthesis of Auxin via a New Tryptophan-Dependent Pathway Is Required for Shade Avoidance in Plants”. *Cell*, vol. 133: pp. 164–176, **2008**. doi:10.1016/j.cell.2008.01.049.
- THEISSEN, G. “Development of floral organ identity: stories from the MADS house.” *Current opinion in plant biology*, vol. 4(1): pp. 75–85, **2001**. doi:10.1016/S1369-5266(00)00139-4.
- THOMPSON, J.; HIGGINS, D. and GIBSON, T. “{CLUSTAL W}: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice”. *Nucleic Acids Research*, vol. 22: pp. 4673–4680, **1994**.
- TILNEY-BASSETT, R. A. E. *Plant Chimeras* (Cambridge University Press, **1991**).
- TRAAS, J. and DOONAN, J. H. “Cellular basis of shoot apical meristem development.” *International review of cytology*, vol. 208: pp. 161–206, **2001**.
- ULMASOV, T.; HAGEN, G. and GUILFOYLE, T. J. “Activation and repression of transcription by auxin-response factors.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96(10): pp. 5844–9, **1999a**. doi:10.1073/pnas.96.10.5844.
- ULMASOV, T.; HAGEN, G. and GUILFOYLE, T. J. “Dimerization and DNA binding of auxin response factors.” *The Plant journal : for cell and molecular biology*, vol. 19: pp. 309–319, **1999b**. doi:10.1046/j.1365-313X.1999.00538.x.
- ULMASOV, T.; MURFETT, J.; HAGEN, G. and GUILFOYLE, T. J. “Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements.” *The Plant cell*, vol. 9: pp. 1963–1971, **1997**. doi:10.1105/tpc.9.11.1963.
- VAUGHAN, J. G. “The morphology and growth of the vegetative and reproductive apices of *Arabidopsis thaliana* (L.) Heynh., *Capsella bursa-pastoris* (L.) Medic. and *Anagallis arvensis* L.” *Journal of the Linnean Society of London, Botany*, vol. 55(359): pp. 279–301, **1955**. doi:10.1111/j.1095-8339.1955.tb00014.x.
- VERNOUX, T.; BESNARD, F. and TRAAS, J. “Auxin at the shoot apical meristem.” *Cold Spring Harbor perspectives in biology*, vol. 2: p. a001487, **2010**. doi:10.1101/cshperspect.a001487.
- VERNOUX, T.; BRUNOUD, G.; FARCOT, E.; MORIN, V.; VAN DEN DAELE, H.; LEGRAND, J.; OLIVA, M.; DAS, P.; LARRIEU, A.; WELLS, D.; GUÉDON, Y.; ARMITAGE, L.; PICARD, F.; GUYOMARC’H, S.; CELLIER, C.; PARRY, G.; KOUMPROGLOU, R.; DOONAN, J. H.; ESTELLE, M.; GODIN, C.; KEPINSKI, S.; BENNETT, M.; DE VEYLDER, L. and TRAAS, J. “The auxin signalling network translates dynamic input into robust patterning at the shoot apex.” *Molecular systems biology*, vol. 7(508): p. 508, **2011**. doi:10.1038/msb.2011.39.
- VERNOUX, T.; KRONENBERGER, J.; GRANDJEAN, O.; LAUFS, P. and TRAAS, J. “PIN-FORMED 1 regulates cell fate at the periphery of the shoot apical meristem.” *Development*, vol. 127: pp. 5157–65, **2000**.
-

- WABNIK, K.; KLEINE-VEHN, J.; BALLA, J.; SAUER, M.; NARAMOTO, S.; REINÖHL, V.; MERKS, R. M. H.; GOVAERTS, W. and FRIML, J. “Emergence of tissue polarization from synergy of intracellular and extracellular auxin signaling.” *Molecular systems biology*, vol. 6(447): p. 447, **2010**. doi:10.1038/msb.2010.103.
- WEIGEL, D.; ALVAREZ, J.; SMYTH, D. R.; YANOFSKY, M. F. and MEYEROWITZ, E. M. “LEAFY controls floral meristem identity in Arabidopsis”. *Cell*, vol. 69: pp. 843–859, **1992**. doi:10.1016/0092-8674(92)90295-N.
- WEIJERS, D.; BENKOVA, E.; JÄGER, K. E.; SCHLERETH, A.; HAMANN, T.; KIENZ, M.; WILMOTH, J. C.; REED, J. W. and JÜRGENS, G. “Developmental specificity of auxin response by pairs of ARF and Aux/IAA transcriptional regulators.” *The EMBO journal*, vol. 24: pp. 1874–1885, **2005**. doi:10.1038/sj.emboj.7600659.
- WU, M.-F.; TIAN, Q. and REED, J. W. “Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction.” *Development (Cambridge, England)*, vol. 133: pp. 4211–4218, **2006**. doi:10.1242/dev.02602.
- YADAV, R. K.; PERALES, M.; GRUEL, J.; GIRKE, T.; JÖNSSON, H. and REDDY, G. V. “WUSCHEL protein movement mediates stem cell homeostasis in the Arabidopsis shoot apex.” *Genes & development*, vol. 25(19): pp. 2025–30, **2011**. doi:10.1101/gad.17258511.
- YADAV, R. K.; PERALES, M.; GRUEL, J.; OHNO, C.; HEISLER, M.; GIRKE, T.; JÖNSSON, H. and REDDY, G. V. “Plant stem cell maintenance involves direct transcriptional repression of differentiation program.” *Molecular systems biology*, vol. 9(1): p. 654, **2013**. doi:10.1038/msb.2013.8.
- YADAV, R. K.; TAVAKKOLI, M. and REDDY, G. V. “WUSCHEL mediates stem cell homeostasis by regulating stem cell number and patterns of cell division and differentiation of stem cell progenitors.” *Development (Cambridge, England)*, vol. 137: pp. 3581–3589, **2010**. doi:10.1242/jcs.082420.
- YAMAGUCHI, N.; WU, M. F.; WINTER, C. M.; BERNS, M. C.; NOLE-WILSON, S.; YAMAGUCHI, A.; COUPLAND, G.; KRIZEK, B. A. and WAGNER, D. “A Molecular Framework for Auxin-Mediated Initiation of Flower Primordia”. *Developmental Cell*, vol. 24: pp. 271–282, **2013**. doi:10.1016/j.devcel.2012.12.017.
- YANG, Y.; HAMMES, U. Z.; TAYLOR, C. G.; SCHACHTMAN, D. P. and NIELSEN, E. “High-Affinity Auxin Transport by the AUX1 Influx Carrier Protein”. *Current Biology*, vol. 16: pp. 1123–1127, **2006**. doi:10.1016/j.cub.2006.04.029.
- YOO, T. S.; ACKERMAN, M. J.; LORENSEN, W. E.; SCHROEDER, W.; CHALANA, V.; AYLWARD, S.; METAXAS, D. and WHITAKER, R. “Engineering and algorithm design for an image processing API: A technical report on ITK - The Insight Toolkit”. In *Studies in Health Technology and Informatics*, vol. 85, pp. 586–592 (**2002**). doi:10.3233/978-1-60750-929-5-586.
- YOSHIDA, S.; BARBIER DE REUILLE, P.; LANE, B.; BASSEL, G. W. W.; PRUSINKIEWICZ, P.; SMITH, R. S. S.; WEIJERS, D. and BARBIER DE REUILLE, P. “Genetic Control of Plant Development by Overriding a Geometric Division Rule”. *Developmental Cell*, pp. 1–13, **2014**. doi:10.1016/j.devcel.2014.02.002.
- ZHANG, X. X.; WANG, Z.; YUE, X.; MA, Y.; KIESEWETTER, D. O. and CHEN, X. “pH-sensitive fluorescent dyes: Are they really pH-sensitive in cells?” *Molecular Pharmaceutics*, vol. 10: pp. 1910–1917, **2013**. doi:10.1021/mp3006903.

List of Figures

1	<p>SAM functional and structural organisations in <i>A. thaliana</i>. a. External top view of a –cleaned– SAM displaying the precise spatial disposition of the floral meristems (FM) in green. The SAM is indicated in pink. It is possible to appreciate the three first morphological stages of flower development from the floral initium (i) to a stage 3 FM (largest one at south-east of the picture) by successive counter-clockwise rotations of 137.5°. b. Cross-section view of a SAM highlighting the L1 and L2 tunica layers structure. Floral meristems (FM) and initium (i) can be observed on the sides. AS1, STM, WUS and CLV3 are introduced later. In both figures, CZ refers to the Central Zone and PZ to the Peripheral Zone, i to initium</p>	10
2	<p>Effects of ectopic auxin application (red) on auxin-related mutants in <i>A. thaliana</i>. a, auxin induce lateral organ formation on <i>pid</i> mutant. b, auxin does not induce organ formation on <i>mp</i> mutants. <u>Source:</u> Reinhardt <i>et al.</i> [2003]</p>	12
3	<p>Localisation of PIN1 protein (red) in <i>A. thaliana</i> inflorescence and floral meristems. k, PIN1 localisation in inflorescence meristem (im) cross-section; arrowhead indicate initiation site. l, PIN1 localisation in young (stage 2) FM cross-section; arrowhead indicate sepal initiation site. m, PIN1 localisation in older FM (stage 3) cross-section; arrowhead indicate anthers initiation site. (P) indicate flower primordia, and white arrows indicate vasculature. <u>Source:</u> adapted (white outlines) from Reinhardt <i>et al.</i> [2003]</p>	14
4	<p>Schematic representation of the ARF and Aux/IAA structures as found in <i>Ara-bidopsis thaliana</i>. DBD: DNA binding domain. I: Aux/IAA specific putative homo-dimerisation domain. AD: Activation domain. RD: Repression domain. II: Aux/IAA specific degradation domain. III & IV: protein dimerisation domains. Arrowed lines indicate the extend of each inhibiting ARF structure. <u>Source:</u> adapted from Hagen and Guilfoyle [2002]; Guilfoyle and Hagen [2007] with the notable difference that we found domains III and IV when aligning full length protein sequences for ARF13.</p>	16
5	<p>Model of auxin transduction pathway. Activation and repression activities depend on ARF middle domain amino-acid composition. DBD indicate the DNA binding domain found upstream of auxin-inducible genes. In absence of auxin Aux/IAA are dimerise with the ARFs, preventing them to exert their activating or repressing activity. When auxin is present, it target the Aux/IAA to the proteasome leaving the ARF free to dimerise and exert their regulating activity. <u>Source:</u> adapted from Chapman and Estelle [2009].</p>	17
6	<p>The three first stages of <i>A. th.</i> FM morphogenesis. a. Stage 1, no clear distinction between the FM and the SAM; b. Stage 2, a clear boundary is visible between the SAM and the FM(+26h); c. Early stage 3, sepals start emerging at the sides of the FM (+69h); d. Stage 3, the sepals are clearly visible (+80h). Green fluorescent marker: <i>AHP6::erGFP</i>. <u>Source:</u> Images acquired at the RDP by Pradeep DAS.</p>	21

-
- 7 The ABCE model for floral-organ patterning illustrated with *A. thaliana* flower. **a.** the wild-type display four whorls of organs: the sepals, the petals, the stamens and the carpels. **b.** An *ap2* mutant flower, thus lacking class A genes, show the loss sepals and petals, replaced by carpels and stamens, respectively, subsequent to the expansion of class C genes throughout the flower. **c.** A *pi* mutant flower posses sepals and carpels instead of petals and stamens, respectively, subsequently to the absence of class B genes. **d.** An *ag* mutant flower, lacks class C genes activity, thus displaying sepals, two whorls of petals and further repeating the patterns in interior whorls. **e.** A *SEPALLATA* mutant flower for four *sep* genes (*sep1-4*) present reiterating whorls of leaf-like (le) organs subsequently to the lack of class E genes impairing the activity of ABC class genes, ultimately leading to loss of floral organ determinacy. Source: Krizek and Fletcher [2005]. 22
- 8 Contribution of the SAM histogenetic layers to floral organ tissues. L1 is in green, L2 in red and L3 in blue. The L1 gives rise to the epidermis of all floral organs and to some other tissues. L2 and L3 give rise to the mesophyll and other internal tissues. Source: Jenik and Irish [2000] 24
- 9 Theoretical example of a molecular universe for *in toto* imaging of zebrafish embryogenesis. The ‘xyz’ space correspond to spatial dimension, ‘t’ to temporal dimension and ‘g’ to genomic dimension. Source: Megason and Fraser [2007]. 30
- 10 Illustration of intensity based image (left) transformed into a cell-based image (right). Source: Images and segmentation were generated at the RDP. 34
- 1.1 Flowchart representation of the strategy and main findings. The experimental parts of the work are shown in light green boxes. The connections between the different parts of the work are represented by directed arrows. 42
- 1.2 Spatial regulation of Aux/IAA-ARF signalling in the inflorescence. (**A-D**). Expression patterns of TIR1/AFB F-box co-receptors. Expression was analyzed using GUS translational fusions for TIR1, AFB1 and AFB3 and *in situ* hybridisation for *AFB5*. The relative levels of the protein are indicated for AFB1 and TIR1, as revealed by GUS activity detection time (TIR1: 48h; AFB1: 8h).(**E-ZC**) Expression patterns of *ARFs* and *Aux/IAAs* revealed by RNA *in situ* hybridization. (**ZD**) Detection of *Aux/IAA* and *ARF* expression by RT-qPCR in the inflorescence meristem. The analysis was done in duplicate on meristem mRNAs. Error bars represents the range of values.(**ZE**) Schematic representation of the *Aux/IAAs* and *ARFs* distribution in the meristem. The meristem is represented as a dome (PZ: peripheral zone; CZ: central zone, in grey; SC: stem cells; OC: organizing centre). Global tendency in expression levels are indicated by the size of the + sign. A dashed line was drawn between the upper and lower part of the centre to indicate differences in signalling capacities since *ARF4* and *IAA18* are expressed in the inner core. The primordia (P) have been delineated by a dashed line in the PZ to indicate that several *Aux/IAAs* and *ARFs* shows an even high accumulation in the organ primordia. Several *Aux/IAA* and *ARF* are also more specifically associated with vasculature (V; see main text). Median or near-median sections are shown. Scale bar: 50 μ m. 45
-

- 1.3 Structure the auxin signalling network. **(A)** Visual representation of the Aux/IAA-ARF interactome using Cytoscape (www.cytoscape.org). The proteins are grouped according to their biological identity as indicated. Note the global differences in connectivity of the three biological groups **(B-D)** Connectivity graph and clusters identified by the MixNet algorithm. The probabilities associated with the connectivity structure for the global network are indicated in **(B)**. The three clusters are mainly composed of Aux/IAA (I), ARF activators (II) and ARF repressors (III) as indicated in brackets in **(B)**. The identity of the proteins in these clusters for both the global network **(C)** and the SAM-specific network **(D)** is shown. The proteins are ordered from the most to the least central in each cluster based on the distance of the protein to the cluster. **(E)** The topology of the network relies on stereotypic interaction capacities for the different classes of proteins as represented. ARF+: ARF activators; ARF-: ARF repressors. 47
- 1.4 Mathematical model of the auxin signalling network. **(A)** Reaction scheme considered for the model. The numbers in brackets indicate the 5 populations of molecule described by ODEs in the model. **(B)** Effect of the level of ARF activators (middle panel) and ARF repressors (bottom panel) on target gene induction capacity upon an increase in auxin (upper panel). **(C)** 3D representation of the induction capacity as a function of ARF levels and ARF activator to ARF repressor ratio. The surface has been obtained by calculating the transcription fold-change, i.e. the ratio of the mRNA levels at equilibrium before and after a step increase in auxin. A colour map representing the parameter values is shown. **(D)** Effect of the level of ARF activators (middle panel) and ARF repressors (lower panel) on the stability of target gene induction upon varying auxin level. Transcription in response to sinusoidal changes in auxin levels (upper panel) has been studied. Two situations, corresponding to the centre (CZ) or the periphery (PZ), were considered. The effect of increasing ARF activators was tested for the first situation (CZ + ARF+) and of decreasing ARF repressors for the second (PZ - ARF-). For simplicity mRNA levels are shown in **(B,D)** for only the higher and the lower concentration of the variable parameter used in the simulation. See figures 4 and 5 of Note S1 for the full range of values. 50
- 1.5 Spatial control and dynamics of auxin signalling at the inflorescence meristem. **(A)** Schematic representation of signalling parameters monitored by DII-VENUS as compared to DR5::VENUS. **(B-D)** Expression of DII-VENUS **(B)**, mDII-VENUS **(C)** and DR5::VENUS **(D)** visualized by confocal microscopy. Insets: overlay of the VENUS signal (Green) with the autofluorescence signal (red). In **(B)** and **(D)** the 3 first primordia (P) are indicated and numbered from the youngest to the oldest. Two *initia* (I) are indicated and numbered from the oldest to the youngest following standard nomenclature. **(E)** Auxin-dependent binding of IAA28 domain II to TIR1/AFB auxin co-receptors. Anti-FLAG immunoblots of IAA28 domain II peptide pull-down assay with TIR1-FLAG, AFB1-FLAG or AFB5-FLAG. IAA treatments are as indicated. **(F,G)** Time-course of DII-VENUS **(F)** and DR5::VENUS **(G)** expression followed by confocal microscopy (VENUS fluorescence in green). Images were taken at indicated time after t0. The *initia* (I1 to 3) and the localization of the centre of the meristem (C) are indicated. Scale is the same in all images. Scale bar: 50 μm 52
- S1.1 TIR1/AFB expression patterns in the inflorescence meristem. Serial sections showing expression of TIR1/AFB proteins fused to GUS **(A-D)** or *AFB_{4/5}* mRNA as visualized by RNA *in situ* hybridization **(E,F)**. The name of the gene is indicated on the figure. Scale is identical in all images for each serial section. 79
- S1.2 ARF expression patterns in the inflorescence meristem. **(A-M)** Serial sections showing expression of *ARF* genes as visualized by RNA *in situ* hybridization. The name of the gene is indicated on the figure. Scale is identical in all images for each serial section. 83

- S1.3 Aux/IAA expression patterns in the inflorescence meristem. (**A-L**) Serial sections showing expression of *Aux/IAA* genes as visualized by RNA *in situ* hybridization. The name of the gene is indicated in the figure. For each serial section, scale is identical in all images. 87
- S1.4 *In planta* analysis of Aux/IAA-ARF interactions. Bimolecular Fluorescence complementation (BiFC) was used to test 31 Aux/IAA-ARF interactions *in planta*. The set of interactions included both positive and negative interactions as predicted from the Y2H interactome analysis. (**A**) Summary of the results of the BiFC analysis. All the results were coherent with the results from the Y2H analysis except for 3 interactions (result indicated in bold). (**B**) Confocal images for a selection of the tested interactions showing results of BiFC in *Nicotiana Benthamina* leaves. Yellow: YFP; Red: autofluorescence; Grey levels: transmission channel. Scale is identical in all images. 87
- S1.5 Cluster analysis of the global Aux/IAA-ARF interactome using MixNet. (**A**) Results of the MixNet algorithm for the global interactome. We applied the MixNet algorithm to the ARF-Aux/IAA protein network for $Q = 2$ to 10 clusters. The model selection criterion favours 4 clusters. Since this model selection criterion is only asymptotically valid (i.e. for large N), this number of clusters should only be considered as indicative in our case and we thus explored neighbouring solutions. The results are presented for $Q = 3$ clusters. The proteins are ordered from the most to the least central in each cluster based on the distance of the protein to the cluster ($D(i, q)$; data not shown). The connectivity probability matrix $\hat{\Pi}$ and the cluster distance matrix $\{D(q, \ell); q, \ell = 1, \dots, Q\}$ are given. The three clusters differ strongly in terms of connectivity profiles; compare the rows of matrix $\hat{\Pi}$. The three clusters are also well separated; compare the diagonal elements of the cluster distance matrix $\{D(q, \ell)\}$ corresponding to within-cluster distances to the off-diagonal elements corresponding to between-cluster distances. The proteins in italics form the fourth cluster when the MixNet algorithm is applied for $Q = 4$ clusters. In this case, the 3 most peripheral proteins of cluster I are grouped with the 4 most peripheral proteins of cluster II to form a second Aux/IAA cluster. This fourth cluster is not well defined (the within-cluster distance $D(\text{IV}, \text{IV}) = 0.34$ is larger than the between-cluster distance $D(\text{II}, \text{IV}) = 0.3$ (data not shown), indicating that the 3-cluster solution is the most adequate. (**B**) Visualization of the clusters using MDS. The first two MDS coordinates were deduced from the pairwise distances between proteins $\{D(i, j); i, j = 1, \dots, N\}$. These first two coordinates account for 61.9% of the total variation. The proteins from cluster I, II and III are figured in green, blue and red respectively. Note that the clusters are compact (except for a few outliers) and well separated. The proteins that are in the 4th cluster when the MixNet algorithm is applied for $Q = 4$ are found amongst the outliers (labelled in purple). 88

S1.6 Cluster analysis of the meristem Aux/IAA-ARF interactome using MixNet. (A) Results of the MixNet analysis for the SAM interactome. We applied the MixNet algorithm to the SAM subnetwork for $Q = 3$ clusters. The three clusters obtained are almost nested in the three clusters obtained for the complete network, the only exceptions being IAA29 and IAA26, peripheral in cluster III, which are then assigned to cluster I with almost all the other Aux/IAAs. As in the case of the complete network, the proteins are ordered from the most to the least central in each cluster based on the distance of the protein to the cluster ($D(i, q)$; data not shown) (B). A graphic representation of the structure of the network using Cytoscape (www.cytoscape.org) is given. The Aux/IAA, ARF activator (ARF+) and repressors (ARF-) have been grouped. The vertices are colored according to their MixNet cluster: blue for cluster I, green for cluster II and pink for cluster III. (C) Visualization of the clusters using MDS. The first two MDS coordinates were deduced from the pairwise distances between proteins $\{D(i, j); i, j = 1, \dots, N\}$ and explain 67.9% of the total variation. The proteins assigned to cluster I, II and III are figured in green, blue and red respectively. Note again that the clusters are compact (except for a few outliers) and well separated. 89

S1.7 Distribution of DII-VENUS and mDII-VENUS in vegetative shoot meristem. DII-VENUS (A) and mDII-VENUS (B) expression was analyzed in 5 day old plantlets using confocal microscopy. VENUS fluorescence is in green. The red channel visualizes auto-fluorescence. One cotyledon has been peeled out to reveal the meristem (arrowhead). The two first leaves (L) are indicated. Note the absence of DII-VENUS signal at the centre of the meristem and the homogenous distribution of mDII-VENUS. Scale is identical in (A,B). Scale bar: 40 μm 90

S1.8 Proteasome inhibitors interfere with DII-VENUS auxin-induced degradation in the SAM. Images were obtained by confocal microscopy. DII-VENUS is in green. The red channel visualizes auto-fluorescence in the inflorescence meristem. All the images were taken using the same settings. Plants were treated either with DMSO (control; A), 1 μM IAA (B) or co-treated with IAA and the MG132 proteasome inhibitor (C). Note that the degradation of DII-VENUS induced by auxin is largely blocked by MG132. Scale is the same in (A-C). Scale bar: 40 μm 90

S1.9 Instability of VENUS visualized in DR5::VENUS plants upon chemical inhibition of auxin transport. To verify that the stability of DR5::VENUS fluorescence observed in inflorescence meristem was not solely due to the stability of the VENUS protein, we germinated DR5::VENUS plants on the auxin transport inhibitor NPA to generate pin-like inflorescences in vitro. The plant were then transferred on a new NPA-free medium and DR5::VENUS expression was followed over time as described in Hamant *et al.* [2008]. Due to the perturbation of auxin transport, we often observed in these conditions transient peaks of DR5::VENUS expression that were not followed by organ induction. Representative results from 2 independent time-courses are shown (A,B). Arrows indicates group of nuclei showing rapid decrease in fluorescence. Note that in each cases drastic diminution in VENUS fluorescence could be observed after 4 hours, suggesting a half-life for the protein in that time range. Scale is identical in all images. Scale bar: 20 μm 91

2.1 Model of auxin transduction pathway. Activation and repression activities depend on ARF middle domain amino-acid composition. DBD indicate the DNA binding domain found upstream of auxin-inducible genes. In absence of auxin Aux/IAA are dimerise with the ARFs, preventing them to exert their activating or repressing activity. When auxin is present, it target the Aux/IAA to the proteasome leaving the ARF free to dimerise and exert their regulating activity. Source: adapted from Chapman and Estelle [2009]. 95

2.2	Schematic representation of the ARF and Aux/IAA structures as found in <i>Arabidopsis thaliana</i> . DBD: DNA binding domain. I: Aux/IAA specific putative homodimerisation domain. AD: Activation domain. RD: Repression domain. II: Aux/IAA specific degradation domain. III & IV: protein dimerisation domains. Arrowed lines indicate the extend of each inhibiting ARF structure. Source: adapted from Hagen and Guilfoyle [2002]; Guilfoyle and Hagen [2007] with the notable difference that we found domains III and IV when aligning full length protein sequences for ARF13.	96
2.3	Methodological options to analyse Yeast-2-Hybrid and other interaction data.	96
2.4	Three-component Gaussian mixture model estimated for the the OD ratios. The optical density has here an upper limit of 1.7 for graph readability reasons.	99
2.5	Empirical distribution of the standardised interaction distances for the A, B and C (top-down) networks.	100
S2.1	Histograms of OD ratios per X-Gal marks for the successive marks '+?', '+', '++' and '+++'. The OD ratio values have an upper limit of 1.5 for histogram readability reasons.	120
S2.2	Adjacency matrix of detected interaction (red) the the tresholds: HIS3 > 0,3 and X-Gal > '+?'.	121
S2.3	Adjacency matrix of detected interaction (red) the the tresholds: HIS3 > 0,65 and X-Gal > '+?'.	122
S2.4	Adjacency matrix of detected interaction (red) the the tresholds: HIS3 > 0,45 and X-Gal > '+?'.	123
S2.5	Ranked vertex distances to assigned cluster for network A based on a 4-clusters GM model.	124
S2.6	Ranked vertex distances to assigned cluster for network A based on a 5-clusters GM model.	124
S2.7	Connectivity graph and associated probabilities for the 4-cluster BM model restricted to the ARF+, ARF- and Aux/IAA clusters. The connectivity probability matrix describes the topology of the network at the cluster scale ($\pi_{q\ell}$ is the probability for a protein of cluster q to be connected with a protein of cluster ℓ). This matrix show the heterogeneous behaviour of the 3 clusters corresponding to ARF+, ARF- and Aux/IAA. Instead of classical graph clustering methods which would present highly intra-connected and poorly inter-connected clusters (high probabilities on the diagonal and low elsewhere), the 3-clusters model captured the latent structure of the binary graph. The connectivity probabilities highlight that ARF- are very poorly connected to the rest of the network, ARF+ being on the other hand well-connected to Aux/IAAs, themselves well-connected to each other.	125
S2.8	Ranked vertex distances to assigned cluster center for network A based on a LM model with one explanatory variable.	125
S2.9	Linear regressions for network A based on a 4-clusters model with one explanatory variable for the different pair of groups: ARF+ v.s. ARF+, Aux/IAA v.s. Aux/IAA and ARF+ v.s. Aux/IAA. The outliers cluster is also presented in the graphic.	126
S2.10	Details of the linear regression for network A based on a 4-clusters model with one explanatory variable for the interaction of clusters ARF+ v.s. ARF+.	127
S2.11	Ranked vertex distances to assigned cluster center for network A based on a LM model with two explanatory variables.	128
S2.12	Linear regressions for network A based on a 4-clusters model with one explanatory variable for the different pair of groups: ARF+ v.s. ARF+, Aux/IAA v.s. Aux/IAA and ARF+ v.s. Aux/IAA. The outliers cluster is also presented in the graphic.	129

S2.13	Detail of the multiple alignment curated by Gblocks showing the domains III and IV. Every proteic sequences of the 49 AGI-accessible Aux/IAA and ARF have been used. The blue boxes indicate the conserved sequences found by Gblocks. Coloured letter represent highly conserved amino-acids. The dashes (-) represent gaps created by the multiple-alignment method (CLUSTAL-W) to align conserved segments of the sequence.	130
3.1	<i>A. thaliana</i> early and mature flower structure. a. IM structural organisations, CZ refers to the Central Zone (organising center) and PZ to the Peripheral Zone; i indicate the initium and FM the developing floral meristems. b. Whorled organisation of the flower made of 4 sepals, 4 petals, 6 stamens and a gynoecium made of 2 carpels.	135
3.2	The three first stages of <i>A. th.</i> FM morphogenesis. a. Stage 1, no clear distinction between the FM and the IM; b. Stage 2, a clear boundary is visible between the IM and the FM(+26h); c. Early stage 3, sepals start emerging at the sides of the FM (+69h); d. Stage 3, the sepals are clearly visible (+80h). Vital dye FM4-64 used for wall staining is in red. Green fluorescent marker: <i>AHP6::erGFP</i> . Source: Images acquired at the RDP lab by Pradeep DAS.	136
3.3	Example of segmented and lineaged time-series. Each color giving the lineage from a mother cell at the initial time point passed on to all its daughter up until the last segmented and lineaged time point. Only cells with a complete lineage from the initial to the last time point have been coloured.	140
3.4	Illustration of the spatio-temporal graph creation from segmented and lineaged images. a. Topological graph extraction form segmented images; Coloured circles represent the vertices; Plain green links are the topological relations between cells. b. Spatio-temporal graph framework; Dashed red links are temporal relations between cells (lineage); In black we indicate the graph structured that will receive the cell features introduced later.	141
3.5	2D example of cell vertices used as landmarks (green points) associated between time points t_n (a) -before- and t_{n+1} (b) -after- observed growth. The black lines indicates the cell outlines. Coloured lines indicate the surrounding cells that have divided and the resulting new adjacent walls. This representation is a simplification of a real tissue.	145
3.6	2D example of segmentation error leading to impossibility to associate landmarks at time $t_n: (i, j)$ with those at $t_{n+1}: (k, l)$. Coloured cells indicate their lineage. The black lines indicates the cell outlines. This representation is a simplification of a real tissue.	146
3.7	Example of a segmentation error leading to the indeterminacy of vertices. Red and blue cells in the center of the image (gray dots) are below the epidermal layer. Since the orange cell present a missing portion we cannot detect the vertices marked by white arrows. This representation is a real tissue where only wall associated voxels are presented as coloured points indicating cells.	147
3.8	Example of a segmentation error leading to the indeterminacy of two vertices by presenting only one. The indeterminacy is shown in the yellow dotted ellipse. This representation is a real tissue where only wall associated voxels are presented as coloured cubes indicating cells.	147
3.9	2D example of cell wall median used as landmarks (green points) associated between time points t_n (a) -before- and t_{n+1} (b) -after- observed growth. The black lines indicates the cell outlines. Coloured lines indicate the surrounding cells that have divided and the resulting new adjacent walls. This representation is a simplification of a real tissue.	148

3.10	Illustration of the inertia tensor shape descriptors: detailed illustration of a 2D inertia tensor for a cell wall (a) and the representation of computed 3D inertia tensors at t_4 (b). a. Representation of the 2D inertia tensor with the main and secondary axes in blue and green respectively, obtained using the rank-2 subspace projection (plane defined by the red dots) of the 3D cell wall (gray). b. Inertia tensor with decreasing sorted vectors in red, green and blue; the tensors are centred on the barycentre of the cells. The epidermis cells outlines is given in white.	149
3.11	Detailed illustration of principal curvature descriptors: the principal curvatures of a saddled surface (a). a. Saddle illustrating the principle curvature estimation; the origin point where to do the estimation must be provided; the red lines indicate the direction of maximal curvature in both directions, defining therefore a rank-3 tensor made of the two vectors of principal curvature (blue and green arrows) and the normal vector (black arrow). b. Epidermis curvature computed for a radius of 70 voxels from the epidermis wall median (origin of the vectors); the normal to the surface is not represented here for readability reasons; the epidermis cells outlines are given in white in both figures.	151
3.12	Illustration of four different local sampling radius around the geometric median (green cube) of two epidermal cell walls. The yellow, magenta, blue and red discs are voxels selected at a distance of 60, 70, 80 and 100 voxels respectively.	152
3.13	Examples of spatial differentiation functions <i>Laplacian</i> and <i>mean absolute deviation</i> applied on AHP6 signal (genetic marker) for rank-1 and 2 (topological distance) at t_2 (+44h). Top view projection of a. AHP6 signal feature; b. rank-1 laplacian; c. rank-2 laplacian; d. rank-1 mean absolute deviation; e. rank-2 mean absolute deviation. The colormaps of each illustration (a. to e.) are scale ranging their max and min values over the whole time series.	153
3.14	Examples of temporal differentiation functions <i>temporal change</i> (a) and <i>relative temporal change</i> (b) applied on 'volume' between t_3 and t_4 , represented at t_n and t_{n+1} (same data). When using the relative temporal change instead of the temporal change, the extrema of the observed changes (volumetric growth) are highlighted, here showing the boundary between the central dome of the FM and its emerging sepals. The colormaps for (a) and (b) both range between their respective minima and maxima. a. Observed <i>temporal change</i> in volume. b. Observed <i>relative temporal change</i> in volume.	154
3.15	Examples of temporal differentiation functions <i>temporal rate</i> (a) and <i>log temporal rate</i> (b) applied on 'volume' between t_3 and t_4 , represented at t_n and t_{n+1} (same data). Using the log ratio impose a symmetry between increase and decrease growth rate for temporally differentiated values. The log temporal rate(b) representation seems less interesting, but only because of the biased repartition of the values around the mean value. The colormaps for (a) and (b) both range between their respective minima and maxima. a. Observed <i>temporal rate</i> in volume. b. Observed <i>log temporal rate</i> in volume.	155
3.16	Observed standardized distributions of temporally differentiated volume values computed with the given four temporal differentiation functions. a. Comparing <i>temporal change</i> and <i>relative temporal change</i> highlight the redistribution of the values by dividing by the initial values allowing to compare the effective volumetric growth independently for this initial values. b. Comparing <i>temporal rate</i> and <i>log temporal rate</i> show that the log function restore a symmetry in the distribution allowing for a better relation between increase and decrease.	156
3.17	Example of tranformation of a tensor (3D) shape descriptors into a scalar (1D) shape descriptor for t_4 segmented image. a. Inertia tensor with decreasing sorted vectors in red, green and blue; the tensors are centered on the barycenter of the cells. b. Fractional anisotropy describing shape anisotropy defined in $[0, 1]$	157

3.18 Example of a rank-2 projection (green) computed from the set of voxels defining the real initial shape of the wall (blue). In blue are the voxels defining the true wall. In green are the voxels position computed through rank-2 subspace projection of the blue ones. This representation has been generated from rank-2 computation and representation of real wall from our data. 159

3.19 Stretch matrix 2D illustration with cell wall medians used as landmarks before deformation at t_n (green) and after at t_{n+1} (blue). The red line illustrate the error made by the affine regression (rotation and non-isotropic dilation) that may not be able to match perfectly t_n and t_{n+1} landmarks (obviously in the case of a non-affine deformation). After deformation, the initial centered black cross, now in dashed grey, has been affected by a rotation \vec{R} and a dilation d_1 along the main deformation axis and d_2 along an orthogonal axis to \vec{d}_1 . The black lines does not indicates the cell outlines, but represent the closed shape. Coloured point indicate the wall medians. This representation is a simplification of a real tissue. 160

3.20 Aeral strain rate computational methods comparison between the ‘LTC version’ (eq. 3.12) and the ‘SVD version’ (eq. 3.11). The ‘LTC version’ slightly over-estimating the areal strain rate since using non-flattened surfaces of the external wall of epidermal cells. 161

3.21 Comparison of sibling volumes similarity computed over the whole time-series. **a.** Each coloured point represent the coordinate defined by the two siblings volumes; the blue line represent the coordinates for a perfect symmetry in terms of volume. The abscissa present the sibling with the minimum volume and the ordinate the one with the maximum volume, thus all the information is on the same side of the ‘prefect ratio line’. However, the orthogonal distance to this ‘prefect ratio line’ is not linear for the sibling volumes ratios, since bigger cells can be further away from this line than smaller cells but present the same ratio. I thus coloured the points according to their respective volume ratios and given three others ratios lines: 1/3, 1/2 and 2/3 to help reading the illustration. **b.** Boxplot per time-point (top), distribution (middle) and cumulative distribution (bottom) of the sibling volume ratios over the whole time-series; the vertical lines indicate the percentage of sibling under the curve starting from 1 (the perfect ratio) to the line defined ratio (e.g. 42.6% of the ratios are superior to 0.8). The box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data as follow: $Q3 + 1.5 * IQR$, where $IQR = \text{interquartile range}$ (i.e. $Q3 - Q1$). Flier points are those past the end of the whiskers. 173

3.22 Spatial projection (top and abaxial views) of the siblings volume ratio at $t_n + 1$. In order to be thorough we uses the sublineage (hand-made) to compute volumetric ratios for all cells presenting a division The t4 present a limited amount of data, but this is consistent with the low division number observed between t3 and t4 (see supplementary figure S3.5). 174

3.23 Representation of the sibling volume ratio correlation to **(a.)** the sibling volume sums and **(b.)** the 3D shape anisotropy obtained from the inertia tensor computed with the fused-siblings. 175

3.24 Main axis of inertia and division wall normal vectors correlations at t_{n+1} . Collinear vectors have a value close to 1 and orthogonal vectors close to 0. We use the inertia axis obtained with the fusion of the daughters since the division wall is defined at t_{n+1} . Only the cells presenting a binary division (i.e. only for two siblings) are presented. 176

3.25 Distribution of main axis of inertia and division plane normal vector collinearity computed over the whole time-series and attributed at t_n . Boxplot per time-point (left), distribution (middle) and cumulative distribution (right) of the sibling volume ratios over the whole time-series; the vertical lines indicate the percentage of sibling under the curve starting from 1 (the perfect ratio) to the line defined ratio (e.g. 92.6% of the ratios are superior to 0.8). The box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data as follow: $Q3 + 1.5 * IQR$, where $IQR =$ interquartile range (i.e. $Q3 - Q1$). Flier points are those past the end of the whiskers. 177

3.26 Representation of the vectors collinearity in correlation to **(a.)** the sibling volume sums and **(b.)** the 3D shape anisotropy obtained from the inertia tensor computed with the sibling fused. 177

3.27 Example of morphology related regions defined on the last time point (t4). We manually defined the boundary (dark blue) region between the central dome and the sepals. The central zone (blue) is surrounded by a peripheral zone (light blue). Each sepals are reported in different colors (abaxial: lighth green; adaxial: yellow; left: dark red; right: orange). 178

3.28 Example of features distributions by morphologically defined regions (at t4). . . . 178

3.29 Example of an acquired genetical marker: AHP6 signal (top) and the associated 4-groups clustering by ward method (bottom). In the top row the colormap range from low-intensity signal in dark blue to high-intensity signal in dark red. The highest intensities have not been reproduce for lower signal readability reasons. The threshold used is presented in supplementary figure S3.1. In the bottom row the colormap is made as follow: no-signal = dark blue; low-intensity signal = orange; medium-intensity signal = light green; high-intensity signal = light blue. 182

3.30 Cell features boxplot per groups obtained from 4-clusters Ward clustering of AHP6 signal. We give the distribution of AHP6 signal, 3D shape anisotropy, volume, log volumetric growth rate and division rate. It seems that the log volumetric growth rate and division rate are related to AHP6 signal level since both show an increase relate to the increase of signal. Volume property seems inversely proportional to AHP6 signal levels, but it might be a secondary effect of increased growth and division rate. 182

3.31 Spatial projection of the clustering obtained using a 4-cluster Ward model based on ‘epidermis area’ feature **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**. . 185

3.32 Ordered vertex distance to their cluster center **(a)** and clusters feature distributions **(b)** obtained using a 4-cluster Ward model based on ‘epidermis area’ features. . . . 185

3.33 Spatial projection of the clustering obtained using a 4-cluster Ward model based on ‘volumetric log temporal rate’ feature at t_{n+1} **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**. 186

3.34 Ordered vertex distance to their cluster center **(a)** and clusters feature distributions **(b)** obtained using a 4-cluster Ward model based on ‘volumetric log rate’ features. . 187

3.35 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 80% ‘volumetric log temporal rate’ + 20% ‘topology’ features features **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**. 188

3.36 Ordered vertex distance to their cluster center **(a)** and clusters feature distributions **(b)** obtained using a 4-cluster Ward model based 80% ‘volumetric log temporal rate’ + 20% ‘topology’ features. 188

3.37 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 50% ‘volumetric log temporal rate’ + 50% ‘Gaussian curvature’ features **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**. 189

3.38	Ordered vertex distance to their cluster center (a) and clusters feature distributions (b) obtained using a 4-cluster Ward model based on 50% ‘volumetric log temporal rate’ + 50% ‘Gaussian curvature’ features.	189
3.39	Spatial projection of the clustering obtained using a 5-cluster Ward model based on 30% ‘volumetric log temporal rate’ + 30% ‘Gaussian curvature’ + 30% ‘AHP6’ + 10% ‘spatial topology’ features (a-e,g-j), and the cluster distances associated heatmap (f).	191
3.40	Clusters feature distributions comparison for the 4- (a) and 5-cluster (b) Ward model based on 30% ‘volumetric log temporal rate’ + 30% ‘Gaussian curvature’ + 30% ‘AHP6’ + 10% ‘spatial topology’ features. In the 5-cluster model, cluster 1 and 4 have been obtained splitting cluster 1 from the 4-cluster model.	191
3.41	Spatial projection of the four states obtained using a multivariate hidden multi-type branching process computed using the epidermis area, the internal area (i.e. between the L1 and L2), the volumes, the Gaussian curvature, the shape anisotropy and the AHP6 signal. State 0 is in dark blue, state 1 in light blue, state 2 in yellow and state 3 in dark red. Sepals are mostly identified by considering state 0 and 3, the dome by state 2 and boundary cells by state 1.	195
3.42	Observed histogram and mixture of observed distributions for each structuring feature. State 0 is in dark blue, state 1 in light blue, state 2 in yellow and state 3 in dark red (same than in the spatial projections). Areas and volumes have been modelled by Gamma distributions and curvatures by Gaussian distribution. Combining separations induced by areas and volume in one hand and curvatures in the other hand indicates that states are well separated using only these features.	196
3.43	Representation of the generation distributions by states. Generations distribution relates to the evolution to the cell state in time, meaning the state of its descendants (either binary or no division). For graph readability reasons, only generation laws with a probability higher than 0.1 where represented. A single line indicates generation law of undivided cell when double lines and two parts arrows indicate a binary division. The thickness of the line is proportional to its associated generation probability, which are also given as plain numbers. The state colors are the same than in the spatial projection and the distribution histograms.	197
S3.1	Distribution of AHP6 signal by time point. The red dots on the boxplot (right) and upper limits on the histogram (left) give the outliers positions estimated from a <i>mean absolute deviation</i> estimator.	206
S3.2	Main axis of inertia and division wall normal vectors correlations projected at t_n . Collinear vectors have a value close to 1 and orthogonal vectors close to 0. We use the inertia axis obtained with the fusion of the daughters since the division wall is defined at t_{n+1} . To be coherent with the assumption of a ‘small deformation’ between successive time-points, we only represented cells presenting a binary division (i.e. only for two siblings).	206
S3.3	Spatial projection (top and abaxial views) of the siblings volume ratio at $t_n + 1$. In order to be thorough we uses the sublineage (hand-made) to compute volumetric ratios for all cells presenting a division The t4 present a limited amount of data, but this is consistent with the low division number observed between t3 and t4 (see supplementary figure S3.5).	207
S3.4	Main axis of inertia and division wall normal vectors correlations projected at t_{n+1} . Collinear vectors have a value close to 1 and orthogonal vectors close to 0. We use the inertia axis obtained with the fusion of the daughters since the division wall is defined at t_{n+1} . Only the cells presenting a binary division (i.e. only for two siblings) are presented.	207
S3.5	Top and abaxial view of the division rates projected at t_{n+1}	208

S3.6 Examples of spatial differentiation functions *Laplacian* and *mean absolute deviation* applied on AHP6 signal (genetic marker) for rank-1 and 2 (topological distance). 209

S3.7 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 90% ‘epidermis area’ + 10% ‘spatial topology’ features, **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**, ordered vertex distance to their cluster center **(k)** and clusters feature distributions **(l)**. 210

S3.8 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 80% ‘epidermis area’ + 20% ‘spatial topology’ features, **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**, ordered vertex distance to their cluster center **(k)** and clusters feature distributions **(l)**. 210

S3.9 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 70% ‘epidermis area’ + 30% ‘spatial topology’ features, **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**, ordered vertex distance to their cluster center **(k)** and clusters feature distributions **(l)**. 211

S3.10 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 60% ‘epidermis area’ + 40% ‘spatial topology’ features, **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**, ordered vertex distance to their cluster center **(k)** and clusters feature distributions **(l)**. 211

S3.11 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 90% ‘volumetric log rate’ + 10% ‘spatial topology’ features, **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**, ordered vertex distance to their cluster center **(k)** and clusters feature distributions **(l)**. 212

S3.12 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 80% ‘volumetric log rate’ + 20% ‘spatial topology’ features, **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**, ordered vertex distance to their cluster center **(k)** and clusters feature distributions **(l)**. 212

S3.13 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 70% ‘volumetric log rate’ + 30% ‘spatial topology’ features, **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**, ordered vertex distance to their cluster center **(k)** and clusters feature distributions **(l)**. 213

S3.14 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 60% ‘volumetric log rate’ + 40% topological features, **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**, ordered vertex distance to their cluster center **(k)** and clusters feature distributions **(l)**. 213

S3.15 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 30% ‘volumetric log temporal rate’ + 30% ‘Gaussian curvature’ + 30% ‘AHP6’ + 10% ‘spatial topology’ features, **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**, ordered vertex distance to their cluster center **(k)** and clusters feature distributions **(l)**. 214

S3.16 Spatial projection of the clustering obtained using a 4-cluster Ward model based on 30% ‘volumetric log temporal rate’ + 30% ‘Gaussian curvature’ + 30% ‘AHP6’ + 10% ‘spatial topology’ features, **(a-e,g-j)**, and the cluster distances associated heatmap **(f)**, ordered vertex distance to their cluster center **(k)** and clusters feature distributions **(l)**. 214

S3.17 Backward temporal projections of two manually defined regions: boundary and central zone. On the left we can observe the backward projection of the regions onto the 4 time points preceding t4 presented in figure 3.27. on the right are represented the distributions of the volumetric growth rate for those two regions. In both case, we can observe a dramatic reduction of the growth rate over time, with the most significant reduction between t4 and t5. The time points are here numbered from t1 to t5 unfortunately. 215

List of Tables

S1.1	Results of the yeast two-hybrid analysis.	72
S1.2	Summary of Aux/IAA-ARF interactions tested in the literature.	72
S1.3	Oligonucleotides used in this study.	77
2.1	Example of Yeast-2-Hybrid data, with the name of the tested proteins, the side they were attached to and the each output provided by the reporter genes.	97
2.2	ICL and corresponding posterior model probabilities for the BM model.	102
2.3	Composition of the four clusters obtained using the BM model. The ARF activators are in bold. The distance $D(i, q)$ between protein i and cluster q to which its is assigned is given for the most central, the most peripheral and some other proteins of interest for interpretation.	102
2.4	ICL and corresponding posterior model probabilities for network A based on a GM model.	102
2.5	ICL and corresponding posterior model probabilities for network B based on a GM model.	102
2.6	Composition of the four clusters obtained using the GM model applied to network A. The ARF activator are in bold. The proteins that are attributed to two clusters in the 5-clusters GM model are in blue and cyan. The distance $D(i, q)$ between protein i and its cluster q is given for the most central, the most peripheral and some other proteins of interest for interpretation purposes. See supplementary figures S2.5 and S2.6 for the distance plot.	103
2.7	ICL and corresponding posterior model probabilities for network A based on a LM model with III & IV as a single explanatory variable.	106
2.8	Proteins clustering for network A based on a 4-clusters LM model with one explanatory variable. The distance $D(i, q)$ between protein i and cluster q to which its is assigned is given for the most central, the most peripheral and some other proteins of interest for interpretation purposes. See supplementary figure S2.8 for the distance plot.	106
2.9	ICL criterion and corresponding posterior model probabilities for network A based on a LM model with two explanatory variables.	108
2.10	Proteins clustering for network A based on a 4-clusters LM model with two explanatory variables. The distance $D(i, q)$ between protein i and cluster q to which its is assigned is given for the most central, the most peripheral and some other proteins of interest for interpretation purposes. See supplementary figure S2.11 for the distance plot.	108
2.11	Cluster composition matching for 4-clusters models.	110
2.12	Table for decision rule 1.	114
2.13	Tables for decision rule 2.	114
2.14	Tables for decision rule 3.	115
2.15	Tables for decision rule 4.	115
2.16	Tables for decision rule 5.	115
S2.1	Distance between clusters matrix $D(q, \ell)$ for network A based on a GM model. . . .	121

S2.2 Estimated model parameters for the 4-cluster BM model. $\hat{\alpha}$ are the cluster proportions and $\hat{\Pi}$ is the connectivity probability matrix. 122

S2.3 ICL and corresponding posterior model probabilities for network B based on a LM model with III & IV as one global explanatory variable. 122

3.1 Examples of basic spatial information extractable from segmented image. 143

