Quantitative analysis of animal morphogenesis : from high-throughput laser imaging to 4D virtual embryo in ascidians

Léo Guignard

Collège Doctoral du Languedoc-Roussillon

THÈSE

Pour obtenir le grade de

Docteur

Délivré par l'**Université de Montpellier**

Préparée au sein de l'école doctorale **CBS2**
Et des unités de recherche **CRBM UMR 5237**
et **EP Inria Virtual Plants**

Spécialité: **Biologie cellulaire**

Présentée par **Léo GUIGNARD**

Analyse quantitative de la morphogenèse animale: de l'imagerie laser haut débit à l'embryon virtuel chez les ascidies

Soutenue le 09 Décembre 2015 devant le jury composé de

Prof. María J. Ledesma-Carbayo	DIE UP de Madrid	Rapporteur
Dr. Hitoyoshi Yasuo	LBDV Villefranche s/mer	Rapporteur
Prof. William Puech	LIRMM Montpellier	Examinateur
Prof. Carl-Philipp Heisenberg	IST Austria Klosterneuburg	Examinateur
Dr. Charles Kervrann	Inria Rennes	Examinateur
Dr. Édouard Bertrand	IGMM Montpellier	Examinateur
Dr. Patrick Lemaire	CRBM Montpellier	Directeur de thèse
Dr. Christophe Godin	Inria Montpellier	Co-directeur de thèse
Dr. Grégroire Malandain	Inria Sophia-Antipolis	Membre invité

Université Montpellier

"Lay back, relax and talk mathematics"

Tyga - Hookah

Remerciements

Quatre années de thèse, commençant informaticien, terminant biologiste, au moins sur le papier. Beaucoup de chemin parcouru, je ne pense pas avoir autant appris que durant ces quatre années. Tout ça ne s'est évidemment pas fait tout seul, j'aimerai remercier les personnes qui m'ont permis de naviguer entre le traitement d'image, la modélisation et la biologie.

Je voudrais tout d'abord remercier mes trois encadrants, Christophe Godin, Patrick Lemaire et Grégoire Malandain. Vous avez su me faire confiance, j'ai pu m'exprimer comme je l'aimais, j'ai pu faire ce qu'il me plaisait. Il m'a été très agréable de pouvoir arpenter ce chemin à vos côtés.

Je voudrais aussi remercier Hitoyoshi Yasuo et María Ledesma-Carbayo pour avoir accepté de relire et corriger ma thèse ainsi que William Puech, Carl-Philipp Heisenberg et Charles Kervrann pour avoir participé au jury de thèse. Édouard Bertrand pour avoir accepté de présider le jury de ma soutenance et pour avoir fait parti de mon comité de suivi de thèse et finalement, merci à Laure Blanc-Feraud et Pierre-François Lenne pour avoir participé a mon comité de suivi de thèse.

J'ai passé de très bons moments dans les deux équipes de Christophe et Patrick, ça n'aurait pas été le cas sans tous leurs membres. Je voudrais particulièrement remercier Eugenio, Jonathan, Olivier, les Juliens, Guillaume, Jean-Philippe, Christophe P., Yann, Romain, Yoan, Matthieu et Matija. Merci aussi à Laurence pour toute l'aide que tu m'as apporté lorsqu'il fallait que je parte en mission. Je voudrait remercier particulièrement Fred, Emmanuel et Ibrahim pour toutes les discussions professionnelles ou non qu'on a partagé.

Je voudrais aussi remercier mes amis avec qui j'ai passé beaucoup de temps quand je n'étais pas à travailler. Jörg et Sabrina, j'ai particulièrement aimé passer du temps avec vous. Pierre et Jean, partenaires dans la pause café et la détente de fin de journée. J'ai toujours aimé discuter avec vous, autour d'une bière plus particulièrement ! Thibault et Cédric, merci à tout les deux, merci pour le bon temps ! Merci aussi à mes amis de Bordeaux, Marc, Laura, Marylis, Gaultier, Geoffrey et surtout Romain qui a réussi à se déplacer pour assister à ma soutenance thèse.

Merci aussi à ma famille pour m'avoir toujours soutenu et pour toujours avoir cru en moi. Merci à mon parrain et à mes sœurs. Merci à mon père et à ma mère, merci de vos conseils, de votre écoute et compréhension. Merci d'avoir fait le déplacement pour assister à ma soutenance de thèse.

Finalement, je voudrais terminer en remerciant Alicia, merci pour ta patience, pour tes concessions. Merci aussi pour m'avoir soutenu, écouté, encouragé, remonté le moral quand il le fallait. Merci pour m'avoir accompagné durant cette fin de thèse. Merci pour m'avoir inspiré.

Résumé

Les embryons d'ascidies se développent avec un lignage cellulaire stéréotypé et évolutionairement conservé pour produire en quelques heures ou jours un têtard comportant un petit nombre de cellules. De ce fait, ils fournissent cadre intéressant pour décrire avec une résolution cellulaire le programme de développement d'un organisme complet. Pendant mon doctorat, j'ai développé une approche quantitative pour décrire l'évolution morphologique embryonnaire pendant le développement de Phallusia mammillata. J'ai ensuite utilisé cette approche pour systématiquement caractériser en détail les logiques des événements de spécifications de destin cellulaire.

Pour caractériser quantitativement les comportements cellulaires pendant l'embryogenèse, nous avons utilisé de la microscopie à feuille de lumière multi-angles pour imager des embryons entiers à haute résolution spatio-temporelle. Les membranes plasmiques étaient marquées pour permettre l'identification des cellules. Pour extraire les informations biologiques de ce jeu de donnés, j'ai développé une nouvelle méthode pour segmenter les cellules en 4D, ASTEC. Une fois appliquée aux embryons de Phallusia mammillata imagés pendant 6 heures entre le stade 64 cellules et le début des stades bourgeon caudal, cette méthode a permis de récupérer la forme et de suivre 1030 cellules pendant 640 divisions. L'embryon digital 4D résultant peut être formalisé par un graphe dynamique, dans lequel les cellules sont représentées par des sommets reliés par des arrêtes représentant au sein d'un point de temps leur voisinage spatial, et entre différents points de temps leur lignage cellulaire.

Basé sur cette représentation digitale et quantitative, nous avons systématiquement identifié les événements de spécification cellulaire jusqu'au dernier stade de la gastrulation. Des simulations informatiques ont révélé que des règles remarquablement simples intégrant les aires de contacts cellulaires et les expressions spatio-temporelles booléennes de signaux moléculaires extracellulaires sont suffisantes pour expliquer les inductions cellulaires au cours du développement précoce. Ce travail suggère que pour les embryons établissant des contacts stéréotypés et précis entre cellules voisines, les contraintes génomiques sont relâchées, ce qui permet une évolution plus rapide du génome.

Titre Analyse quantitative de la morphogenèse animale : de l'imagerie laser haut-débit à l'embryon virtuel chez les ascidies

Résumé

Les embryons d'ascidies se développent avec un lignage cellulaire stéréotypé et évolutionairement conservé pour produire en quelques heures ou jours un têtard comportant un petit nombre de cellules. De ce fait, ils fournissent cadre intéressant pour décrire avec une résolution cellulaire le programme de développement d'un organisme complet. Pendant mon doctorat, j'ai développé une approche quantitative pour décrire l'évolution morphologique embryonnaire pendant le développement de Phallusia mammillata. J'ai ensuite utilisé cette approche pour systématiquement caractériser en détail les logiques des événements de spécifications de destin cellulaire.

Pour caractériser quantitativement les comportements cellulaires pendant l'embryogenèse, nous avons utilisé de la microscopie à feuille de lumière multi-angles pour imager des embryons entiers à haute résolution spatio-temporelle. Les membranes plasmiques étaient marquées pour permettre l'identification des cellules. Pour extraire les informations biologiques de ce jeu de donnés, j'ai développé une nouvelle méthode pour segmenter les cellules en 4D, ASTEC. Une fois appliquée aux embryons de Phallusia mammillata imagés pendant 6 heures entre le stade 64 cellules et le début des stades bourgeon caudal, cette méthode a permis de récupérer la forme et de suivre 1030 cellules pendant 640 divisions. L'embryon digital 4D résultant peut être formalisé par un graphe dynamique, dans lequel les cellules sont représentées par des sommets reliés par des arrêtes représentant au sein d'un point de temps leur voisinage spatial, et entre différents points de temps leur lignage cellulaire.

Basé sur cette représentation digitale et quantitative, nous avons systématiquement identifié les événements de spécification cellulaire jusqu'au dernier stade de la gastrulation. Des simulations informatiques ont révélé que des règles remarquablement simples intégrant les aires de contacts cellulaires et les expressions spatio-temporelles booléennes de signaux moléculaires extracellulaires sont suffisantes pour expliquer les inductions cellulaires au cours du développement précoce. Ce travail suggère que pour les embryons établissant des contacts stéréotypés et précis entre cellules voisines, les contraintes génomiques sont relâchées, ce qui permet une évolution plus rapide du génome.

Mots-clés Développement ; Segmentation ; Suivi cellulaire ; Atlas 4D ; Ascidies

Title Quantitative analysis of animal morphogenesis: from high-throughput laser imaging to 4D virtual embryo in ascidians

Abstract

Ascidian embryos develop with stereotyped and evolutionarily conserved invariant cell lineages to produce in a few hours or days tadpole larvae with a small number of cells. They thus provide an attractive framework to describe with cellular resolution the developmental program of a whole organism. During my PhD, I developed a quantitative approach to describe the evolution of embryonic morphologies during the development of the ascidian *Phallusia mammillata*. I then used this approach to systematically characterize in detail the logic of cell fate induction events.

To quantitatively characterize cell behaviors during embryogenesis, we used multi-angle light-sheet microscopy to image with high spatio-temporal resolution entire live embryos with fluorescently labeled plasma membranes. To extract biological information from this imaging dataset, I then developed a conceptually novel automated method for 4D cell segmentation, ASTEC. Applied to a *Phallusia mammillata* embryo imaged for 6 hours between the 64-cell and the initial tailbud stages, this method allows the accurate tracking and shape analysis of 1030 cells across 640 cell divisions. The resulting 4D digital embryo can be formalized as a dynamic graph, in which cells are represented by nodes, linked within a time point by edges that represent their spatial neighborhood, and between time points by temporal edges describing cell lineages.

Based on this quantitative digital representation, we systematically identified cell fate specification events up to the late gastrula stage. Computational simulations revealed that remarkably simple rules integrating measured cell-cell contact areas with boolean spatio-temporal expression data for extracellular signalling molecules are sufficient to explain most early cell inductions. This work suggests that in embryos establishing precise stereotyped contacts between neighboring cells, the genomic constraints for precise gene expression levels are relaxed, thereby allowing rapid genome evolution.

Keywords Development; Segmentation; Cell Tracking; Atlas 4D; Ascidians

Virtual Plants
Campus St Priest
860 rue de St Priest, Bat. 5
39095 Montpellier Cedex 5, France

CRBM
1919, route de Mende
34293 Montpellier Cedex 5, France

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Contents

1.1 Understanding embryo development

Understanding the mechanisms that lead to the formation of a complex organism from a simple egg is a question that has long attracted the attention of philosophers and scientists. As early as the 5th century BC in Greece, Hippocrates made one of the first recorded observations on the heredidity of developmental traits [Stent, 1971]. Nowadays, developmental processes are considered to be one of the greatest results of Evolution: how can arms, wings, eyes, heart and brain be formed from one fertilized egg?

1.1.1 From preformation to epigenetics

In the IVth century BC, Aristotle suggested that organs could be formed from a fertilized egg in two contrasted ways [Wolpert and Tickle, 2011; Gilbert, 2006]. According to the philosopher, organisms can either be already preformed in one of the parents or be the result of the progressive formation of the different organs. Aristotle called these two hypotheses *preformation* and *epigenesis*, respectively.

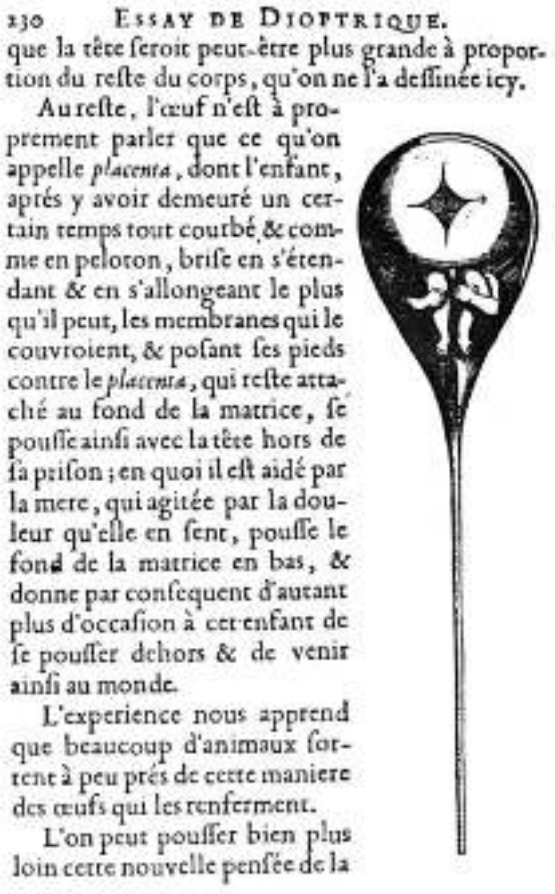

230 ESSAY DE DIOPTRIQUE.
que la tête ſeroit peut-être plus grande à proportion du reſte du corps, qu'on ne l'a deſſinée icy.
Au reſte, l'œuf n'eſt à proprement parler que ce qu'on appelle *placenta*, dont l'enfant, aprés y avoir demeuré un certain temps tout courbé & comme en peloton, briſe en s'étendant & en s'allongeant le plus qu'il peut, les membranes qui le couvroient, & poſant ſes pieds contre le *placenta*, qui reſte attaché au fond de la matrice, ſe pouſſe ainſi avec la tête hors de ſa priſon; en quoi il eſt aidé par la mere, qui agitée par la douleur qu'elle en ſent, pouſſe le fond de la matrice en bas, & donne par conſequent d'autant plus d'occaſion à cet enfant de ſe pouſſer dehors & de venir ainſi au monde.
L'experience nous apprend que beaucoup d'animaux ſortent à peu prés de cette maniere des œufs qui les renferment.
L'on peut pouſſer bien plus loin cette nouvelle penſée de la

Figure 1.1 – Illustration of homunculi in sperm, drawn by Hartsoeker in 1695.

In the XVIIth century, Aristotle's view was still predominant and shaping the debate among scientists. Most were prone to preformation and for them, organisms would already have their final form inside either the oocyte or the spermatozoon. They thought that only growth and unfolding could happen during the development of an organism (Figure 1.1). As a logical consequence of this view, an organism should contain all its future descendants. Going one step further, this means that the first human contained all humanity. This vision fitted well with the entanglement of science and religion that prevailed at that time and with the belief that all living organisms had been already created by God at the time of the creation. Preformation also implies that it is either the male or the female that contains the preformed organism. Therefore, only one of the parents will contribute to the traits of the future organism. By contrast, few scientists at the time were in favour of epigenesis

which required the existence of a force organizing the drive of organ formation. Lack of evidence for such a force provided strong arguments against this theory. The debate continued to rage on throughout the 18th century.

It is during the XIXth century that this debate was finally settled thanks to the invention of the microscope by Zacharias Jansen and its use by Robert Hooke that led to the discovery of cells. The cell theory, initially developed by Matthias Schleinder and Theodor Schwann between 1820 and 1880, had a major impact on the preformationism versus the epigenesis debate. This theory proposed that the cell is the unit of life. It states too that every living organism is composed of cells that arise from the division of already existing cells and that the first cell is the egg. Thus, complex multicellular organisms, such as animals, are not preformed. Based on this new multicellular view of development, epigenesis could now satisfactorily explain various aspects of development. By dividing, changing their shape and moving, the cells that constitute the embryo would eventually form the final organism. It is a combination of cellular changes that drives the development of the embryo. Understanding what these cellular changes are, how they occur and how they are orchestrated should allow to decipher the morphogenesis of an organism, that is how it is made.

1.1.2 Cell specification

To form the various organs of an organism from a fertilized egg, cells need to divide and to differentiate as they fulfill different functions in the organism. For example muscle cells contract, epidermal cells protect, pancreatic cells secrete. These different properties are progressively acquired by cells throughout the course of cell fate specification events. In animal embryos, three germ layers are first to be defined; these are the endoderm (mostly digestive tissues), the mesoderm (blood, muscle and various internal organ tissues) and the ectoderm (skin, nervous tissues and most head structures). Each germ layer is then regionalized into individual tissues, and within these tissues, individual cell types are defined. This process usually takes place rapidly, from a few hours in most invertebrates and anamniote vertebrates to a few days in mammals and birds. The cell fate decisions that are made result in a progressive loss of pluripotency of the cells (an egg can generate all types of cells while ectodermal cells are restricted to only ectodermal derivatives).

Cell fate specification is often the result of communication between neighbouring cells. The fate of a cell is thus defined both by its ancestry (the signals it can respond to) and by its context (which defines the signals the cell is exposed to). Cell communication is in most cases mediated by secreted or transmembrane proteins belonging to a small number of protein families which usually bind to membrane receptors.

For example, the notochord cells and posterior neural plate cells in the embryo of the ascidian *Ciona* arise from a common progenitor and are specified during the division between the 32 and the 64 cell stage of development as a result of the action

of the FGF9/16/20 inducer [Picco et al., 2007]. These cells give rise to different organs and will therefore adopt different behaviours as a result of their specification. The neural plate elongates as a consequence of the preferential antero-posterior division of its constituent cells [Nicol and Meinertzhagen, 1988; Lemaire, 2009]. By contrast, the notochord elongation is a result of cell intercalation [Munro and Odell, 2002].

1.1.3 Morphogenesis as a result of changes in cellular organisation

It is not enough to specify cell fates. It is also important for the cells to be positioned in the correct local neighbourhood (for example, to allow the precise communications and further specification events) and in the correct global neighbourhood to then later organise the organs in their appropriate place and position. Development is therefore not only the result of cell fate specifications but also of the orchestration of the behaviour of cells or group of cells, a process called morphogenesis.

In animal embryos for example, the first major morphogenetic event is gastrulation which takes place immediately after the definition of the three germ layers. Gastrulation leads to the internalization of the mesoderm and endoderm and to the definition of the final antero-posterior, dorso-ventral and left-right axes of the developing organism. In vertebrates, gastrulation is followed by the formation of the neural tube and of the brain vesicles, a process called neurulation. Then, tail elongation is performed.

These morphogenetic events are driven by combinations of cellular changes which are usually classified in 5 categories.

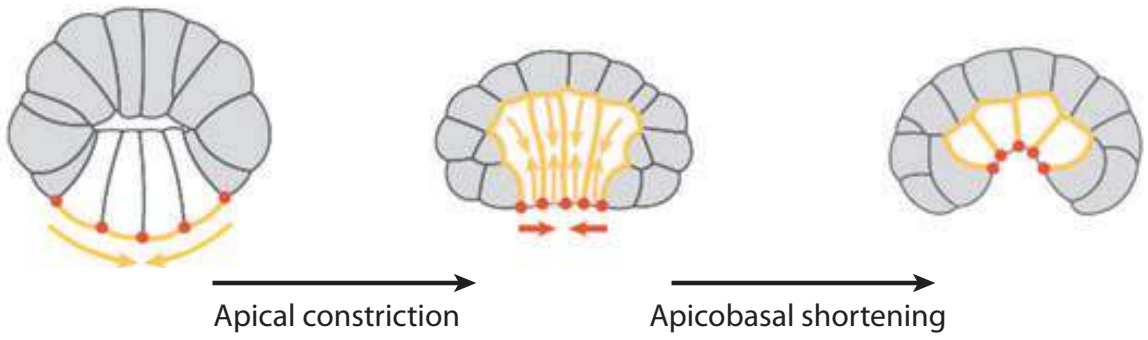


Figure 1.2 – The evolution of endoderm precursor cell shapes during gastrulation of Ciona intestinalis. The vegetative part of the embryo is to the bottom. First, in the round embryo, the endodermal cells (light grey) apically constrict, inducing a flattening of the vegetal pole of the embryo. Then, endodermal cells shorten apicobasally inducing their own internalisation and pulling their neighbouring cells. (adapted from Lecuit et al. [2011])

Isovolumetric shape changes. The shape of cells is determined by two combined effects [Lecuit and Lenne, 2007; Paluch and Heisenberg, 2009]. The first effect is the tension at the surface of the cell. This force minimises the cell surface and therefore drives the cell shape towards roundness. The second effect is cell adhesion between cells, which tends to maximise the surface of contact between cells and, thus, acts against the roundness of a cell. As a first approximation, the global shape of a cell in an organ is determined by the balance between these two seemingly opposing processes. In the case of the invagination of the endoderm of *Ciona intestinalis*, the surface tension of the endodermal cells is locally modified by a change in the activation status of the Myosin on different cell-cell interfaces (in yellow in Figure 1.2). As a result, endodermal cells autonomously change their shape and pull on the abutting epidermal and mesodermal cells (grey cells), thus inducing the global shape change of the embryo.

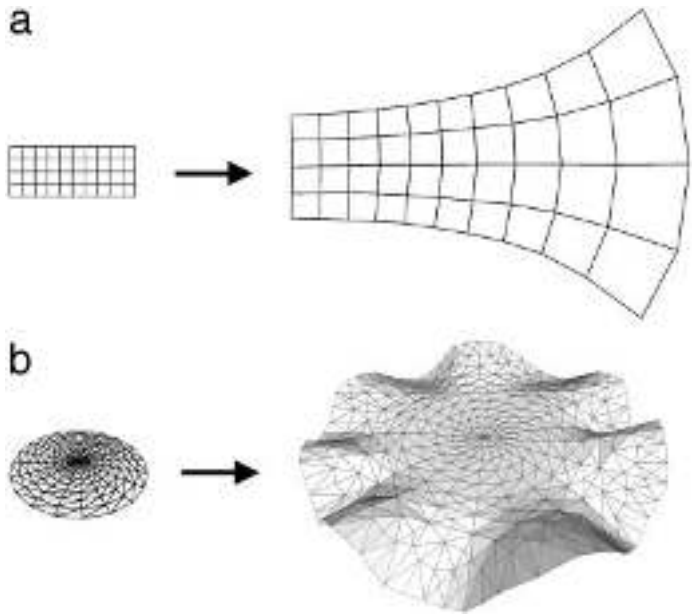


Figure 1.3 – Isotropic growth rate example. (a) Isotropic growth rate increasing from left to right. (b) Flat disk with isotropic growth rate greater at the margins than at the center. (adapted from Coen et al. [2004])

Cell growth. Cells can change their shape through growth. Cell growth happens during the development of most animals. This does not, however, occur in embryos that develop outside their mother and without external source of nutriments (including *Drosophila*, ascidians, sea urchins). Cell growth is particularly conspicuous in plants where it is the major driver of morphogenesis in meristematic tissues. In *Arabidopsis thaliana* the regulation of cell growth is thought to be the major actor of the elongation of the apical meristem, the stem cell structure which generates aerial plant structures [Hamant et al., 2008]. In this case, growth is anisotropic; it is the principal direction of

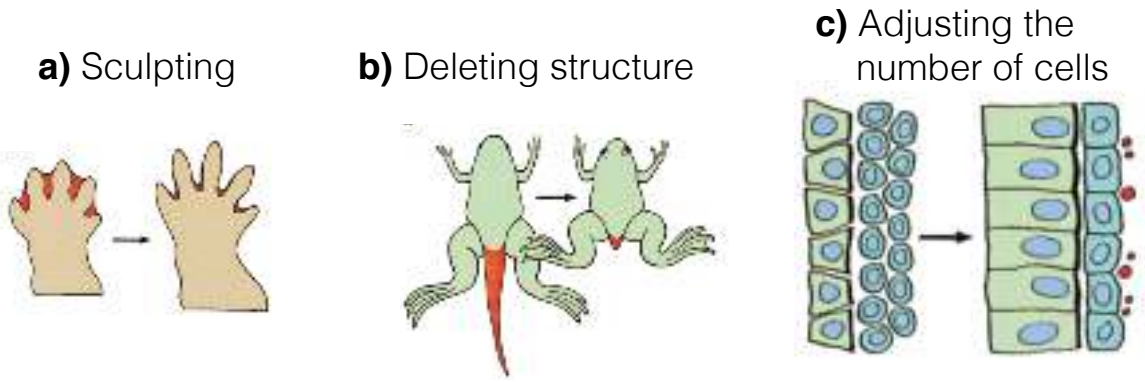


Figure 1.4 – Programed cell death. ***a)*** *Removing epidermal cells to sculpt the future fingers.* ***b)*** *Deleting structure to remove the tail.* ***c)*** *Adjusting the number of cells when they have been over produced. (adapted from Jacobson et al. [1997])*

the growth that is regulated and that allows the apical elongation of the meristem. In other plants, like *Antirrhinum*, differential growth directions may vary over the organ and drive the formation of complex petal shapes [Kennaway et al., 2011].
Cell growth rates can also be isotropic for every cell in a given organ but different between cells. This can result in complex final shapes too (see Figure 1.3) [Coen et al., 2004].

Programmed cell death. In 1977, Sulston and Horvitz [1977] showed that cell death events were reproducible in the cell lineage of *Caenorhabditis elegans*. This reproducibility suggested the idea that cell death was a "fate", hence under genetic control and therefore programmed. Even though cell death was previously known to take place during development, its active role in the morphogenetic program was unexpected. The authors of this discovery were awarded the Nobel price for medicine in 2002. There are two classes of programmed cell death (PCD): apoptosis and autophagy. PCD are often ascribed 4 functions [Jacobson et al., 1997; Conradt, 2009]. (1) Sculpting structures (Figure 1.4 a), for example to form complex structures or organs like the digits of the hand of land vertebrates. (2) Deleting unnecessary structures (Figure 1.4 b)such as during metamorphosis. (3) Controlling cell numbers (Figure 1.4 c) for example during sympathetic nervous system formation in vertebrates and (4) eliminating abnormal, misplaced, non-functional, or harmful cells as in T cell maturation in mammals. These events of PCD are often thought to be controlled transcriptionally as they are known to be affected by a family of dedicated intracellular proteins including several caspase proteases but not only [Broker et al., 2005]. These dedicated proteins have been found in all mammalian genomes that have been analysed and are thought to be present in all animal genomes. As expected for a fate decision, PCD is often triggered by the presence or lack of extracellular signals [Jacobson et al., 1997; Abud, 2004] (Figure 1.4).

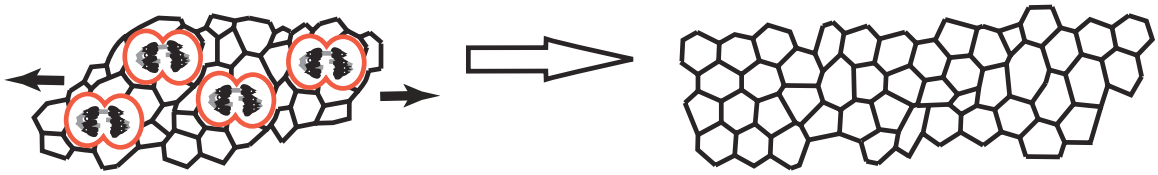


Figure 1.5 – Cell division promotes the elongation of the tissue. (adapted from Lecuit and Le Goff [2007])

Cell division. Cells can change shape during their life span and thereby contribute to the development of their organism. Then, when a cell divides, its contribution to the development stops and it is its two daughter cells that continue to shape the organism. This division not only splits the mother cell into two daughter cells, it can also favour a direction of elongation of a tissue. For example, during the process called epiboly in the *zebrafish* embryo, cells from the presumptive enveloping layer (pre-EVL), initially located at one end of the embryo, flatten and eventually cover the whole embryo. This process is regulated in part by the orientation of the division of the epidermal cells [Xiong et al., 2014]. By dividing within the plane of the epidermal layer or perpendicular to it, the cells will favour respectively more or less spreading.

Different generic rules that would explain the orientation of division planes during cell division have been proposed. In 1884, O. Hertwig, proposed a rule that is thought to represent the default behaviour of animal cells [Minc and Piel, 2012]. Hertwig's rule states that the plane that separates a mother cell in to two daughter cells is perpendicular to the longest axis of the mother and includes its center of mass. Following Hertwig's rule, a global orientation of cell divisions can result from a previous global elongation of the mother cells. In this case, the orientation of the cell divisions results in the elongation of a tissue. Therefore, elongation is a consequence of cell division orientation. Furthermore, interference with the orientation of cell division revealed that this process can have an active role in tissue elongation (see Figure 1.5) [Lecuit and Le Goff, 2007] for instance during *zebrafish* neurulation [Concha and Adams, 1998].

Cell cell rearrangements. Cell rearrangement also plays an important role in morphogenesis. Cell rearrangements can be separated into two categories, cellular intercalation and collective or individual cell migration. For example, in *Ciona intestinalis*, cell intercalation promotes the elongation of the notochord [Munro and Odell, 2002]. The cells start as a cluster of cells that intercalate to eventually create a single row of cells. This elongation was shown to need an extracellular orientation cue that is thought to be given by the presumptive neural plate cells. Because these cellular rearrangements often take place in a constrained space, they are often associated with cell shape changes. As an other example, in *Drosophila*, epithelial elongation during gastrulation is mainly driven by cell-cell reorganisation [Blankenship et al., 2006; Bertet et al., 2004] (Figure 1.6). This specific intercalation is thought to be the result of a polarization of certain molecules along the cell boundaries.

Additional examples of cell migration can be found during the development of the *Drosophila* in the midline central nervous system or in *Zebrafish* lateral line formation [Klämbt et al., 1991; Friedl and Gilmour, 2009].

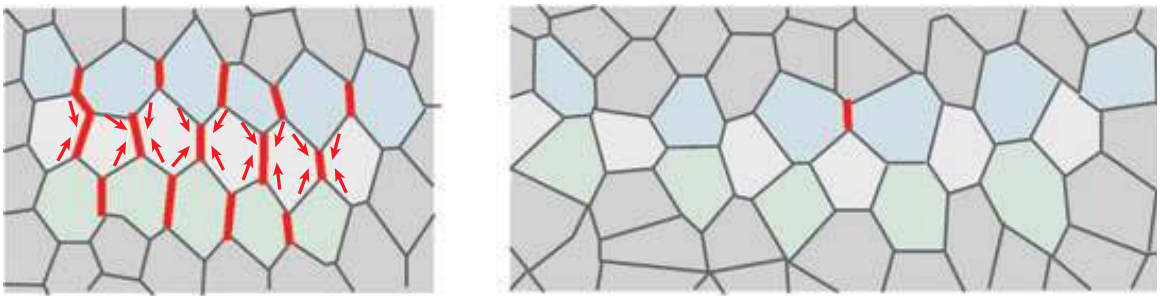


Figure 1.6 – Intercalation of the blue and green cells in between the light grey cells causes the tissue to elongate. (adapted from Lecuit et al. [2011]

In most cases, a combination of individual cell behaviours shapes the embryo. These changes need to be highly coordinated in time and space to promote a successful development. Genetic analysis of ascidian notochord formation reveals that targets of the notochord specifier gene Brachyury control distinct cell behaviours; some affect cell convergence, cell intercalation or cell shape changes [Hotta et al., 2007b]. If the signals are not properly received therefore preventing the necessary morphogenetic processes, the development can be perturbed. For example, if the orientation of the divisions of a tissue is not regulated, it cannot lead to its elongation [Lecuit and Le Goff, 2007]. Or, if differential cell growth is not finely coordinated and oriented, the organ will not elongate [Coen et al., 2004]. These changes are often entangled. For example, in *Arabidopsis thaliana* shoot apical meristem, the growth direction is promoted by the orientation of cellulose microfibrils in the cell walls. The growth then creates local mechanical stresses that then impact on the orientation of the microfibrils synthesized by the cells. This event acts as a feedback loop on the direction of the growth of the organ [Hamant et al., 2008].

From what precedes, morphogenesis fulfills two distinct aims. Firstly, it leads to the correct positioning of cells within differentiated organs, thereby ensuring their proper function. Secondly, this first consequence of morphogenesis is preceded during development by a crucial role in the relative positioning of cells emitting cell fate specification signals and cells able to respond to these signals. The challenge of the morphogenesis process is therefore to coordinate both the formation of functional organs and the temporary communications between cells.

1.1.4 Understanding morphogenesis.

Ultimately, understanding the morphogenesis of an organism and its causal forces should allow to predict the consequences of environmental or experimental perturba-

tions. By perturbing the development of an organism and looking at the resulting phenotype, it should be possible to formulate hypotheses on potential phenotypes induced by novel (but similar) perturbations. These local experimental perturbations allow to infer more global rules. This method has been extensively used in biology [Gilbert, 2006; Wolpert and Tickle, 2011]. Another way to predict the consequences of perturbations is to mathematically or computationally model the wild-type behaviour of an organism and then to perturb the model. If the model is correct, it should predict the organism's phenotype following such perturbations.

This method has been used in recent works including Besnard et al. [2013] where a model was used, together with measurements and perturbation tests, to decipher the molecules involved in the phyllotaxis of *Arabidopsis thaliana*. This modelling approach was also used in Xiong et al. [2013] to show how *zebrafish* precursor cells of the neural tube were reorganising after specification or in Sherrard, K. and Robin, F. et al. [2010] to model the gastrulation process of the ascidians.

The measurements of the physical properties of modelled organisms is a critical part of understanding development. This can be restricted to geometrical measurement of the wild-type shape of cells and of how it evolves during development as in Xiong et al. [2014]. Measurements can also report (potentially indirectly) mechanical properties of the cells such as their stress or strain as in Hamant et al. [2008] or in Boudon et al. [2015]. Measurement can also be extended to biochemical or genetic parameters such as temporal and spatial position of proteins or metabolites of gene expression patterns as in Fowlkes et al. [2008]. These measurements are usually extracted from records of the development. The geometry of an organism at the cellular scale during its development can be extracted from recorded images of the nuclei or the membranes. For the mechanical measurements, tools like atomic force microscopy or micropipette aspiration can be used [Paluch and Heisenberg, 2009; Maitre et al., 2012; Milani et al., 2013]. To record gene expression methods such as fluorescent *in-situ* hybridization can be used in order to tag molecules with fluorophores [Luengo Hendriks et al., 2006]. The ways to capture such images of membranes and/or nuclei and to extract relevant information from them will be the subject of the next two sections and the chapter 2 of this thesis.

1.2 Image acquisition

To image live organisms with a cellular level of resolution, organelles can be used as proxies for cell shapes and/or positions. The most frequently imaged organelles are the nuclei and the plasma membranes. The nuclei report cell position and allow to track cell displacements. In some organisms such as the *drosophila* embryo, cells are evenly distributed throughout the surface of the embryo, and cell shapes can be inferred from nuclei positions using computational tools such as voronoï. Depending on the precision of the cell shapes required or of their complexity, detection of nuclei might, however, be insufficient to report precisely enough actual cell shapes. In such cases, direct imaging of cell membranes is necessary to capture accurate cell shape. Imaging these organelles

is rendered difficult when the tissues are not transparent and either absorb or scatter the light, which is often the case. To overcome this issue, different methods can be used, second or third harmonics generation [Olivier et al., 2010; Witte et al., 2011] or optical projection tomography [Arranz et al., 2013, 2014] for example [Ripoll et al., 2015]. When fluorescent markers can be attached to the organelles, it is possible to image them with fluorescence microscopy. This last method is commonly used and has been the subject of recent technological developments [Huisken and Stainier, 2007; Keller et al., 2008; Krzic et al., 2012; Chen et al., 2014; Mahou et al., 2014; Vogt, 2014; Bassi et al., 2015; Jahr et al., 2015]. I will describe the principle and exhibit some of acquisition geometry for this method.

The acquisition process when imaging with fluorescence microscopy can be split into two distinct parts, (1) labelling the organelles with fluorescent markers and (2) acquiring time-series of development. Let us review these two processes.

1.2.1 Labelling organelles

Organelles can be labelled in different ways with fluorescent molecules. First, synthetic, non toxic and permeable fluorescent molecules can be able to bind directly to specific components only present in the targeted organelle. This is called vital staining method. As an alternative, genetically encoded fluorescent fusion proteins specifically addressed to the target organelle can be designed and induced by transgenesis, mRNA injection or protein injection.

Vital staining methods. These methods allow an easy labelling of the organelles since the only requirement is to immerse the organism in the dye. The dye either passes through the membranes or remains in the intra-membrane zone. It then binds on the targeted organelle. Some dyes are composed of fluorescent probes (like Hoechst 33342) that attach directly to the DNA and allow the labelling of the nuclei. Fluorescent dyes can also bind to molecules that have a high affinity with compounds that are mostly present in the structure of interest. For example, styryl dyes are composed of lipophilic molecules with fluorescent properties. They react and attach to the lipids that are present in the cell membranes thereby labelling them.

Using a dye also has limitations. Membrane staining such as FM4-64 can internalize in the cells and decrease the image quality. Nuclei staining such as Hoechst has no parasitic staining drawback but has a short wavelength which prevent deep imaging. Vital staining dyes have three main limitations, (1) the choice of the wavelength is limited, (2) their penetration across the membrane in the case of nuclei staining and (3) the internalization in the case of membrane staining. These limitations can be overcome by genetically-encoded reporters.

Genetically-encoded reporters. The principle of this class of method is to make an organism produce chimeric proteins in which a fluorescent protein is fused to a protein

domain responsible for addressing the chimeric protein to the desired organelle. There are two different methods to make an organism produce such chimeric proteins. One can first induce into the embryo modified mRNA or transgenic constructs. This exogenous mRNA will be translated in addition to the ones usually produced by the transcription of their genes. This modified mRNA is engineered such that it is translated into the a part of the usual protein fused to a fluorescent protein at one of its ends. A second method is to directly modify the genomic DNA of the organism and add at one end of a gene the coding structure of a fluorescent protein. The mRNA resulting from the transcription of the targeted gene is then translated. The translated protein eventually folds to fulfill both its normal function and indicate the position of its target organelle.

Labelled organelles can be imaged with fluorescence microscopes. They first excite fluorescent molecules with a light source. Fluorescent molecules then emit light that is recorded by one or more light sensors in the microscope. All the fluorescent molecules of the organism are excited at the same time; photons in the plane of focus are collected as well as out of focus photons which lead to a blurring of the image. Out of focus photons can then be removed using computational post-treatment (deconvolution algorithms). The blur due to out of focus photons can also be optically removed either by restricting the excitation to a single plane (light-sheet microscopy) or of a small region (multi-photons scanning microscopy) or by filtering out of focus photons using pinholes (confocal and spinning disk microscopy). These strategies allow to capture single 2D images across the sample. To image the whole sample, the focal plane is displaced from one end of the sample to the other. The organism is therefore scanned in depth, section by section, producing a set of 2D slices, which together form a "3D image" of the organism. Note that it must be assumed that the time necessary to capture a 3D image should be sufficiently short to ensure that the geometry of the imaged object does not change significantly between the beginning and the end of acquisition. These 3D acquisitions are then repeated through time to generate a 4D image of the development of the organism.

The next two sections will successively describe first the main determinants of image quality, then the technical solutions at hand to produce suitable images.

1.2.2 Image quality assessment

To extract morphogenesis relevant informations, high quality images are required. When illuminated by a light source, the electrons of a fluorescent molecule go from a low (ground) energy state to a higher energy state. When the illumination is stopped, the electrons go back to the lower energy state, releasing energy as photons, that are captured by the observation device. For a given observation device, the larger the number of emitted photons, the higher and the better the signal. This number of emitted photons is directly linked to the number of fluorophores (density), the size of the sample (object size), the desired spatial resolution (resolution), the excitation intensity (signal), the exposure time of the sample (time period) and the number of repetitions (temporal

resolution times the observation length). This can be captured by the following formula [Stelzer, 2013]:

$$\begin{aligned} \#\,\text{photons} \sim \text{density} \cdot \text{object size} \cdot \text{resolution} \cdot \\ \text{signal} \cdot \text{time period} \cdot \text{repetitions} \end{aligned} \tag{1.1}$$

Fluorescence microscope devices and their imaging protocol are designed to maximize this number of captured photons. Note that these parameters are not independent of each other. Maximizing the number of photons thus requires to understand these trade-offs in order to optimize the acquisition. Last but not least, this equation does not show a major limitation of this optimization, the number of possible photons emitted can be degraded and this degradation can have a drastic photoxic impact on the sample development.

Photobleaching and phototoxicity. When getting back from a higher level of energy, the electrons have a probability to remain stuck in an intermediary third state, instead of going back to the ground state. This probability, usually low, is related to the intensity of the light source. The higher the intensity, the higher the probability. When an electron in a molecule reaches this third state, two major defects can be induced due to photochemical reaction: photobleaching and phototoxicity [Lichtman and Conchello, 2005]. Photobleaching incapacitates further electron emission and phototoxicity harms the development of the organism. It is necessary then to optimize the illumination, enough to get a good signal but not too much to avoid that the electrons fall in this third energy state.

Optical resolution. The shortest distance between two points on an imaged object that is necessary to distinguish them defines the optical resolution. For a 3D stack, X, Y and Z resolutions are usually defined as the distance between two consecutive scanned points in the direction of their respective axis. The resolution is directly linked with the magnification. The resolution decrease is proportional to the increase of magnification. In contrast with the usual screen resolutions for which the higher is the better, the optical resolution is better when its value is low. The lower bound of the resolution is determined by the wavelength (λ) of the laser used to illuminate and by the numerical aperture (**NA**) of the objective lenses. This lower bound is $\frac{\lambda}{2\mathbf{NA}}$ [Hell, 2009]. As **NA** and the size of the field of view are usually inversely proportional, the gain in resolution is often coupled with a diminution of the size of the field of view.

Temporal resolution. During organism development, local changes in cell shapes define the developmental rate. To acquire every change that occurs, it is necessary to acquire faster than the developmental rate. The time separating two successive acquisitions of a 3D snapshot of the imaged organism defines the temporal resolution. Temporal resolution is limited by two factors. First, an increase in the number of imaged time points increases the duration of exposure of the sample to the light thereby increasing the risk of phototoxicity or photobleaching. Second, the time elapsed between two successive acquisitions cannot be inferior to the time needed for one acquisition.

Acquisition time varies with the type of microscope used, with the intensity of the exciting light and with the desired spatial resolution. It is possible to decrease the laser power intensity to decrease the risks of phototoxicity or photobleaching but this will trade-off with the image quality. In externally developing embryos, it is also possible to lower the temperature to decrease the developmental rate [Rombough, 2003; Jarośík et al., 2004]. This latter method implies longer imaging times.

Signal to noise ratio. When imaging membranes or nuclei of an organism, it is possible that unwanted signal is captured as well. This unwanted signal usually comes from fluorescence that is present in parts of the organism that are not of interest. For example, when using a styryl dye to mark the membranes, internalisation of the fluorescent molecules inside the cytoplasm can occur. This interferes with the detection of the actual membranes. This signal is considered as noise and the quality of the acquisition can be quantified by the signal to noise ratio (SNR). It is defined as the mean of the the signal intensity given by the structures over the mean of the background signal intensity. Low exposure time and low laser intensity power are often a reason of poor SNR. Though, increasing laser power or exposure time of the fluorophores might induce photobleaching/toxicity. An other reason of poor SNR is a low specificity of the marker used, it can be present in unwanted organelles creating noise inside them. Poor SNR can also come from crosstalk between the imaged pixels. A signal that is received by the objective can actually come from photons that were not meant to be excited or from diffraction through the tissues the photons cross.

These parameters help us to quantify the quality of time-series. Low temporal and spatial resolution and high SNR are preferred. But both spatial resolution and SNR are antagonistic with keeping a high temporal resolution. All are antagonist to low photobleaching and phototoxicity. Therefore, when imaging a sample, it is necessary to make compromises to select the features that are least crucial for the specific purpose of the imaging experiment.

1.2.3 Different methods to capture a single plane across the sample.

Fluorescence microscopes have been designed to specifically adapt optimally to the three features mentioned previously. The microscopes are required to be fast, with low SNR, good spatial resolution and low photobleaching and toxicity. The difference between the various microscopes is the scanning method and the excitation source.

Point scanning microscopy, confocal microscopy. In confocal microscopy the sample is scanned point by point [Davidovits, 1969; Pawley, 2006] (Figure 1.7 **a**). The laser beam is focused on one point of the sample. To illuminate this focused point, however, the laser beam traverses the whole sample illuminating all the molecules in its path. To avoid capturing the photons that do not come from the focused point, a pinhole is added to the set-up between the sample and the camera. The pinhole

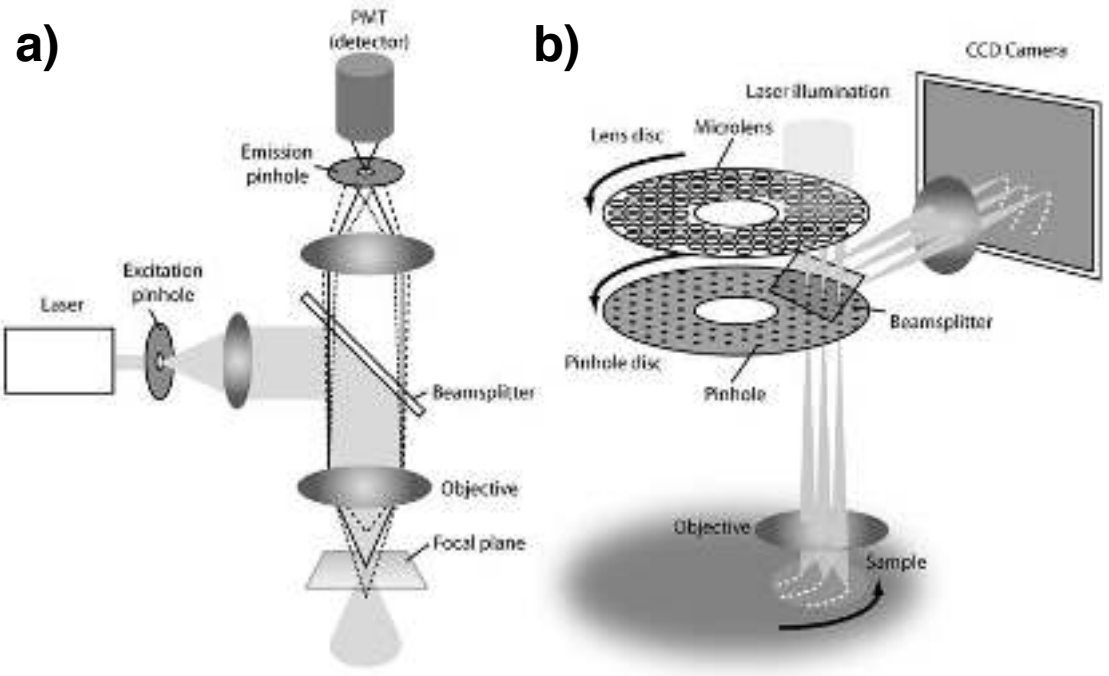


Figure 1.7 – ***a)*** *Principle of confocal microscopy.* ***b)*** *Principle of spinning disk microscopy (adapted from [Graf et al., 2005])*

blocks the out of focus photons emitted by the molecules that have been illuminated by the laser beam. The pinhole is designed to block all the photons that are not emitted from the focal point of the lenses. This method allows precise scan of the sample. However, the necessary illumination of the whole thickness of the sample for each point may lead to an increased photobleaching/toxicity. Also, in point scanning methods, the acquisition time is linearly dependent on the number of points to be scanned. This makes it relatively slow and limits temporal resolution.

Improving confocal microscopy. Confocal microscopy has several limitations: it has a low temporal resolution due to the single point scanning procedure, it causes photobleaching and its tissue penetration is usually limited to $\approx 100\mu m$. To overcome the first issue, spinning disc microscopes have been developed. In this set-up, the laser beam is split to enable multiple point scanning at the same time. This increases the speed of acquisition but do not the other two issues [Graf et al., 2005] (Figure 1.7 **b**). To bypass the two other issues, bi-photon microscopes illuminate the fluorescent molecules with two photons of two wavelengths that equals to two times the excitation wavelength of the fluorophore [Denk et al., 1990; Konig, 2000]. This 2 or multi-photon excitation takes advantage of the fact that the fluorescent molecules are almost not excited where the beam is not focused. Therefore the out-of-focus molecules are usually not excited. This allows to get rid of the pin-hole for these configurations. An other advantage of

multi-photon microscopes is that they allow a better penetration thanks to the higher wavelength used to excite the fluorescent molecules [Ripoll et al., 2015]. However, these microscopes remain relatively slow and phototoxic.

Figure 1.8 – Principle of light-sheet microscopy. The light-sheet goes through the illumination objective. The detection objective is perpendicular to the illumination objective and allows to image the whole field of view created by the light-sheet. (adapted from [Weber et al., 2014])

Plane scanning microscopy, light-sheet microscope. To increase scanning speed, like the spinning disk methods, while keeping a low phototoxicity/bleaching, light-sheet fluorescence microscopes (LSFM) selectively illuminates a whole plan across the sample [Huisken et al., 2004] (Figure 1.8). Conversely to the confocal microscopes that use spherical lenses to focus the laser beam in a single point, LSFM use cylindrical lenses to focus the beam in only one direction which creates a thin sheet of light. Then, by positioning the detection axis perpendicularly to this light sheet, it is possible to simultaneously acquire the illuminated plane.
This method was initially imagined in 1903 [Ripoll et al., 2015]. However, it is only since Huisken et al. [2004] showed its application to the *in-vivo* imaging of the *zebrafish* that it really took off. This was mainly possible thanks to the recent improvements of the captor sensitivity (CCD and sCMOS cameras).
This technique allows high speed acquisition since all points of a 2D slice are imaged simultaneously. All the illuminated fluorophores are recorded. This reduces photobleaching rsuch as with confocal microscopes where, to acquire a sample through n planes, each molecule is excited at least n times. With light sheet microscopy, each molecule is excited only once, when it is imaged. The acquisition velocity is more than $x \times y$ time faster than point scanning microscopy (where x and y are the number of

pixels and n the number of planes).
However, the lower photobleaching and acquisition velocity gains trade-off with sample penetration. The laser beam is less focused (a plane instead of a point) and therefore, if the sample is not transparent enough (like *drosophila*) more laser power is needed to go through the sample. That can counterbalance the gain of phototoxicity mentioned before. Another issue is that the light sheet is created from one source on one side of the sample. If the light sheet encounters less transparent structures, they can block the light and the excitation of fluorophores behind it leaves shadow stripes. Finally, there is a lot more crosstalk between the imaged points since they are acquired at the same time which might increase the SNR and make thin structures, such as membranes, harder to detect.

Line scanning microscopy, scanned light-sheet microscopy. A hybrid method that combines the advantages of both confocal and light-sheet microscopy was subsequently introduced. Digital scanned light-sheet microscopy (DSLM) [Keller et al., 2008] uses the laser beam focused on a line instead of a plane to generate the light-sheet by rapid scanning.

Emitted photons from this line are then recorded by the camera, the detection axes being perpendicular to the laser line. This method allows to record all the molecules that are illuminated while maintaining a low crosstalk between the pixels and a high rate of acquisition. The crosstalk between pixels due to the light diffraction is also better than the LSFM, but not as good as in the confocal microscopy. To improve this latest drawback, the pinhole behaviour of the confocal can be emulated thanks to the latest camera technologies [Yang et al., 2015]. These newest cameras allow to individually select the sensor lines that are turned on or off inside the camera. By synchronising the camera and the scanning laser line, only the sensors that should receive photons for a given laser beam position are turned on, limiting the effects of the light scattering (Figure 1.10). This set-up can still struggle to record thick samples and is still subject to the potential shadow stripes.

Figure 1.9 – DSLM [Weber et al., 2014]

Improving image acquisition. Finally, all of these microscopes share two disadvantages. Firstly, they all have an anisotropic resolution which means that the lateral resolutions (in the scanned planes) are usually a lot better than the depth resolution. Secondly, they struggle to image thick or deep organisms. To overcome these issues, the sample can be rotated to allow the light to penetrate from different angles. This allows more homogeneous illumination, compensates for the shadow stripes in the cases of LSFM and DLSM systems. A post-process is then needed to fuse the different im-

ages obtained from acquisition from the different angles. The composite resulting fused image usually has isotropic resolution. Usually, for light sheet based microscopes, 4 to 8 angles are required [Krzic et al., 2012] and for confocal microscopes 3 or 4 angles can be used in the case of slowly developing organisms such as plants [Fernandez et al., 2010]. Multiplying the angles of acquisitions reduces the time necessary to acquire one time-point. This increases the advantage of the parallel scanning microscopes with respect to the confocal ones. This advantage is emphasized by the possibility to parallelize this rotation process. LSFM and DLSM microscopes can be built in order to illuminate and detect from two opposite sides [Huisken and Stainier, 2007; Krzic et al., 2012]. This allows to acquire the equivalent of 2 angles (0° and 180°) illuminated from two opposite sides in one single acquisition. It therefore divides the number of required rotations and the number of acquisitions by two.

These microscopes produce 4D images (3D+time) stacks that have a specific temporal and spatial resolution. These data-sets usually contain nuclei or membrane information. Due to the typical observation time lapse and high temporal resolutions needed to report the development, the data-set often contains from tens to hundreds of 3D stacks each ranging from 500Mb to 10Gb in size after fusion [Amat et al., 2015] creating 4D images that can have sizes up to dozens of terabytes. From these stacks, shape and position information of the cells can be extracted. These extractions are usually done using fully/semi-automatic image processing algorithms.

1.3 Extracting information from images

4D image datasets contain a very rich description of the position of all labelled structures. Their biological analysis requires the identification of individual structures in

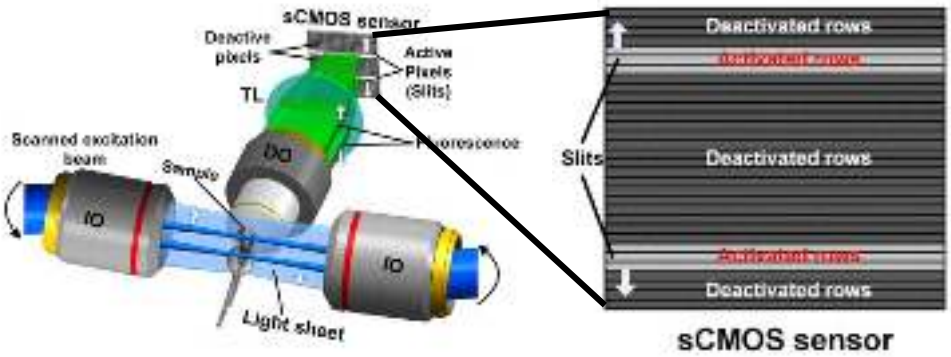


Figure 1.10 – Principle of slit mode in scanned light-sheet microscopy. The lasers, seen here coming from both sides of the embryo, are focused by corresponding illumination objectives (IO). The emitted photons then go through the detection objective (DO) to be captured by the sCMOS camera where only the rows that are meant to recieve light are activated. Adapted from [Yang et al., 2015]

order to characterize their individual behaviours. This process is called segmentation. It can be performed manually on small datasets [Tassy et al., 2006; Sherrard, K. and Robin, F. et al., 2010; Nakamura et al., 2012]. These manual segmentations are probably the most trustworthy, however, light-sheet datasets are generally of very large size, often larger than a terabyte. In such cases, the amount of time necessary to manually segment the dataset is not affordable. For example, following a *Ciona intestinalis* embryo during 6 hours of development from the 64 cell stage would yield around 60000 cell snapshots if an image was taken every two minutes (see Chapter 2). To manually segment such a data-set would take 3 years and a half of work 24/7. It is therefore not feasible to manually segment such dense time-series. This information has to be then extracted automatically or at least semi-automatically.

Computational extraction usually involves four major consecutive steps. (1) A segmentation algorithm identifies and finds the boundaries of the labelled cellular organelles in all cells of each 3D stack. (2) A tracking algorithm tracks all segmented objects from one time point to the next in the time-series. (3) The data are organised in a dynamic graph. And (4) metrics are built to measure positions and shape evolutions of the different cell organelles, tissues and possibly the whole organism. Each of these steps can be performed with different classes of algorithms. Together, they produce a digitalized and formalized representation of the embryo.

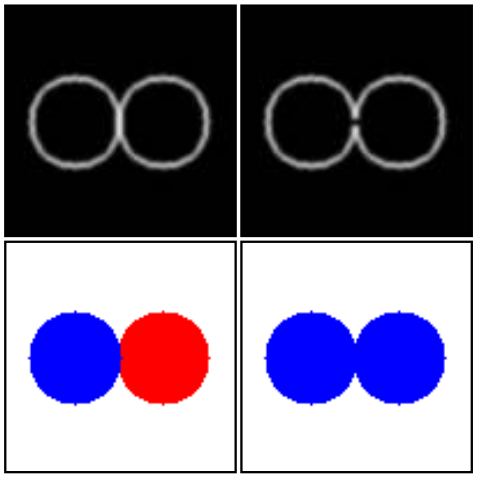


Figure 1.11 – How a small modification of a given image can change the resulting segmentation. Top row, two synthetic images of membrane marked images; to the right, a small part of the intensity have been removed. Bottom row, the results of the segmentations of the two upper intensity image.

1.3.1 Segmentation of an image into regions of biological object of interest.

The segmentation problem consists in deciding, for each coordinate of the space (or voxel) of a given image $\mathcal{I}$, which object (nuclei, cell, tissue) it belongs to. A segmentation partitions the image $\mathcal{I}$ into either regions of a biological object of interest or exterior. The result of these kind of algorithms can have different forms. It can be an image $\mathcal{S}_\mathcal{I}$ that has the same dimensions of $\mathcal{I}$. In $\mathcal{S}_\mathcal{I}$, to every coordinate $(x, y, z) = \mathrm{v}$ is attached the label of the object it belongs to.

Why is the segmentation problem hard? The partitioning of an image into discrete objects can be done easily and quickly by the human visual system. Computation-

ally, it is, however, a complex problem, as it is ill-posed [Khairy and Keller, 2011; Vu, 2008; Micusik and Hanbury, 2005]. A well posed problem satisfies three conditions, (1) it has a solution, (2) this solution is unique and (3) the result is robust to small variations in the initial conditions (in the case of the segmentation, the image). The condition (1) is assumed to be true. But the conditions (2) and (3) are often not respected. In the case of image segmentation, two experts can give significantly different segmentations of a given image [Amat et al., 2014], which invalidate condition (2). Figure 1.11 gives two example of how the condition (3) can be broken.

To be respected, it is possible to include prior knowledge to the process. This prior knowledge, often depends on the organism that is to be segmented. It can also vary depending on the period of development for a given organism. That is, in part, what makes the segmentation problem hard. More specifically, the signal or the noise in a time-series can be different across time or space. For example, the signal diffraction can increase if the sample becomes thicker because of internal cell divisions changing the SNR. The differences can also be biological. For example, the shape or size of the cells or nuclei differ depending on the developmental stage and within a stage, it can be affected by the cell cycle. The signal intensity can even vary according to this biological diversity, for example in animal embryos, the signal in labelled membranes is less intense in apical membranes facing the exterior than at the interface between two cells. This comes from the fact that the number of fluorescent molecules is higher for the internal membranes due to the double contribution of the two cells in contact (see Figure 1.12). The differences in results in temporal and spatial inhomogeneities across the time-series.

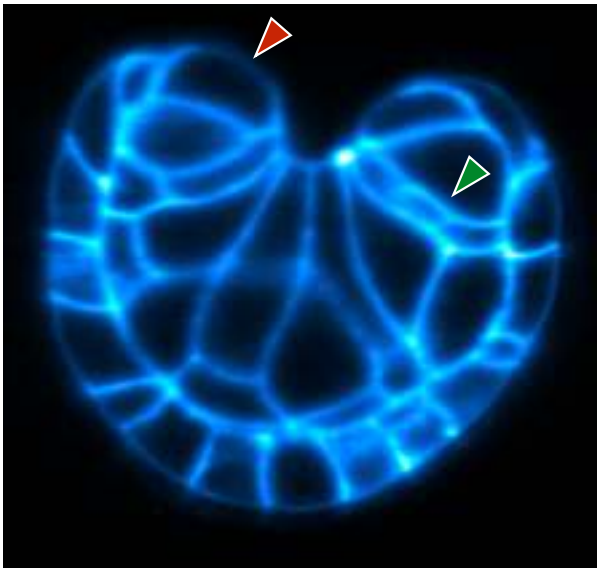


Figure 1.12 – Example of image inhomogeneity. 2D optical section through a Phallusia mammillata embryo during gastrulation. The red arrow points to an outer membrane with fainter signal than the inner membrane indicated by the green arrow.

These fluctuations coupled to the noise induced by the imaging procedure generally make

a generic segmentation algorithm impossible. Moreover, it induces the algorithms to fail to detect some real cell structures, (problems of under-segmentation, US, Figure 1.13 c) or to miss cells (Figure 1.13 d). They can also induce wrong detection of noise as actual structures (problems of over-segmentation, OS, Figure 1.13 b). Additionally, even when the cells are correctly detected, it can happen that their shape is not accurately segmented, again because of these noises (problems of shape detection, SD Figure 1.13 e).

To deal with these issues, different algorithms have been developed which can be classified in to 3 classes, (i) pre-treatment to partially remove the noise of the acquisition, (ii) identification and shape reconstruction and (iii) post-treatment. These methods are often adapted to a specific type of acquisition (membrane or nuclei), microscope and organism. It is then a combination of 2 or more of these algorithms that makes up a segmentation pipeline.

1.3.1.1 Pre-treatment algorithms.

An image can be subjected to random noise (see previous paragraph). Pre-treatment algorithms can be applied beforehand to $\mathcal{I}$ in order to partially remove it. Usually, this noise is supposed to be linear and additive [Bloch et al., 2005], meaning that, the image $\hat{\mathcal{I}}(x, y, z)$ with no noise, is affected by the additive noise $n(x, y, z)$ to give the noised image $\mathcal{I}$:

$$\mathcal{I}(x, y, z) = \hat{\mathcal{I}}(x, y, z) + n(x, y, z) \tag{1.2}$$

These issues can be resolve by using the continuous property of the imaged sample and by using filters. The gaussian, mean and median filters are the most commonly used [Bloch et al., 2005; Al-Kofahi et al., 2006; Bao et al., 2006; Long et al., 2009; Keller et al., 2008]. They all consist of a similar treatment whereby a voxel v is modified according to its

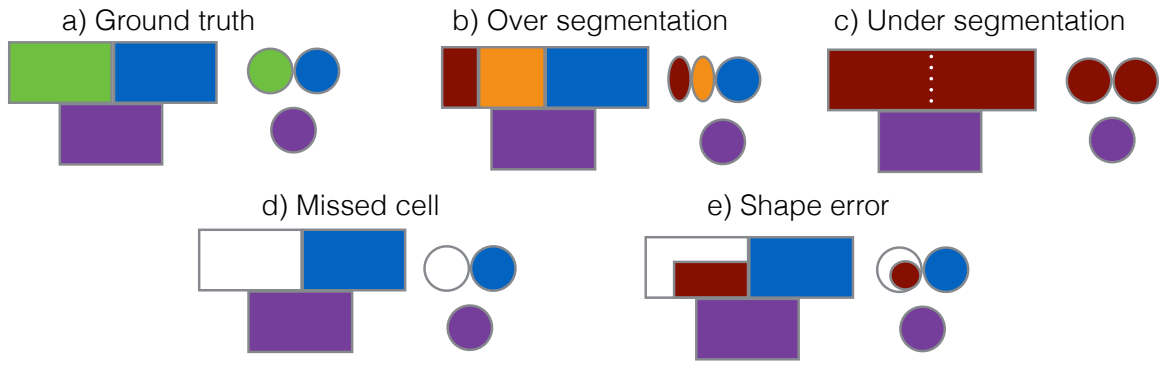


Figure 1.13 – Illustration of the possible mistakes in a segmentation. The green cell in the ground truth a) can be split into two or more cells (over segmententation) b), fused with an other cell (under segmentation) c), not segmented at all (under segmentation) d) or its shape can be wrongly segmented (shape detection) e).

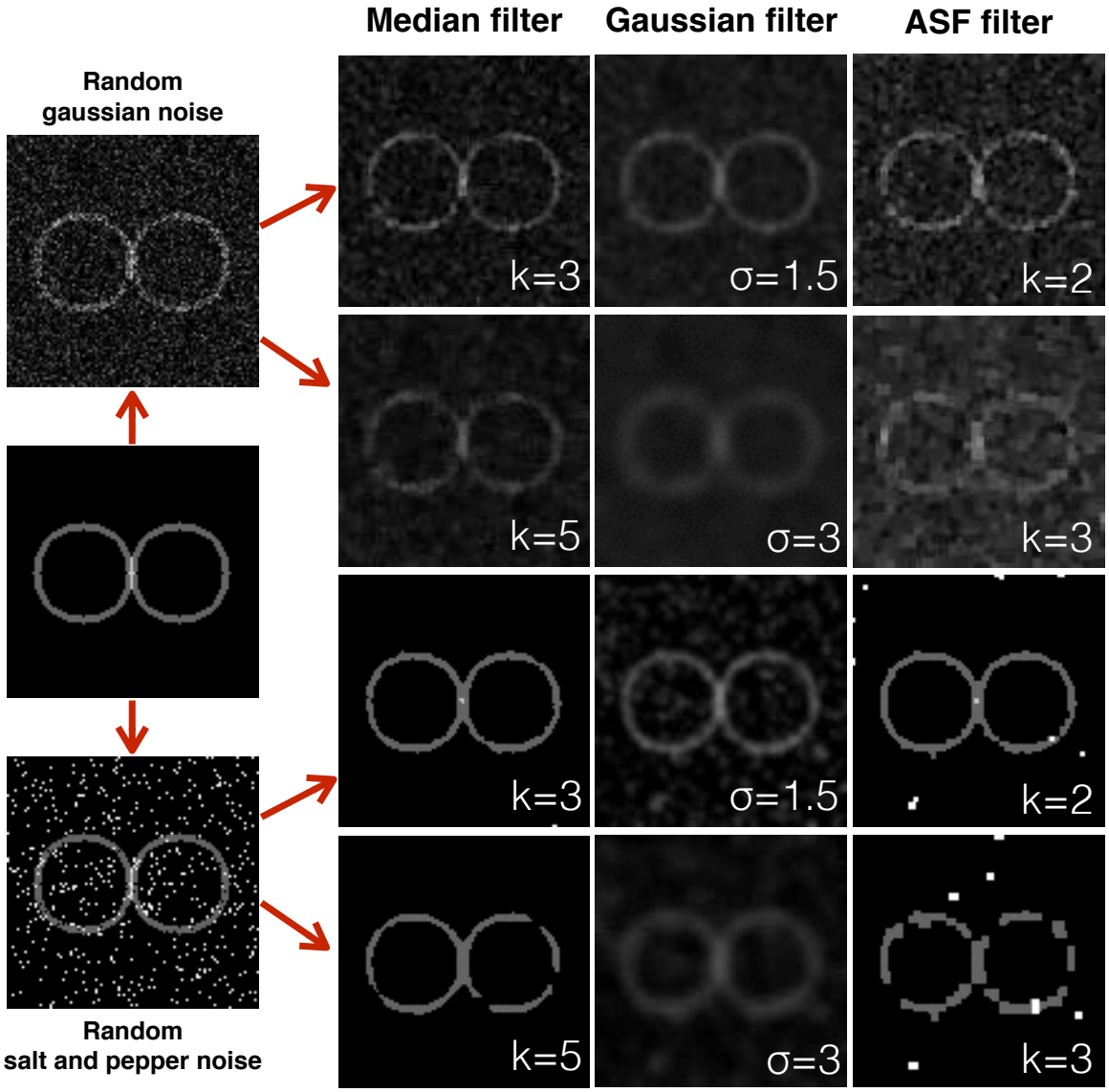


Figure 1.14 – Example of intensity image pre-treatments. Random gaussian and salt and pepper noise have been added to a synthetic toy image. Median, gaussian and alternative-sequential filters have been applied with different parameter values.

neighbourhood. The median (resp. mean, gaussian) filter with a neighbourhood defined by the shape of $\mathcal{K}$ (the kernel) transforms a voxel $\mathrm{v} = (x, y, z)$ of an image $\mathcal{I}$ to the values of the median (resp. mean, mean weighted by a gaussian distribution) of the set of values that are in $\mathcal{K}$ centered on v. These three filters have at least one parameter that is the size of $\mathcal{K}$ to use or the value of the σ of the gaussian distribution. When the structures to be imaged are too close to each other (nuclei packed) or discontinued by imaging defects (holes in membranes), morphological filters can be used. The morphological filters allow to close or create gaps by performing a series of opening and closing on the image (e.g. in Fernandez et al. [2010]). The opening procedure separate the structures

that are linked by thin structures and the closing procedure fill the small holes and link close structures.

In addition to these standard filters, more sophisticated pre-treatment algorithms have been designed to answer specific issues. For example, pre-treatments have been proposed in Mosaliganti et al. [2012] and Michelin et al. [2014] (see annex A for detail on the method). These methods have been designed to detect and enhance membrane-like structures in intensity images based on the study of second derivatives of the image.

1.3.1.2 Nuclei and cell segmentation.

When the nuclei are imaged, the objects to segment are thus local peaks of intensity. If membranes are imaged, the cells are delimited by local ridges of intensity. These two types of labels induce major differences in the principle of the segmentation algorithms.

Nuclei segmentation. For nuclei images, the segmentation problem can be reduced to the detection of local maxima of intensity in the image. Several methods have been designed to do so. Methods such as in Al-Kofahi et al. [2006]; Bao et al. [2006]; Long et al. [2009]; Keller et al. [2008] are based on a thresholding of the nuclei image. These methods binarize the intensity image $\mathcal{I}$, creating an image $\mathcal{B}$ where the 1s are putative nuclei and 0s are the exterior of the nuclei. A threshold value th is either given by the user or computed using global approaches such as Otsu's method [Otsu, 1975] for example. This thresholded image $\mathcal{B}$ is defined by $\mathcal{B}(\mathrm{v}) = 0$ if $\mathcal{I}(\mathrm{v}) < th(\mathrm{v})$ and 1 otherwise. The threshold $th(\mathrm{v})$ can be global to the image (i.e. $th(\mathrm{v}) = th_0 = \mathbf{cste}$)or a function of v. Then, the connected components of $\mathcal{B}$ are extracted and seen as candidate nuclei.

A-priori knowledge on the objects to segment is also often used to help with nuclei detection. For example in Bao et al. [2006], the nuclei are known to be spherical and the threshold algorithm is implemented to give spherical segmented nuclei. In Al-Kofahi et al. [2006], the size of the nuclei are known to be in a given range. This information is used to discard over-segmented nuclei that are too small and to split big nuclei into two nuclei. Keller et al. [2008] uses morphological knowledge of the nuclei to discard the wrongly segmented ones.

Other methods than thresholding methods can also be used as in Keller et al. [2010] and Sommer et al. [2011] or in Santella et al. [2010] for example. In Keller et al. [2010] and Santella et al. [2010] a laplacian-of-gaussian blob detection [Marr and Hildreth, 1980] was used. In Sommer et al. [2011]; Kausler et al. [2012]; Schiegg et al. [2015] a more statistical approach was used. In these studies, each voxel is classified as belonging to a nucleus or not using random forest classifier [Breiman, 2001], markov chain models or graphical models. These latter methods result in a binary image, as in the classic threshold methods. The connected components are then identified as unique cells.

Another way to constrain the shape of the segmented nuclei is to use deformable models. These models start from a predetermined rough contour of a given object. A tension and curvature of this contour can be processed given a mechanical model. By minimizing the energy function of the contour, given by the tension and the curvature, with respect to the intensity profile of a given image, the contour tends to match the shape of the targeted nuclei. This method has been successfully used when the nuclei are well separated from each others such as in Pecreaux et al. [2006]. One of the strengths of these active contour methods is that the physical tensions and bending forces that are necessary to parametrize the contours can sometimes actually be experimentally measured [Pecreaux et al., 2006; Paluch and Heisenberg, 2009; Maitre et al., 2012]. Deformable models, however, tend to fuse nuclei that are too close to one another (under-segmentations) [Khairy and Keller, 2011]. To avoid this, fine pre-treatment of the intensity image is often required. In addition, these methods are often slower to execute than thresholding based ones.

	[Schiegg et al., 2013]	Statistical [Schiegg et al., 2015]	Propagation [Amat et al., 2014]
Precision	0.82	0.97	0.99
Recall	0.93	0.93	0.88
F-measure	0.87	0.95	0.93

Table 1.1 – Results of different algorithms performed on a similar data set of nuclei stained recordings of Drosophila embryos consisting in 65821 cell snapshots. The precision is the number of correctly found nuclei over the number of correctly found + the number of artefactual nuclei found. The recall is the number of correctly found nuclei over the number of correctly found + the number of nuclei not found. The f-measure is twice the precision times the recall over the sum of the precision and the recall. [Amat et al., 2014] is descirbed in section 1.3.3.2. These results are extracted from [Schiegg et al., 2015]

Membrane segmentation. Different methods have been developed to segment cells on the basis of the labelling of enveloping cell membrane. The most common class of algorithm used to segment such images in developmental biology are watershed-based algorithms. This method considers the image as a topological landscape (Figure 1.15). High staining intensities form ridges while the low intensities are found in valleys or basins. The watershed method identifies cells by partitioning the image into groups of voxels on the basis of their belonging to the same basin. Two voxels belong to the same basin if and only if their path of steepest descent ends in the same point. This definition can be implemented by simultaneously filling all the basins; when done, the basins meet at the ridges, the membranes. This basic implementation of the watershed algorithm is very little used in biology. The major flaw of this implementation is its high sensitivity to noise, which often leads to over-segmentations. Many small and shallow basins can

arise from noise in the cytoplasm for example.

To decrease this sensitivity to noise, Adams and Bischof [1994] proposed that seeds could be put in the image in order to initiate and drive the watershed. These seeds initiate each basin and define the number of cells that will be segmented. Accurate seed detection is therefore crucial. A similar automatic method to detect seeds has been implemented in Fernandez et al. [2010]; Mosaliganti et al. [2012]; Michelin et al. [2014]. The basins are pre-filled and if two or more basins end up, after the pre-fill, touching each other, they are then merged. In other terms, if the ridges that separate two basins are too low, this is considered noise and the basins are fused. The seeds found are thus crucial for the segmentation. Too many seeds will induce OS and too few US.

Another issue with a watershed algorithm is that if a membrane is not well defined, segmented cells can leak into the outside or into other cells. This can be overcome by adding physical constraints on the possible shape of the cells. As for the nuclei, deformable models can be used to render the best possible shapes [Cilla et al., 2015]. The restriction of the application of this approach to membranes segmentation is that the surface tension of the cell membranes can sometimes not be uniform. This limitation restricts this kind of methods to more rounded shaped cells such as epithelial cells or plant cells.

Finally, as for nuclei, statistical classification methods can be used to first separate the membranes from the cytoplasm. The connected parts of cytoplasm are then attributed to a specific cell [Lièvre, 2014]. In Delibaltov et al. [2013], a set of segmented images of a given image $\mathcal{I}$, where the seeds have been randomly positioned, is used. Then, by looking at the stability of the shapes among all the segmented, it is possible to correctly delimit the cells and to fuse the over segmented cells (see Figure 1.16).

Due to the large diversity of segmentation methods available, one needs to carefully choose the method to use for a specific biological and imaging situation. Even when a segmentation method is optimized for a given situation, errors will remain in the

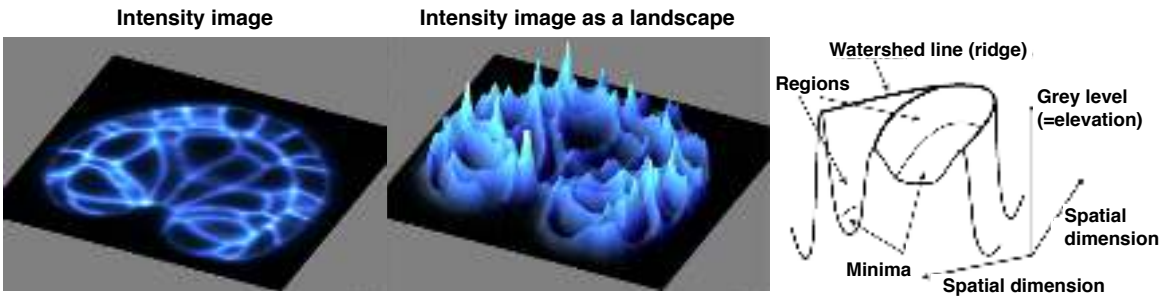


Figure 1.15 – Example of landscape representation of an intensity image.

resulting segmentation, (see Tables 1.1 and 1.2), which can be corrected using post-treatment algorithms.

	ACME [Mosaliganti et al., 2012]	EDGE4D [Khan et al., 2014]	MARS-ALT [Fernandez et al., 2010]
False discovery rate	0.26	0.14	0.30
False negative rate	0.16	0.16	0.46
Estimated Processing time	245 days	710 days	119 days

Table 1.2 – Results of different algorithms performed on a similar data set of membrane-stained light-sheet microscopy recordings of Drosophila, mouse embryo and confocal fluorescence microscopy image data of a Drosophila embryo [Stegmaier et al., 2016]

1.3.1.3 Segmentation post-treatment.

Post-treatment methods can be manual semi-automated and automated treatments. Manual correction consist in spotting errors and fusing the over-segmented cells and splitting the under-segmented cells, identifying missed cells and correcting the shape of the cells. Manual curration can be aided by computational metrics that report cells that deviate from the expected properties. These metrics are usually based on biological knowledge as in Amat et al. [2014] or in Fernandez et al. [2010]. The consistency of cell volumes across time or the minimum time that is expected between two division events can be computed and then used to tag cells that are potentially poorly segmented.

These metrics can also be used to automatically correct the segmentation. For example, if two neighbouring cells are much smaller than expected, it is likely that they

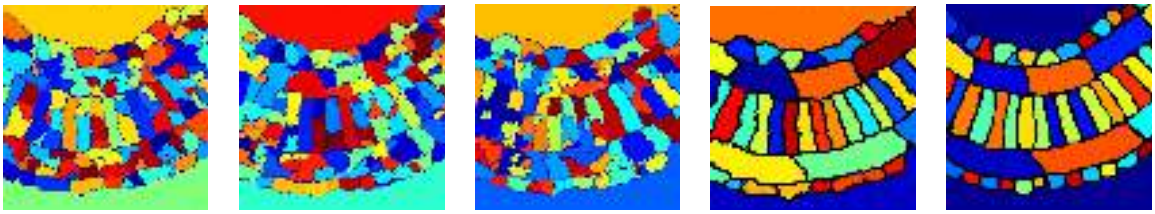


Figure 1.16 – From left to right: Three randomly initialized watershed over-segmentations. Joint correction using method in Delibaltov et al. [2013]. Ground truth segmentation. (adapted from Delibaltov et al. [2013])

should be fused together. In Fernandez et al. [2010], when a cell is too small, the seed that leads to it is removed and a watershed is performed *de novo*. In specific organisms, specific cell shapes are expected, for example in *A. thaliana*, cells usually do not have protrusions. From such specific knowledge, specific filters can be designed. In the previous case, specific filters can locally improve the shape of the cells by removing the protrusions or by smoothing the surface of a cell. These algorithms can be morphological operators on labelled images. These types of automatic corrections are possible if prior knowledge on the studied system is accessible and if it is homogeneous across the organism.

1.3.2 Cell tracking

Each segmented image thus provides snapshots of the cells or nuclei present at this time point. To track the development in time of a cell and its progeny, it is necessary to follow the segmented objects from one time-point to the next. Since the segmentations are usually done independently, it is necessary to associate each segmented cell/nuclei snapshot from one image at a given time t to its counterpart in the next segmented image in the time-series. This problem can be mathematically formulated. The track to find is a function $\mathcal{T}$ that best maps cells snapshot from $\mathcal{S}_t$ to $\mathcal{S}_{t+1}$. Rules can be added to this mapping for particular types of acquisitions or embryos. If there is no cell fusion during the time-series studied, a cell in $\mathcal{S}_{t+1}$ can only have at most one image by $\mathcal{T}^{-1}$. Biologically, this means that a cell can only arise from a unique cell but that a cell can appear in the field of view if only part the organism is imaged. If the whole organism is imaged, then every cell in $\mathcal{S}_{t+1}$ has exactly one antecedent. If the organism is known to have no cell death during its development, then $\mathcal{T}^{-1}$ is surjective. Which biologically means that every cell from $\mathcal{S}_t$ is mapped to at least one cell from $\mathcal{S}_{t+1}$.

Nearest antecedent. A first approach to building $\mathcal{T}$ is to register $\mathcal{S}_t$ onto $\mathcal{S}_{t+1}$ (see chapter 3 for registration algorithm examples), to compare the cells of $\mathcal{S}_t$ to the cells of $\mathcal{S}_{t+1}$ and to match the ones that correspond best. To compare two cells in $\mathcal{S}_t$ and in $\mathcal{S}_{t+1}$ a distance between them can be computed. This distance can be the physical euclidean distance between the two cells. Keller et al. [2008] or Brown et al. [2010] added a correlation analysis of the shape of the nuclei to refine the distance. Cell motions can also be exploited in order to refine this distance metric by giving a probability of the position of a cell from $\mathcal{S}_t$ in $\mathcal{S}_{t+1}$. From these metrics, a distance d can be built, which associates a numeric value to a pair of cells from two consecutive time-points. Then an inverse track corresponding to $\mathcal{T}^{-1}$ can be built by mapping every cell in $\mathcal{S}_{t+1}$ to its nearest one in $\mathcal{S}_t$, $\mathcal{T}(i) = \text{argmin}_{j \in \mathcal{S}_{t+1}}\{d(i,j)\}$. This method was used in Bao et al. [2006] coupled to heuristics based on biological knowledge to correct potential mistakes that were made. The heuristics that helped the tracking process were the minimal time between two consecutive divisions, the shape of the cell before the division and the size of the potential sister nuclei.

Nearest antecedent methods usually map every cell in $\mathcal{S}_{t+1}$ to a cell in $\mathcal{S}_t$; therefore, it works best when there are no cells coming in or out of the field of view between two

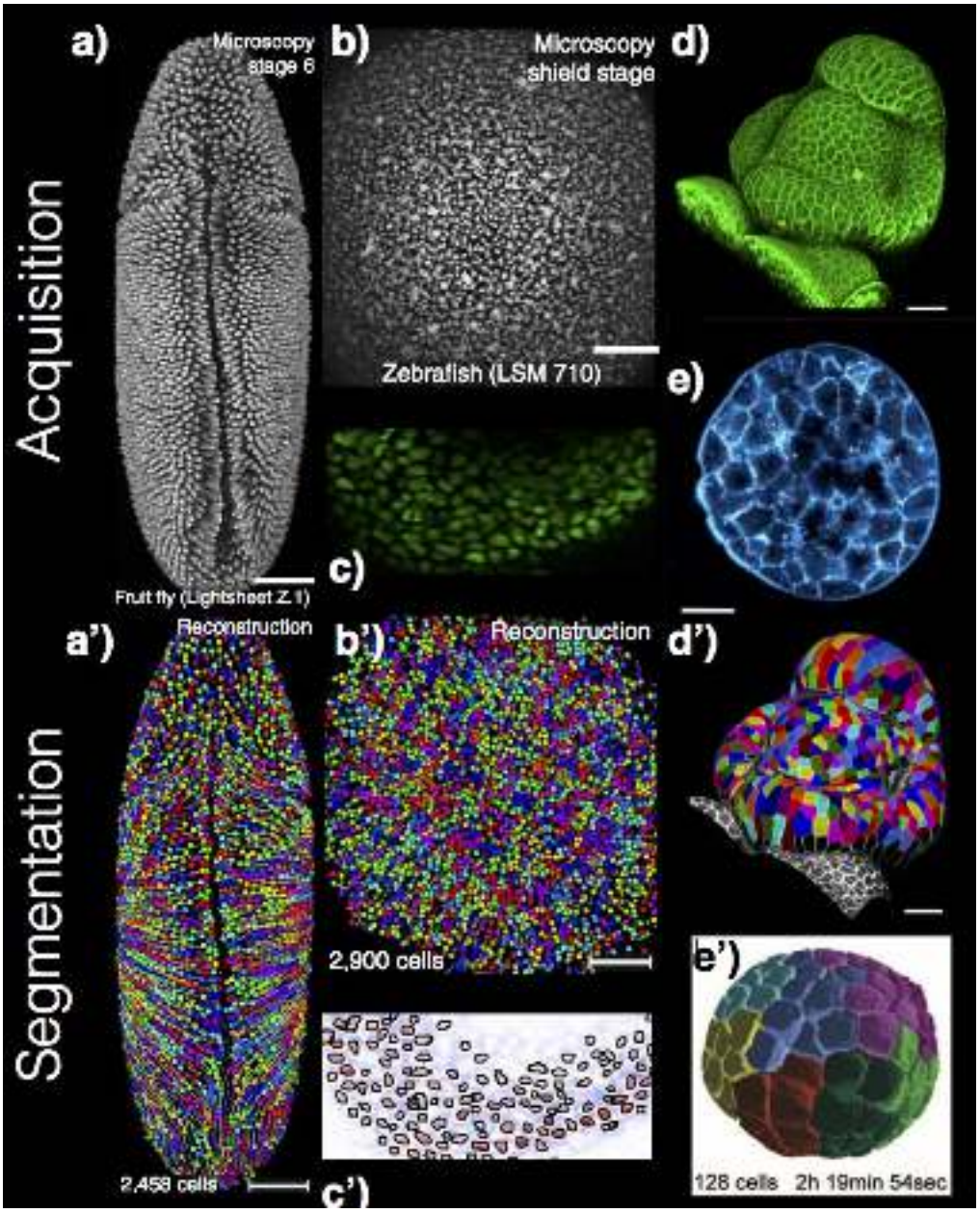


Figure 1.17 – Examples of segmentations. Acquisition images and their corresponding segmentations (' images). Drosophila ***a)*** *and zebrafish* ***b)*** *embryos where the nuclei have been stained and their corresponding segmentations* ***a')***, ***b')*** *from Amat et al. [2014]. A slice of a Caenorhabditis elegans embryo where the nuclei have been stained* ***c)*** *and its corresponding segmentation from Santella et al. [2010]. Membrane labeled Arabidopsis thaliana apical meristem* ***d)*** *and its corresponding segmentation from Barbier de Reuille et al. [2015]. A zebrafish embryo* ***e)*** *where the membranes have been labeled and its corresponding segmentation from Olivier et al. [2010]*

consecutive time points and if there are no cell death events. A sensitive parameter

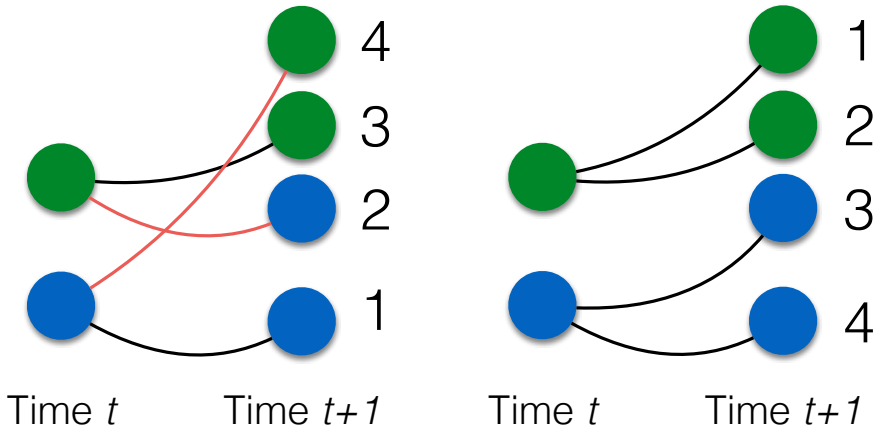


Figure 1.18 – This example shows two results given by a similar greedy algorithm where the order of treatment of the cells was changed. The algorithm maps every cell at t+1 to the cells at t, a cell at t can have at most two daughter cells. The number of the cells corresponds to their treatment order. On the left, the tracking algorithm matches the cell 1 to its corresponding blue cell. Then the cell 2 is wrongly assigned to the green cell since it is closest. The cell number 3 is matched to the green cell which locks it (the maximum number of daughters for the green cell is reached). Ultimately, the cell 4 is matched to the last possible cell, the blue, raising an second error. The second example, on the right, raises a correct mapping.

of this algorithm is how the distance between two cells in consecutive segmentations is computed.

The tracking as an optimization problem. Nearest antecedent algorithms are greedy algorithms. If the problem is constrained, for example if a cell $i \in \mathcal{S}_t$ has a maximum number of descendants in $\mathcal{S}_{t+1}$, then the result can depend on the order of treatment of the cells. This can result in non optimal pairing and therefore errors in the tracking (Figure 1.18). To overcome this issue, Fernandez et al. [2010] proposed to formulate the tracking problem as an optimization problem. The tracking $\mathcal{T}$ has to minimize a given cost function $\mathcal{C}(\mathcal{T})$, which for a given mapping can be described by the sum of the distances between cells that are paired, $\mathcal{C}(\mathcal{T}) = \sum_{(i,j)\in\mathcal{E}} d(i,j)$ where i is a cell in $\mathcal{S}_t$, j is a cell in $\mathcal{S}_{t+1}$ and $(i,j) \in \mathcal{E} \Leftrightarrow \mathcal{T}(i) = j$. In this special given case, Fernandez et al. [2010] formulated this optimization problem as a minimal cost maximum flow problem (as in Cilla et al. [2015]) which, using linear programming formulation, can always be resolved in a polynomial time (when $d \in \mathbb{N},\ \forall (i,j) \in \mathcal{S}_t \times \mathcal{S}_{t+1}$).

The tracking problem is even harder when cells can enter or leave the field of view, in the case of biological samples too large to be imaged in their entirety. It also becomes harder when the cells move significantly faster than the rate of 3D image acquisition. The tracking problem can be greatly eased by decreasing the time between two consecutive segmented images and by imaging the whole organism. Because of their high

acquisition rate and ability to image whole organisms up to several hundred μm, light-sheet microscopes are particularly useful when cell segmentation and tracking are the main purpose of the study.

1.3.3 Coupling segmentation and tracking

Tracking methods described above also highly depend on the segmentation quality of the images, since segmented cells are used as an input for the mapping algorithms. Tracking algorithms are highly sensitive to segmentation mistakes (Figure 1.19). For example, when 1000 cells are followed over many time points, if the segmentation at each individual time-point has an accuracy of 98% (which is higher than the best published algorithms), after 50 time points, assuming that the errors are randomly distributed, the number of good tracks remaining would be on average $0.98^{50} * 1000 \approx 364$. Only 36.4% of the tracks would be accurate from the beginning to the end of the acquisition process. To have 95% of the tracks correctly tracked over the whole imaging period, the error in the segmentation should be less than 0.1%. Therefore, the segmentation algorithms would have to yield extremely accurate results. Which is, as seen previously, extremely hard to achieve without extensive manual correction.

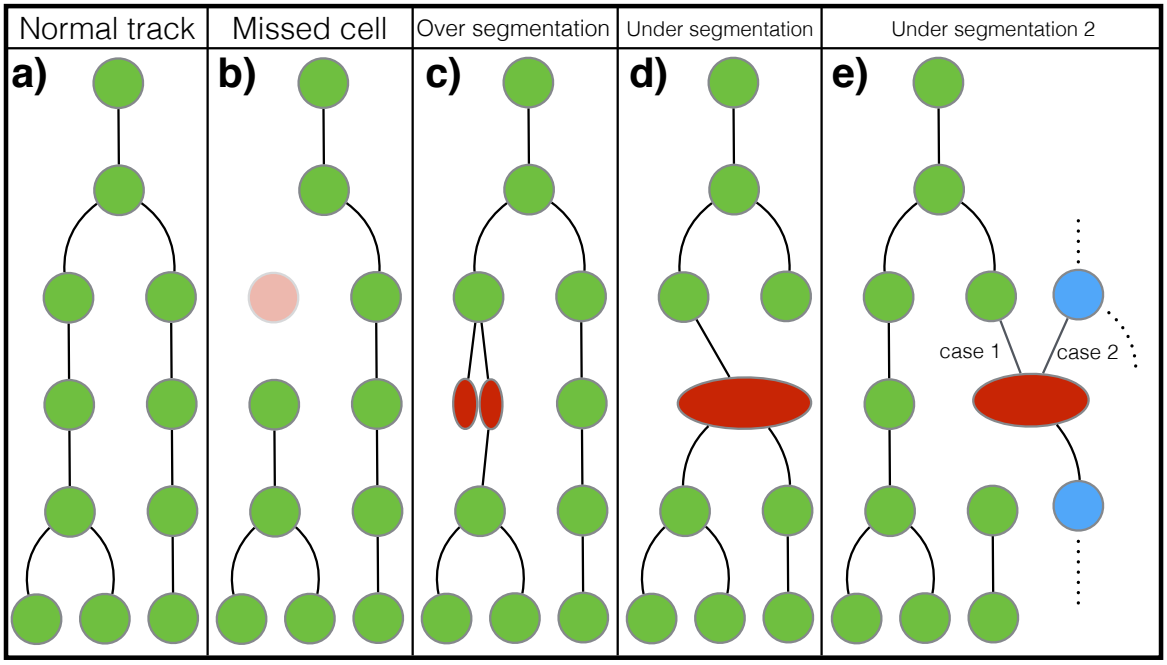


Figure 1.19 – Possible tracking mistakes induced by segmentation mistakes. ***a)*** *Ground truth track.* ***b)*** *a cell is missed by the segmentation, the track is broken.* ***c)*** *a cell is temporarily over segmented, a fake division is induced followed by a cell death.* ***d)*** *two sister cells are temporarily fused, resulting in a fake cell death followed by a division.* ***e)*** *Two cells from different tracks are fused.*

1.3.3.1 Relationship between segmentation and tracking

The error model described in the previous paragraph assumes that one mistake in the segmentation leads to the interruption of one cell track (Figure 1.19). It can actually be worse. The most common segmentation errors are over and under segmented cells. A cell that is over-segmented (namely artefactually split into two cells) creates a fake division followed by an artefactual cell death (Figure 1.19 **c**). An under-segmentation event (cell fusion), at best, creates a cell death followed by a division (Figure 1.19 **d**) when the fusion occurs between sister cells. If the fusion, however, links two cells belonging to different lineages (Figure 1.19 **e**) it links two unrelated lineages with a significant impact on the global topologies of the cell lineage of the organism. To avoid such mistakes, some algorithms detect segmentation errors and exclude these cells from the tracking. This reduces the number of tracked cells but ensures that the constructed cell lineages can be trusted [Al-Kofahi et al., 2006].

1.3.3.2 Performing segmentation and tracking in parallel.

Liu et al. [2014] proposed a method that couples segmentation and tracking. This methods relies on the fact that mistakes are due to stochastic fluctuations of imaging quality and are thus not expected to be stable in time. Therefore, generating differently parametrized segmentations of a given time-series and looking at the stability of the shape of the cells through time should yield a good segmentation. To do so, a set of segmentations with different parametrization is generated for each time-point of the series, using a seeded watershed. These different segmentations are linked through time using standard tracking algorithms as described above. The links are weighted by the percentage of coverage of two consecutive cell snapshots. Then, for each cell snapshot from the first time-point, the most likely track is chosen among all the possible tracks. Therefore, each final segmented cell snapshot of a given time-point can originate from the segmentations obtained with different parametrization. Unfortunately, this method is highly expensive in computing time when the dataset increases in size.

Amat et al. [2014] propose an alternative strategy to couple segmentation and tracking and applied it to nuclei segmentation. This algorithm propagates segmentations at one time t to the next time $t+1$. Using deformable models and division detection algorithms, the segmentation from time t is deformed to match the intensity image of time $t+1$, yielding a predicted segmentation at time $t+1$. Assuming that the segmentation is accurate at time t, the propagation greatly helps the segmentation of time $t+1$ and simultaneously builds it at the same time as the lineage tracks of the cells. Then, the temporal context is incorporated in the form of biological *a-priori* to constrain the segmentation and the cell tracking. This temporal context allows to compute a score for each cell reflecting their likelihood to be accurately segmented. When cells are tagged with a low score, the source of the error can either be automatically corrected by the algorithm, or the cell is tagged to be manually checked and potentially corrected.

These two methods take advantage of the low temporal resolution permitted by the latest microscopy techniques. They consider each time-series as a coherent 4D object to segment and track rather than as a set of independent 3D images. This new formalisation of the segmentation and tracking problem can help release some constraints on the prior knowledge of the system and can decrease the impact of the inhomogeneity of the images.

1.3.4 From the segmentation to the 4D digitalized embryo

From the segmentations, metrics on the segmented object can be extracted such as the number of cells, the division timings, the life span of the cells [Santella et al., 2010; Fernandez et al., 2010; Amat et al., 2014]. For the nuclei images, the positions of the nuclei, the pairwise euclidian distances and the orientation of division are also usually extracted [Amat et al., 2014; Moore et al., 2013]. In the case of cells snapshots, multiple metrics on the shape of the cells can be added such as their volume, anisotropy, compactness and sphericity [Tassy et al., 2006; Blanchard et al., 2009]. Moreover, the membrane images allow to compute the cell-cell area of contact [Tassy et al., 2006]. All these metrics then allow quantitative description of the imaged organisms.

1.4 Ascidians as model organisms in developmental biology

During my PhD, I focused my work on the analysis of the early development of the ascidian *Phallusia mammillata*.

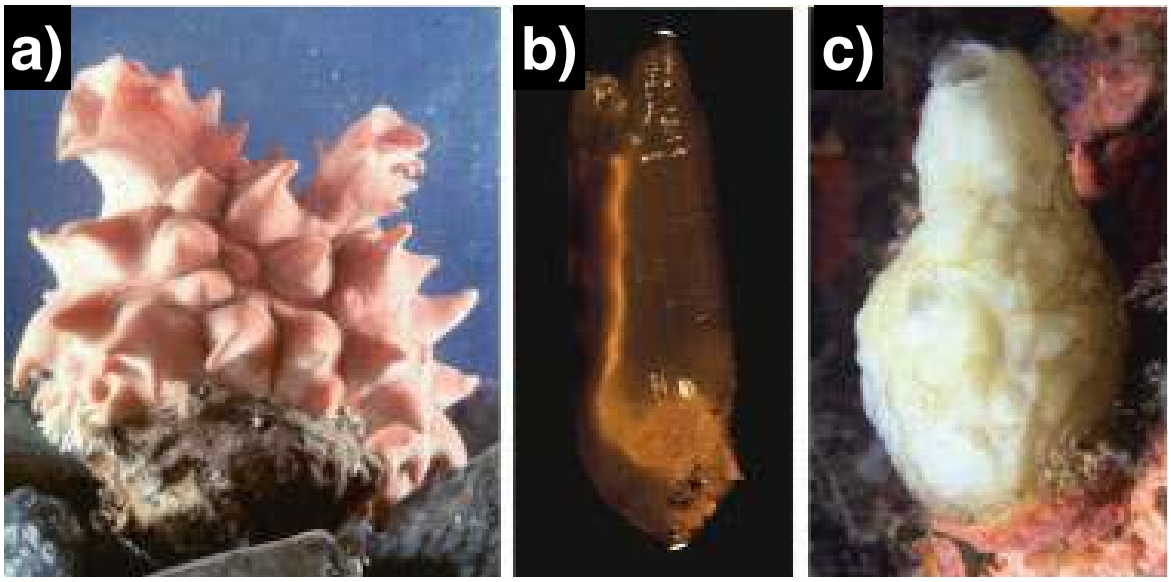


Figure 1.20 – Photographies of adult ascidians. ***a)*** *Halocynthia roretzi,* ***b)*** *Ciona intestinalis,* ***c)*** *Phallusia mammillata. (adapted from Kumano and Nishida [2007])*

Ascidians are marine invertebrates, which live all around the world, usually in shallow water (some can live in deep sea). They belong to the tunicates, the sister group of vertebrates. Ascidians undergo a rapid and stereotyped embryonic development based on invariant cell lineages from an egg to a tadpole larva [Lemaire, 2009]. They then go through a complex metamorphosis resulting in quite differently shaped adult animals (Figure 1.20). They also share a bilateral symmetry during their embryonic development. All studied solitary ascidian species share an almost identical embryonic development. Depending on the species, embryo development lasts from a few hours to several days at a constant volume, a small number of cells (around 2600 cells at larval stage) with few described instances of cell rearrangement. Due to the invariant cell lineages, the developmental stages can be described by the number of cells of the embryo up to the onset of the gastrulation. The developmental stages are then defined by the major events of the development: the internalisation of the mesendoderm (gastrulation), the closure of the neural tube (neurulation), and the extension of the tail (tailbud stages) and the onset of swimming behaviour (hatched larva) (Figure 1.21).

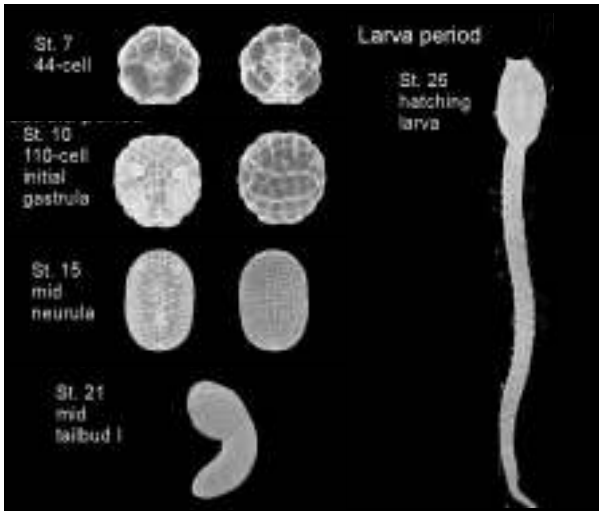


Figure 1.21 – A few developmental stages of C. intestinalis. Left: Vegetal side, right: Animal side. Top is anterior (adapted from Hotta et al. [2007a])

These developmental stages are conserved among most ascidians. Conklin showed in 1905 that ascidian embryos of the Styela fenus develop with a precise stereotyped cell lineage, which he described [Conklin, 1905]. This precise cell lineage allowed the naming of every cell using a simple rule up to the 112 cell stage. Cell lineage trees were subsequently reconstructed up to the same stage for the two species *Halocynthia roretzi* and *Ciona intestinalis* [Nishida, 1987; Kumano and Nishida, 2007]. In the case of *Halocynthia*, Nishida and Satoh [Nishida, 1987; Satoh, 2001] showed that by the onset of gastrulation, most embryonic blastomeres will only give rise to a single type of tissue (e.g. Notochord, muscle, epidermis, neural tissue). This early fate restriction suggests that a large fraction of the fate specification events have taken place by the onset of

gastrulation (reviewed in [Lemaire, 2009].

In parallel, genomes for several ascidian species have been sequenced and are available to the community through the ANISEED portal. This sequencing allowed a characterization of genomic differences within and between ascidian species. A high level of polymorphism was found within *Ciona intestinalis* and *Ciona savignyi* species [Dehal et al., 2002; Small et al., 2007] and also between them [Kim et al., 2007]. This surprisingly high genomic divergence is translated into distinct regulatory logics. For instance, Stolfi et al. [2014] showed that the regulatory syntax between *Molgula* and *Ciona* is different but triggers the same expression profiles. This striking dichotomy between the slow morphological evolution of ascidians and their rapid molecular evolution raises a paradox. A similar phenomenon can also be observed in nematodes, which also develop with invariant cell lineages and a small cell number [Yanai and Hunter, 2009].

While the genomic differences have started to be quantified, it is still necessary to carefully quantify the morphogenesis and the similarities and differences within and between ascidian species. Therefore a quantitative description of the development of the concerned species is necessary in order to quantitatively assess how constrained the development is.

The invariant cell lineage, early fate specification, small number of cells, small size of the embryos (especially *C. intestinalis* and *P. mammillata*) and their transparency (especially *P. mammillata*) make ascidians good model organisms for developmental biology. The invariant lineage and the simplicity of the development allow to describe and compare different embryos. The transparency and the small size allow high speed (because of the small volume to image) and precise in-depth imaging.

1.4.1 Ascidian cell lineage

Ascidian lineages have been extensively studied. In *Ciona* and *Halocynthia*, cell division timings are known for all the cells up to at least the 112 cell stage [Lemaire, 2009]. In *C. intestinalis* neural plate cell lineages have been inferred up to the tailbud stages by comparison of carefully staged fixed embryos [Nicol and Meinertzhagen, 1988; Hudson and Lemaire, 2001; Cole and Meinertzhagen, 2004; Hudson and Yasuo, 2005; Lemaire, 2009]. Moreover, the time of fate restriction of each cell during the early development is in most cases precisely known (Figure 1.22) [Lemaire, 2009].

At the 112 cell stage, only few bilateral cells do not have their fate restricted. Trunk lateral and ventral cells, mesenchyme cells and neural plate cells are incompletely restricted and will give rise to several tissues. Moreover, the B7.2 cell pair will give rise to the posterior head endoderm and endodermal strand, b8.17 cell pair will give rise to secondary muscle and secondary notochord and A8.16 that will result in secondary muscle and tail lateral neural plate. All the other cells have their fate restricted to a simple larval tissue by this 112 cell stage.

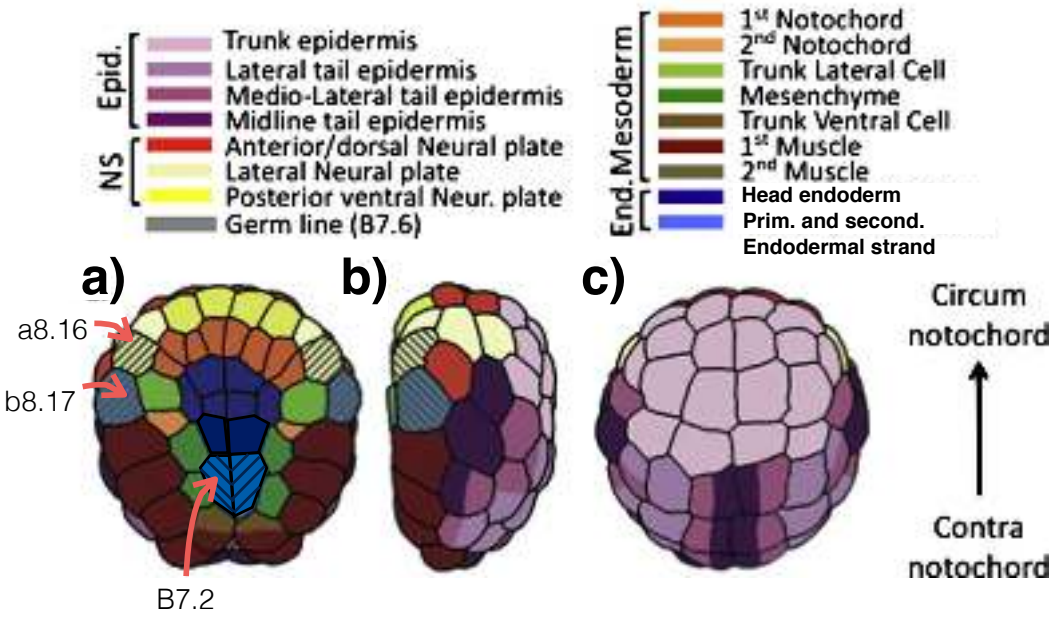


Figure 1.22 – Fate map of the 112 cell stage embryo, the colors represent the different fates. ***a)*** *Vegetal view,* ***b)*** *Lateral view,* ***c)*** *Animal view. (adapted from Lemaire [2009])*

Initial work by Chabry (1887) and Conklin (1905) suggested that ascidian embryos develop in a mosaic manner: cell fates result from the cell-autonomous inheritance of localized maternal determinants. This initial view has changed and it is now thought that most cell fates are achieved via cell induction processes, as in more complex embryos. One exception to this rule appears to be the primary muscle lineage, studied by Chabry and Conklin, which forms as a result of the inheritance of a localized maternal transcription factor, macho-1 [Nishida and Sawada, 2001].

The causes that lead to a mother cell to divide into two daughter cells belonging to different tissues have been elucidated for some cases (see Table 1.3). These inductions are often the result of differential signalling from the direct neighbouring cells. These signals can either induce a fate in one of the two daughter cells or polarize the mother cell cortex into regions that will be differentially inherited by the two daughters. For example the mother cell B6.4 divides into B7.7 that gives mesenchyme and B7.8 that gives primary muscle. Imai et al. [2002] showed that the FGF pathway is responsible for the polarization of the mother through the activation of the ERK kinase.

1.4.2 Quantification of ascidian embryonic morphogenesis

In vertebrates, inducers usually act at long range [Gurdon et al., 1994; McDowell et al., 1997], but in the simple ascidian embryos made of very few cells, such a mechanism would probably lead to the induction of all competent cells. An initial suggested answer came from work carried out by Tassy et al. [2006]. They developed a manual segmentation protocol from 3D two-photon images of fixed embryos (see also section 1.3). This manual

Mother cell	Fate daughter 1	Fate daughter 2	Induction timing	Pathway	Ref
A6.2	Post ventral NP (A7.4)	Notochord (A7.3)	Mother	FGF/ERK	(1)
			Daughter 2	FGF/ERK	(2)
			Daughter 1	Nodal	(2)
A6.4	Lateral NP/2^{nd} muscle (A7.8)	Notochord (A7.7)	Mother	FGF/ERK	(1)
			Daughter 2	FGF/ERK	(2)
			Daughter 2	Nodal	(3)
A6.3	TLC (A7.6)	Head Endoderm (A7.5)	Mother	FGF/ERK	(4)
			Daughter 1	Nodal	(5)
B6.3	Germ cell (B7.5)	TVC (B7.6)	Cell autonomous		
B6.4	Primary Muscle (B7.8)	Mesenchyme (B7.7)	Mother (Daugthers)	FGF/ERK	(6)
a6.7	Head epidermis (a7.14)	Lateral NP (a7.13)	Daughter 2	BMP	(7)
A7.4	NP (col 1) (A8.8)	NP (col 2) (A8.7)	Daughter 1	Notch	(8)
A7.8	Lat NP (A8.15)	2^{nd} muscle/Lat NP (A8.16)	Mother	Notch	(8)
B7.3	Notochord (B8.6)	Mesenchyme (B8.5)	Mother	Notch	(5)
			Daughters	Notch	(5)
a7.9	NP (row III/IV)	NP (row V/VI)	Daughters	ERK[1]	(9)
a7.10	NP (row III/IV)	NP (row V/VI)	Daughters	ERK[1]	(9)

Table 1.3 – Table of the fate inductions known in ascidians in the transitions from the 32 to the 64 cell stage (first part of the table) and from the 64 to the 112 cell stage (second part of the table) (table built by P. Lemaire). [1]*: ligands unknown. References: (1): [Picco et al., 2007], (2): [Yasuo and Hudson, 2007], (3): [Hudson and Yasuo, 2005], (4): [Shi and Levine, 2008], (5): [Hudson and Yasuo, 2006], (6): [Imai et al., 2002], (7): [Ohta and Satou, 2013], (8) [Hudson et al., 2007], (9): [Wagner and Levine, 2012]*

segmentation method allowed to reconstruct 19 3D digital embryos between the 2 and 44 cell stages. Shape descriptors added to these 3D reconstructions allowed to quantify the cell shapes and cell-cell contacts in *C. intestinalis* embryos. This analysis revealed that during the induction of ectodermal cells by underlying FGF expressing vegetal cells, the area of contacts between inducing cells and cells competent to respond to FGF is a major determinant of induction.

Thus, in these simple embryos, it appears that inducing signals may only act at a very short range (i.e. in a juxtacrine manner). Whether this is a specific feature of neural induction, or a more general property of embryonic inductions remained unknown at the

time. Interestingly, the percentage of surface of induced cells contacting the inducing cells is evolutionarily conserved between distantly-related embryos [Tassy et al., 2006].

Manual segmentation was also used in Nakamura et al. [2012] to segment the 1579 cells of a *C. intestinalis* tailbud embryo from 3D confocal images (Figure 1.23). The number of cells of the different tissues were extracted from this segmentation allowing a precise and quantitative description of the embryo. This manual segmentation also lead to the discovery of two populations of cells that were not previously described.

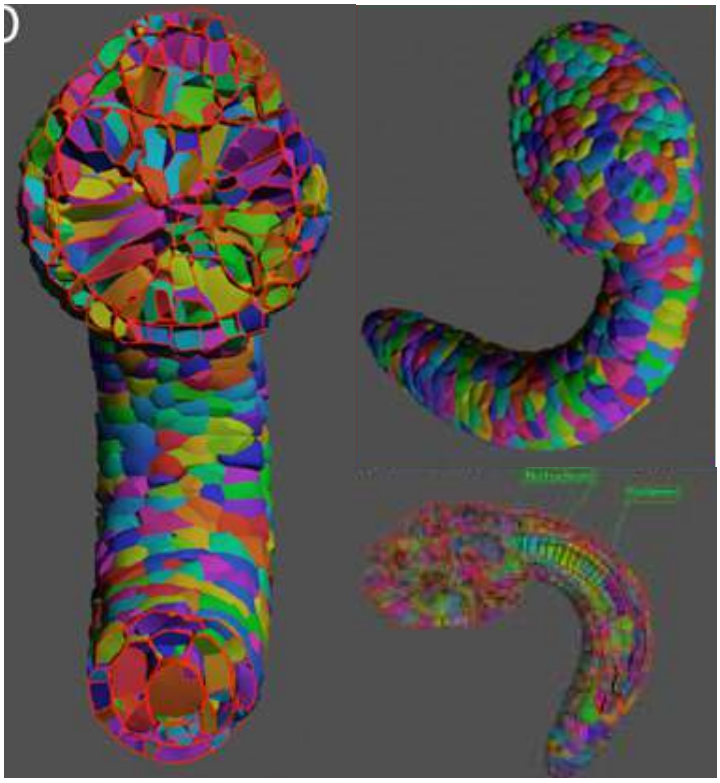


Figure 1.23 – Result of the manual segmentation of C. intestinalis done in Nakamura et al. [2012] (adapted from Nakamura et al. [2012])

Manual segmentation was also used in Sherrard, K. and Robin, F. et al. [2010] to quantitatively describe the shape of the endodermal cells during the first step of gastrulation in *C. intestinalis*, *P. mammillata* and *B. villosa* embryos. Quantitative description of the apical surface area and the apico-basal cell height in different cells was used as input data in a computational model of the mechanical drivers of the gastrulation; the endodermal cells flatten and then invaginate which pulls on the surrounding cells (see Figure 1.2).

1.4.3 From manual to semi-automated segmentation

Thus, in ascidian embryos, manual segmentation allowed to discover new cell groups, induction mechanisms and to model the invagination process. Yet, the time expense of the manual reconstruction ($\approx$ 30 minutes/cell) limits reconstruction efforts to a small number of early embryos (reconstructing a single embryo every two minutes until the tailbud stage would take 3 years of manual work, see section 1.3). To solve this problem and enable a better comprehension of notochord elongation during *C. savigny* embryogenesis, Veeman and Smith [2013] proposed a semi-automated segmentation procedure for cell membranes using a manually seeded 3D watershed method (see section 1.3.1.2). This allowed to quantify the evolution in time of the shape of the notochord cells which, coupled to the quantification of the cell positions, led to a quantitative model of the elongation of the notochord. This method was then improved in Delibaltov et al. [2013] (see section 1.3.1.2) to avoid the use of manual seeding but this method has so far not been used to quantify morphogenesis (note that later in Carlson et al. [2015] they manually tracked the cells without using any segmentation tool).

This semi-automated method is still time-consuming and computationally intensive. It was so far only used to segment a part of an embryo (notochord cells). To fully describe and extract shape quantifiers from a set of multiple embryos, a fully automated method would be necessary.

1.5 Aims of the PhD work

The work presented in this manuscript was carried out with a joint affiliation to the laboratories of my two supervisors, P. Lemaire and C. Godin. The initial aim was to develop a fully automated segmentation pipeline able to cope with the large volume of light-sheet imaging data that a post-doctoral scientist in the Lemaire laboratory, Ulla-Maj Fiuza, was starting to generate using live *Phallusia mammillata* embryos.

In order to exploit these unique datasets, I developed a pipeline to automatically segment and track *Phallusia mammillata* cell shapes throughout a long period of imaging. The cell lineage tree built with this pipeline allowed to study the division patterns of the different cells. This study showed that these patterns alone allow to cluster the cells according to their fate and then allow to pinpoint new fate specification events. Ultimately, the segmentation output coupled to ANISEED database [Brozovic et al., 2015] allowed to model the inductions during the cell specification events. This model shows that a precise control of cell-cell area of contact is necessary to induce and restrict cell fates. The segmentation and tracking pipelines and the associated results are described in Chapter 2.

4D digital embryos are produced by the previous segmentation and tracking algorithm. During my PhD, I mostly analyzed a single embryo. The ultimate aim, however, is to quantify embryonic variability between and within species, in wild-type and in

experimentally manipulated embryos. For this, it is necessary to compare the development of several embryos. The first step of this procedure is to align and register two embryos in space and time. I therefore developed a method to spatio-temporally register two intensity image sequences based on the analysis of the deformations the embryos undergo. I also participated in the development of a pipeline to spatially register two segmented embryos based on the bilateral symmetry and the stereotypy of ascidian embryonic development. These two methods are described in Chapter 3.

Altogether, these pipelines establish a basis to build average 4D digital embryos of ascidians and open the field to an extensive quantitative description of embryonic morphogenesis in ascidians embryos.

References

Abud, H. E. (2004). Shaping developing tissues by apoptosis. Cell Death Differ, 11(8):797–799. 7

Adams, R. and Bischof, L. (1994). Seeded region growing. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 16(6):641–647. 25

Al-Kofahi, O., Radke, R. J., Goderie, S. K., Shen, Q., Temple, S., and Roysam, B. (2006). Automated cell lineage construction: A rapid method to analyze clonal development established with murine neural progenitor cells. Cell Cycle, 5(3):327–335. 21, 23, 31

Amat, F., Höckendorf, B., Wan, Y., Lemon, W. C., McDole, K., and Keller, P. J. (2015). Efficient processing and analysis of large-scale light-sheet microscopy data. Nature protocols, 10(11):1679–1696. 18

Amat, F., Lemon, W., Mossing, D. P., McDole, K., Wan, Y., Branson, K., Myers, E. W., and Keller, P. J. (2014). Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. Nat Meth, 11(9):951–958. 20, 24, 26, 28, 31, 32

Arranz, A., Dong, D., Zhu, S., Rudin, M., Tsatsanis, C., Tian, J., and Ripoll, J. (2013). Helical optical projection tomography. Opt Express, 21(22):25912–25925. 11

Arranz, A., Dong, D., Zhu, S., Savakis, C., Tian, J., and Ripoll, J. (2014). In-vivo optical tomography of small scattering specimens: time-lapse 3d imaging of the head eversion process in drosophila melanogaster. Scientific Reports, 4:7325 EP –. 11

Bao, Z., Murray, J. I., Boyle, T., Ooi, S. L., Sandel, M. J., and Waterston, R. H. (2006). Automated cell lineage tracing in caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America, 103(8):2707–2712. 21, 23, 27

Barbier de Reuille, P., Routier-Kierzkowska, A.-L., Kierzkowski, D., Bassel, G. W., Schüpbach, T., Tauriello, G., Bajpai, N., Strauss, S., Weber, A., Kiss, A., Burian, A., Hofhuis, H., Sapala, A., Lipowczan, M., Heimlicher, M. B., Robinson, S., Bayer, E. M.,

Basler, K., Koumoutsakos, P., Roeder, A. H., Aegerter-Wilmsen, T., Nakayama, N., Tsiantis, M., Hay, A., Kwiatkowska, D., Xenarios, I., Kuhlemeier, C., and Smith, R. S. (2015). Morphographx: A platform for quantifying morphogenesis in 4d. eLife, 4. 28

Bassi, A., Schmid, B., and Huisken, J. (2015). Optical tomography complements light sheet microscopy for in toto imaging of zebrafish development. Development, 142(5):1016–1020. 11

Bertet, C., Sulak, L., and Lecuit, T. (2004). Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. Nature, 429(6992):667–671. 8

Besnard, F., Refahi, Y., Morin, V., Marteaux, B., Brunoud, G., Chambrier, P., Rozier, F., Mirabet, V., Legrand, J., Laine, S., Thevenon, E., Farcot, E., Cellier, C., Das, P., Bishopp, A., Dumas, R., Parcy, F., Helariutta, Y., Boudaoud, A., Godin, C., Traas, J., Guedon, Y., and Vernoux, T. (2013). Cytokinin signalling inhibitory fields provide robustness to phyllotaxis. Nature, advance online publication:–. 10

Blanchard, G. B., Kabla, A. J., Schultz, N. L., Butler, L. C., Sanson, B., Gorfinkiel, N., Mahadevan, L., and Adams, R. J. (2009). Tissue tectonics: morphogenetic strain rates, cell shape change and intercalation. Nat Meth, 6(6):458–464. 32

Blankenship, J. T., Backovic, S. T., Sanny, J. S. P., Weitz, O., and Zallen, J. A. (2006). Multicellular rosette formation links planar cell polarity to tissue morphogenesis. Dev Cell, 11(4):459–470. 8

Bloch, I., Gousseau, Y., Maître, H., Matignon, D., Pesquet-Popescu, B., Schmitt, F., Sigelle, M., and Tupin, F. (2005). Le traitement des images, volume 1, 2, 3. 21

Boudon, F., Chopard, J., Ali, O., Gilles, B., Hamant, O., Boudaoud, A., Traas, J., and Godin, C. (2015). A computational framework for 3d mechanical modeling of plant morphogenesis with cellular resolution. PLoS Comput Biol, 11(1):e1003950. 10

Breiman, L. (2001). Random forests. Machine Learning, 45(1):5–32. 23

Broker, L. E., Kruyt, F. A., and Giaccone, G. (2005). Cell death independent of caspases: A review. Clinical Cancer Research, 11(9):3155–3162. 7

Brown, K. E., Keller, P. J., Ramialison, M., Rembold, M., Stelzer, E. H., Loosli, F., and Wittbrodt, J. (2010). Nlcam modulates midline convergence during anterior neural plate morphogenesis. Developmental Biology, 339(1):14 – 25. 27

Brozovic, M., Martin, C., Dantec, C., Dauga, D., Mendez, M., Simion, P., Percher, M., Laporte, B., Scornavacca, C., Gregorio, A. D., Fujiwara, S., Gineste, M., Lowe, E. K., Piette, J., Racioppi, C., Ristoratore, F., Sasakura, Y., Takatori, N., Brown, T. C., Delsuc, F., Douzery, E., Gissi, C., McDougall, A., Nishida, H., Sawada, H., Swalla, B. J., Yasuo, H., and Lemaire, P. (2015). Aniseed 2015: a digital framework for the comparative developmental biology of ascidians. Nucleic Acids Research. 38

Carlson, M., Reeves, W., and Veeman, M. (2015). Stochasticity and stereotypy in the ciona notochord. Dev Biol, 397(2):248–256. 38

Chen, B.-C., Legant, W. R., Wang, K., Shao, L., Milkie, D. E., Davidson, M. W., Janetopoulos, C., Wu, X. S., Hammer, J. A., Liu, Z., English, B. P., Mimori-Kiyosue, Y., Romero, D. P., Ritter, A. T., Lippincott-Schwartz, J., Fritz-Laylin, L., Mullins, R. D., Mitchell, D. M., Bembenek, J. N., Reymann, A.-C., Böhme, R., Grill, S. W., Wang, J. T., Seydoux, G., Tulu, U. S., Kiehart, D. P., and Betzig, E. (2014). Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. Science, 346(6208). 11

Cilla, R., Mechery, V., Hernandez de Madrid, B., Del Signore, S., Dotu, I., and Hatini, V. (2015). Segmentation and tracking of adherens junctions in 3d for the analysis of epithelial tissue morphogenesis. PLoS Comput Biol, 11(4):e1004124. 25, 29

Coen, E., Rolland-Lagan, A.-G., Matthews, M., Bangham, J. A., and Prusinkiewicz, P. (2004). The genetics of geometry. Proc Natl Acad Sci U S A, 101(14):4728–4735. 6, 7, 9

Cole, A. G. and Meinertzhagen, I. A. (2004). The central nervous system of the ascidian larva: mitotic history of cells forming the neural tube in late embryonic ciona intestinalis. Developmental Biology, 271(2):239 – 262. 34

Concha, M. and Adams, R. (1998). Oriented cell divisions and cellular morphogenesis in the zebrafish gastrula and neurula: a time-lapse analysis. Development, 125(6):983–994. 8

Conklin, E. G. (1905.). The organization and cell-lineage of the ascidian egg / by Edwin G. Conklin. Philadelphia :[Academy of Natural Sciences],. 33

Conradt, B. (2009). Genetic control of programmed cell death during animal development. Annual review of genetics, 43:493–523. 7

Davidovits, P. (1969). Scanning Laser Microscope. Nature, 223:831. 14

Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D. M., Harafuji, N., Hastings, K. E. M., Ho, I., Hotta, K., Huang, W., Kawashima, T., Lemaire, P., Martinez, D., Meinertzhagen, I. A., Necula, S., Nonaka, M., Putnam, N., Rash, S., Saiga, H., Satake, M., Terry, A., Yamada, L., Wang, H.-G., Awazu, S., Azumi, K., Boore, J., Branno, M., Chin-bow, S., DeSantis, R., Doyle, S., Francino, P., Keys, D. N., Haga, S., Hayashi, H., Hino, K., Imai, K. S., Inaba, K., Kano, S., Kobayashi, K., Kobayashi, M., Lee, B.-I., Makabe, K. W., Manohar, C., Matassi, G., Medina, M., Mochizuki, Y., Mount, S., Morishita, T., Miura, S., Nakayama, A., Nishizaka, S., Nomoto, H., Ohta, F., Oishi, K., Rigoutsos, I., Sano, M., Sasaki, A., Sasakura, Y., Shoguchi, E., Shin-i, T., Spagnuolo, A., Stainier, D., Suzuki, M. M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Hyatt, P. D., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine,

M., Satoh, N., and Rokhsar, D. S. (2002). The draft genome of ciona intestinalis: Insights into chordate and vertebrate origins. Science, 298(5601):2157–2167. 34

Delibaltov, D., Ghosh, P., Rodoplu, V., Veeman, M., Smith, W., and Manjunath, B. (2013). A linear program formulation for the segmentation of ciona membrane volumes. In Mori, K., Sakuma, I., Sato, Y., Barillot, C., and Navab, N., editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, volume 8149 of Lecture Notes in Computer Science, pages 444–451. Springer Berlin Heidelberg. 25, 26, 38

Denk, W., Strickler, J. H., and Webb, W. W. (1990). Two-photon laser scanning fluorescence microscopy. Science, 248(4951):73–76. 15

Fernandez, R., Das, P., Mirabet, V., Moscardi, E., Traas, J., Verdeil, J.-L., Malandain, G., and Godin, C. (2010). Imaging plant growth in 4d: robust tissue reconstruction and lineaging at cell resolution. Nat Meth, 7(7):547–553. 18, 22, 25, 26, 27, 29, 32

Fowlkes, C. C., Hendriks, C. L. L., Keränen, S. V., Weber, G. H., Rübel, O., Huang, M.-Y., Chatoor, S., DePace, A. H., Simirenko, L., Henriquez, C., Beaton, A., Weiszmann, R., Celniker, S., Hamann, B., Knowles, D. W., Biggin, M. D., Eisen, M. B., and Malik, J. (2008). A quantitative spatiotemporal atlas of gene expression in the drosophila blastoderm. Cell, 133(2):364 – 374. 10

Friedl, P. and Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. Nat Rev Mol Cell Biol, 10(7):445–457. 9

Gilbert, S. F. (2006). Developmental Biology., volume 99. Sinauer Publisher Inc. 3, 10

Graf, R., Rietdorf, J., and Zimmermann, T. (2005). Live cell spinning disk microscopy. Adv Biochem Eng Biotechnol, 95:57–75. 15

Gurdon, J. B., Harger, P., Mitchell, A., and Lemaire, P. (1994). Activin signalling and response to a morphogen gradient. Nature, 371(6497):487–492. 35

Hamant, O., Heisler, M. G., Jonsson, H., Krupinski, P., Uyttewaal, M., Bokov, P., Corson, F., Sahlin, P., Boudaoud, A., Meyerowitz, E. M., Couder, Y., and Traas, J. (2008). Developmental patterning by mechanical signals in arabidopsis. Science, 322(5908):1650–1655. 6, 9, 10

Hell, S. W. (2009). Microscopy and its focal switch. Nat Meth, 6(1):24–32. 13

Hotta, K., Mitsuhara, K., Takahashi, H., Inaba, K., Oka, K., Gojobori, T., and Ikeo, K. (2007a). A web-based interactive developmental table for the ascidian ciona intestinalis, including 3d real-image embryo reconstructions: I. from fertilized egg to hatching larva. Dev Dyn, 236(7):1790–1805. 33

Hotta, K., Yamada, S., Ueno, N., Satoh, N., and Takahashi, H. (2007b). Brachyury-downstream notochord genes and convergent extension in ciona intestinalis embryos. Development, Growth & Differentiation, 49(5):373–382. 9

Hudson, C. and Lemaire, P. (2001). Induction of anterior neural fates in the ascidian ciona intestinalis. Mech Dev, 100(2):189–203. 34

Hudson, C., Lotito, S., and Yasuo, H. (2007). Sequential and combinatorial inputs from nodal, delta2/notch and fgf/mek/erk signalling pathways establish a grid-like organisation of distinct cell identities in the ascidian neural plate. Development, 134(19):3527–3537. 36

Hudson, C. and Yasuo, H. (2005). Patterning across the ascidian neural plate by lateral nodal signalling sources. Development, 132(6):1199–1210. 34, 36

Hudson, C. and Yasuo, H. (2006). A signalling relay involving nodal and delta ligands acts during secondary notochord induction in ciona embryos. Development, 133(15):2855–2864. 36

Huisken, J. and Stainier, D. Y. R. (2007). Even fluorescence excitation by multidirectional selective plane illumination microscopy (mspim). Opt. Lett., 32(17):2608–2610. 11, 18

Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J., and Stelzer, E. H. K. (2004). Optical sectioning deep inside live embryos by selective plane illumination microscopy. Science, 305(5686):1007–1009. 16

Imai, K. S., Satoh, N., and Satou, Y. (2002). Early embryonic expression of fgf4/6/9 gene and its role in the induction of mesenchyme and notochord in ciona savignyi embryos. Development, 129(7):1729–1738. 35, 36

Jacobson, M. D., Weil, M., and Raff, M. C. (1997). Programmed cell death in animal development. Cell, 88(3):347–354. 7

Jahr, W., Schmid, B., Schmied, C., Fahrbach, F. O., and Huisken, J. (2015). Hyperspectral light sheet microscopy. Nat Commun, 6. 11

Jarośík, V., Kratochvíl, L., Honék, A., and Dixon, A. F. G. (2004). A general rule for the dependence of developmental rate on temperature in ectothermic animals. Proceedings of the Royal Society of London B: Biological Sciences, 271(Suppl 4):S219–S221. 14

Kausler, B., Schiegg, M., Andres, B., Lindner, M., Koethe, U., Leitte, H., Wittbrodt, J., Hufnagel, L., and Hamprecht, F. (2012). A discrete chain graph model for 3d+t cell tracking with high misdetection robustness. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, Computer Vision – ECCV 2012, volume 7574 of Lecture Notes in Computer Science, pages 144–157. Springer Berlin Heidelberg. 23

Keller, P. J., Schmidt, A. D., Santella, A., Khairy, K., Bao, Z., Wittbrodt, J., and Stelzer, E. H. K. (2010). Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. Nat Meth, 7(8):637–642. 23

Keller, P. J., Schmidt, A. D., Wittbrodt, J., and Stelzer, E. H. (2008). Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. Science, 322(5904):1065–1069. 11, 17, 21, 23, 27

Kennaway, R., Coen, E., Green, A., and Bangham, A. (2011). Generation of diverse biological forms through combinatorial interactions between tissue polarity and growth. PLoS Comput Biol, 7(6):e1002071. 7

Khairy, K. and Keller, P. J. (2011). Reconstructing embryonic development. genesis, 49(7):488–513. 20, 24

Khan, Z., Wang, Y.-C., Wieschaus, E. F., and Kaschube, M. (2014). Quantitative 4d analyses of epithelial folding during drosophila gastrulation. Development, 141(14):2895–2900. 26

Kim, J. H., Waterman, M. S., and Li, L. M. (2007). Diploid genome reconstruction of ciona intestinalis and comparative analysis with ciona savignyi. Genome Research, 17(7):1101–1110. 34

Klämbt, C., Jacobs, J., and Goodman, C. S. (1991). The midline of the drosophila central nervous system: A model for the genetic analysis of cell fate, cell migration, and growth cone guidance. Cell, 64(4):801 – 815. 9

Konig, K. (2000). Multiphoton microscopy in life sciences. J Microsc, 200(Pt 2):83–104. 15

Krzic, U., Gunther, S., Saunders, T. E., Streichan, S. J., and Hufnagel, L. (2012). Multiview light-sheet microscope for rapid in toto imaging. Nat Meth, 9(7):730–733. 11, 18

Kumano, G. and Nishida, H. (2007). Ascidian embryonic development: An emerging model system for the study of cell fate specification in chordates. Developmental Dynamics, 236(7):1732–1747. 32, 33

Lecuit, T. and Le Goff, L. (2007). Orchestrating size and shape during morphogenesis. Nature, 450(7167):189–192. 8, 9

Lecuit, T. and Lenne, P.-F. (2007). Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis. Nat Rev Mol Cell Biol, 8(8):633–644. 6

Lecuit, T., Lenne, P.-F., and Munro, E. (2011). Force generation, transmission, and integration during cell and tissue morphogenesis. Annu Rev Cell Dev Biol, 27:157–184. 5, 9

Lemaire, P. (2009). Unfolding a chordate developmental program, one cell at a time: invariant cell lineages, short-range inductions and evolutionary plasticity in ascidians. Dev Biol, 332(1):48–60. 5, 33, 34, 35

Lichtman, J. W. and Conchello, J.-A. (2005). Fluorescence microscopy. Nat Meth, 2(12):910–919. 13

Lièvre, M. (2014). Analysis and multiscale modelling of foliar growh in Arabidopsis thaliana in response to environmental stresses. Implication of the floral transition in the foliar expansion. PhD thesis, Université Montpellier 2. 25

Liu, K., Lienkamp, S., Shindo, A., Wallingford, J., Walz, G., and Ronneberger, O. (2014). Optical flow guided cell segmentation and tracking in developing tissue. In Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on, pages 298–301. 31

Long, F., Peng, H., Liu, X., Kim, S. K., and Myers, E. (2009). A 3d digital atlas of c. elegans and its application to single-cell analyses. Nat Meth, 6(9):667–672. 21, 23

Luengo Hendriks, C., Keranen, S., Fowlkes, C., Simirenko, L., Weber, G., DePace, A., Henriquez, C., Kaszuba, D., Hamann, B., Eisen, M., Malik, J., Sudar, D., Biggin, M., and Knowles, D. (2006). Three-dimensional morphology and gene expression in the drosophila blastoderm at cellular resolution i: data acquisition pipeline. Genome Biology, 7(12):R123. 10

Mahou, P., Vermot, J., Beaurepaire, E., and Supatto, W. (2014). Multicolor two-photon light-sheet microscopy. Nat Meth, 11(6):600–601. 11

Maitre, J.-L., Berthoumieux, H., Krens, S. F. G., Salbreux, G., Julicher, F., Paluch, E., and Heisenberg, C.-P. (2012). Adhesion functions in cell sorting by mechanically coupling the cortices of adhering cells. Science, 338(6104):253–256. 10, 24

Marr, D. and Hildreth, E. (1980). Theory of edge detection. Proc R Soc Lond B Biol Sci, 207(1167):187–217. 23

McDowell, N., Zorn, A. M., Crease, D. J., and Gurdon, J. B. (1997). Activin has direct long-range signalling activity and can form a concentration gradient by diffusion. Curr Biol, 7(9):671–681. 35

Michelin, G., Guignard, L., Fiuza, U.-M., and Malandain, G. (2014). Embryo Cell Membranes Reconstruction by Tensor Voting. In ISBI - International Symposium on Biomedical Imaging, Beijing, China. IEEE. 23, 25

Micusik, B. and Hanbury, A. (2005). Steerable semi-automatic segmentation of textured images. In Kalviainen, H., Parkkinen, J., and Kaarna, A., editors, Image Analysis, volume 3540 of Lecture Notes in Computer Science, pages 35–44. Springer Berlin Heidelberg. 20

Milani, P., Braybrook, S. A., and Boudaoud, A. (2013). Shrinking the hammer: micromechanical approaches to morphogenesis. Journal of Experimental Botany, 64(15):4651–4662. 10

Minc, N. and Piel, M. (2012). Predicting division plane position and orientation. Trends in Cell Biology, 22(4):193 – 200. 8

Moore, J. L., Du, Z., and Bao, Z. (2013). Systematic quantification of developmental phenotypes at single-cell resolution during embryogenesis. Development, 140(15):3266–3274. 32

Mosaliganti, K. R., Noche, R. R., Xiong, F., Swinburne, I. A., and Megason, S. G. (2012). *ACME: Automated Cell Morphology Extractor* for comprehensive reconstruction of cell membranes. PLoS Comput Biol, 8(12):e1002780. 23, 25, 26

Munro, E. M. and Odell, G. (2002). Morphogenetic pattern formation during ascidian notochord formation is regulative and highly robust. Development, 129(1):1–12. 5, 8

Nakamura, M. J., Terai, J., Okubo, R., Hotta, K., and Oka, K. (2012). Three-dimensional anatomy of the ciona intestinalis tailbud embryo at single-cell resolution. Dev Biol, 372(2):274–284. 19, 37

Nicol, D. and Meinertzhagen, I. A. (1988). Development of the central nervous system of the larva of the ascidian, ciona intestinalis l. ii. neural plate morphogenesis and cell lineages during neurulation. Dev Biol, 130(2):737–766. 5, 34

Nishida, H. (1987). Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. Developmental Biology, 121(2):526 – 541. 33

Nishida, H. and Sawada, K. (2001). macho-1 encodes a localized mrna in ascidian eggs that specifies muscle fate during embryogenesis. Nature, 409(6821):724–729. 35

Ohta, N. and Satou, Y. (2013). Multiple signaling pathways coordinate to induce a threshold response in a chordate embryo. PLoS Genet, 9(10):e1003818. 36

Olivier, N., Luengo-Oroz, M. A., Duloquin, L., Faure, E., Savy, T., Veilleux, I., Solinas, X., Débarre, D., Bourgine, P., Santos, A., Peyriéras, N., and Beaurepaire, E. (2010). Cell lineage reconstruction of early zebrafish embryos using label-free nonlinear microscopy. Science, 329(5994):967–971. 11, 28

Otsu, N. (1975). A threshold selection method from gray-level histograms. Automatica, 11(285-296):23–27. 23

Paluch, E. and Heisenberg, C.-P. (2009). Biology and physics of cell shape changes in development. Curr Biol, 19(17):R790–9. 6, 10, 24

Pawley, J. (2006). Handbook of Biological Confocal Microscopy. Springer. 14

Pecreaux, J., Zimmer, C., and Olivo-Marin, J. (2006). Biophysical active contours for cell tracking i: Tension and bending. In Image Processing, 2006 IEEE International Conference on, pages 1949–1952. 24

Picco, V., Hudson, C., and Yasuo, H. (2007). Ephrin-eph signalling drives the asymmetric division of notochord/neural precursors in ciona embryos. Development, 134(8):1491–1497. 5, 36

Ripoll, J., Koberstein-Schwarz, B., and Ntziachristos, V. (2015). Unleashing optics and optoacoustics for developmental biology. Trends in Biotechnology, pages –. 11, 16

Rombough, P. (2003). Development rate (communication arising): Modelling developmental time and temperature. Nature, 424(6946):268–269. 14

Santella, A., Du, Z., Nowotschin, S., Hadjantonakis, A.-K., and Bao, Z. (2010). A hybrid blob-slice model for accurate and efficient detection of fluorescence labeled nuclei in 3d. BMC Bioinformatics, 11:580. 23, 28, 32

Satoh, N. (2001). Ascidian embryos as a model system to analyze expression and function of developmental genes. Differentiation, 68(1):1–12. 33

Schiegg, M., Hanslovsky, P., Haubold, C., Koethe, U., Hufnagel, L., and Hamprecht, F. A. (2015). Graphical model for joint segmentation and tracking of multiple dividing cells. Bioinformatics, 31(6):948–956. 23, 24

Schiegg, M., Hanslovsky, P., Kausler, B. X., Hufnagel, L., and Hamprecht, F. A. (2013). Conservation tracking. In IEEE International Conference on Computer Vision (ICCV 2013), pages 2928–2935, Sydney, Australia. 24

Sherrard, K. and Robin, F., Lemaire, P., and Munro, E. (2010). Sequential activation of apical and basolateral contractility drives ascidian endoderm invagination. Curr Biol, 20(17):1499–1510. 10, 19, 37

Shi, W. and Levine, M. (2008). Ephrin signaling establishes asymmetric cell fates in an endomesoderm lineage of the ciona embryo. Development, 135(5):931–940. 36

Small, K., Brudno, M., Hill, M., and Sidow, A. (2007). A haplome alignment and reference sequence of the highly polymorphic ciona savignyi genome. Genome Biology, 8(3):R41. 34

Sommer, C., Straehle, C., Kothe, U., and Hamprecht, F. (2011). Ilastik: Interactive learning and segmentation toolkit. In Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, pages 230–233. 23

Stegmaier, J., Amat, F., Lemon, W. C., McDole, K., Wan, Y., Teodoro, G., Mikut, R., and Keller, P. J. (2016). Real-time three-dimensional cell segmentation in large-scale microscopy data of developing embryos. Developmental Cell, 36(2):225–240. 26

Stelzer, E. H. (2013). Light sheet based fluorescence microscopy. In Conference: Reducing Photobleaching and Phototoxicity in three-dimensional Imaging. 13

Stent, G. S. (1971). Molecular genetics. An introductory narrative. W.H. Freeman and Company. 3

Stolfi, A., Lowe, E. K., Racioppi, C., Ristoratore, F., Brown, C. T., Swalla, B. J., and Christiaen, L. (2014). Divergent mechanisms regulate conserved cardiopharyngeal development and gene expression in distantly related ascidians. Elife, 3:e03728. 34

Sulston, J. E. and Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, caenorhabditis elegans. Dev Biol, 56(1):110–156. 7

Tassy, O., Daian, F., Hudson, C., Bertrand, V., and Lemaire, P. (2006). A quantitative approach to the study of cell shapes and interactions during early chordate embryogenesis. Curr Biol, 16(4):345–358. 19, 32, 35, 37

Veeman, M. T. and Smith, W. C. (2013). Whole-organ cell shape analysis reveals the developmental basis of ascidian notochord taper. Developmental Biology, 373(2):281 – 289. 38

Vogt, N. (2014). Microscopy: Light microscopy with lattices. Nat Meth, 11(12):1191–1191. 11

Vu, N. B. S. (2008). Image Segmentation with Semantic Priors: A Graph Cut Approach. PhD thesis, University of California, Santa Barbara. 20

Wagner, E. and Levine, M. (2012). Fgf signaling establishes the anterior border of the ciona neural tube. Development, 139(13):2351–2359. 36

Weber, M., Mickoleit, M., and Huisken, J. (2014). Light sheet microscopy. Methods Cell Biol, 123:193–215. 16, 17

Witte, S., Negrean, A., Lodder, J. C., de Kock, C. P. J., Testa Silva, G., Mansvelder, H. D., and Louise Groot, M. (2011). Label-free live brain imaging and targeted patching with third-harmonic generation microscopy. Proceedings of the National Academy of Sciences, 108(15):5970–5975. 11

Wolpert, L. and Tickle, C. (2011). Principles of Development. OUP Oxford. 3, 10

Xiong, F., Ma, W., Hiscock, T. W., Mosaliganti, K. R., Tentner, A. R., Brakke, K. A., Rannou, N., Gelas, A., Souhait, L., Swinburne, I. A., Obholzer, N. D., and Megason, S. G. (2014). Interplay of cell shape and division orientation promotes robust morphogenesis of developing epithelia. Cell, 159(2):415–427. 8, 10

Xiong, F., Tentner, A. R., Huang, P., Gelas, A., Mosaliganti, K. R., Souhait, L., Rannou, N., Swinburne, I. A., Obholzer, N. D., Cowgill, P. D., Schier, A. F., and Megason, S. G. (2013). Specified neural progenitors sort to form sharp domains after noisy shh signaling. Cell, 153(3):550–561. 10

Yanai, I. and Hunter, C. P. (2009). Comparison of diverse developmental transcriptomes reveals that coexpression of gene neighbors is not evolutionarily conserved. Genome Research, 19(12):2214–2220. 34

Yang, Z., Mei, L., Xia, F., Luo, Q., Fu, L., and Gong, H. (2015). Dual-slit confocal light sheet microscopy for in vivo whole-brain imaging of zebrafish. Biomed. Opt. Express, 6(5):1797–1811. 17, 18

Yasuo, H. and Hudson, C. (2007). Fgf8/17/18 functions together with fgf9/16/20 during formation of the notochord in ciona embryos. Dev Biol, 302(1):92–103. 36

Chapter 2

ASTEC: Adaptive Segmentation and Tracking of Embryonic Cells

Contents

2.1 Introduction

This section is the first version of the article *Systematic high-throughput reconstruction and tracking of animal embryonic cells over hundreds of time points.*

Authors and contribution I developed the segmentation and tracking pipeline (ASTEC), did the lineage tree analysis, developed and analyse the induction model. U.-M. Fiuza did all the embryo preparation and image acquisition and the manual curation of the segmentations. L. Hufnagel helped with the image acquisition protocol and provided the MuVi-SPIM. G. Malandain supervised the image analysis part of the work. P. Lemaire and C. Godin wrote the manuscript and supervised the work. All authors contributed to the manuscript.

2.2 Systematic high-throughput digitalization and tracking of live ascidian embryonic cells highlights the importance of the precision of cell-cell contacts areas for cell inductions.

Guignard L.[1,2⋆], Fiuza U.-M[1,⋆], Hufnagel L.[3], Malandain G.[4], Godin C.[2,#], Lemaire P.[1,#]

[1] CRBM, UMR5237, CNRS-U. Montpellier, 1919 Route de Mende F-34293 MONTPELLIER Cedex 5, France
[2] Inria project-team Virtual Plants, joint with CIRAD and INRA, Campus St Priest - BAT 5, CC 05018, 860 rue de St Priest, F-34095 Montpellier Cedex 5, France
[3] EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany
[4] Inria project-team Morpheme, 2000 route des lucioles, les algorithmes - bat. Euclide B - CS 40121 Sophia Antipolis cedex F-06903, France
⋆ equal contribution, # corresponding authors
Contact: lemaire.patrick@crbm.cnrs.fr, christophe.godin@inria.fr

2.2.1 Abstract

We combined light-sheet imaging with computational analysis to achieve a quantitative digital representation of the stereotyped embryogenesis in the ascidian *Phallusia mammillata*, between the cleavage and initial tailbud stages. This dataset gives access to the position, shape, divisions and contacts of 1304 cells with a 2-minutes temporal resolution and across 671 cell divisions. Through the comparison of the cell lineage trees of sister cells, we show that the mitotic history of cells is diagnostic of their cell fate and present a map of cell specification events until the end of gastrulation. To understand the molecular basis of these decisions, we integrated measures of cell volumes, cell-cell contact areas and boolean spatio-temporal expression data for extracellular signalling molecules. Computational simulation reveals that remarkably simple cell induction rules, based on the precision of the measure of cell-cell contacts rather than on the concentration of extracellular ligand explain most cell specification events up to the late gastrula stage. We thus propose the existence of a trade off between constraints on embryo geometry and on quantitative signalling gene expression. This scenario may explain why organisms with an embryogenesis relying on invariant cell lineages combine long-term anatomical evolutionary conservation and rapid genomic divergence.

2.2.2 Introduction

Classical developmental biology approaches have so far led to a mostly qualitative understanding of the regulatory cascades and networks that drive fate specification and morphogenesis during embryogenesis. A major challenge in this field is to extend this coarse understanding to a description that quantitatively links the dynamics of cell behaviour to fate specification and patterns of gene activity.

In nematode or ascidian embryos, the quasi-invariant cell lineages and cell cleavage patterns observed during development should in theory allow tracking the divisions, shape changes and migrations of each cell during development. Indeed, the dynamics of *C. elegans* cell behaviours have been successfully described by imaging individual cell nuclei in live embryos, and computationally extracting from the resulting image stacks the position of each nucleus and the structure of cell lineages. This approach allowed the systematic statistical quantification of both wild-type and mutant phenotypes [Moore et al., 2013] and the systems-level inference of cell fate decision mechanisms [Du et al., 2014]. Nuclei-based cell tracking, however, does not give access to the geometry of individual cells, which can only be obtained by imaging and segmenting plasma membrane-labelled cells. Existing automated [Fernandez et al., 2010; Mikula et al., 2011; Mosaliganti et al., 2012] or semi-automated [Khan et al., 2014; Sommer et al., 2011] solutions are efficient for small datasets but inadequate to reconstruct the many thousand cell "snapshots" generated by the most promising imaging technology in developmental biology, time-lapse light-sheet microscopy [Krzic et al., 2012; Keller, 2013].

2.2.3 Results and discussion

Using multiview lightsheet microscopy [Krzic et al., 2012], we imaged every two minutes and for 6 hours a transparent *Phallusia mammillata* embryo from 4 angles of view, without compromising the development of the embryo (Figure 2.1A-C, Supp. Figure 2.12, Supp. text 2.3.2). The resulting movie extends from the 64-cell stage to the initial tailbud stage [Hotta et al., 2007] and covers two major morphogenetic processes, gastrulation and neurulation. The high acquisition speed (34 frames per seconds) ensured that, for a given time point, embryonic cell geometries do not change between image acquisition along different angles of views (Supp. Figure 2.12-2.11). Images from consecutive time points were similar enough to be efficiently registered, using a non-linear registration algorithm (Figure 2.1D).

To automatically extract the shape of each cell by segmentation and to track cells during their lifetime and across cell divisions, we first attempted to apply our previous MARS-ALT pipeline [Fernandez et al., 2010]. MARS-ALT proceeds in two passes. First, 3D image stacks at each time point are independently segmented by detecting a seed in each cell, which is grown until it reaches the cell boundaries, using a 3D watershed algorithm (MARS). Then cell lineages are tracked between pairs of consecutive segmented images by deforming the image at time $t + 1$ so that it best matches the image at time t and finding an optimal association between cells from time t and $t + 1$ in this common reference frame (ALT). MARS was at best able to detect all the cells with 4% of oversegmented cells at time $t = 152$ minutes when the embryo counts 218 cells (see Supp. Figure 2.13). However, this algorithm turned out to be insufficient to reconstruct faithfully lineage sequences over the 180 time points of the developmental sequence, as the few segmentation errors made at each time point resulted in a high frequency of cell lineage errors (Supp. Figure 2.13-2.14).

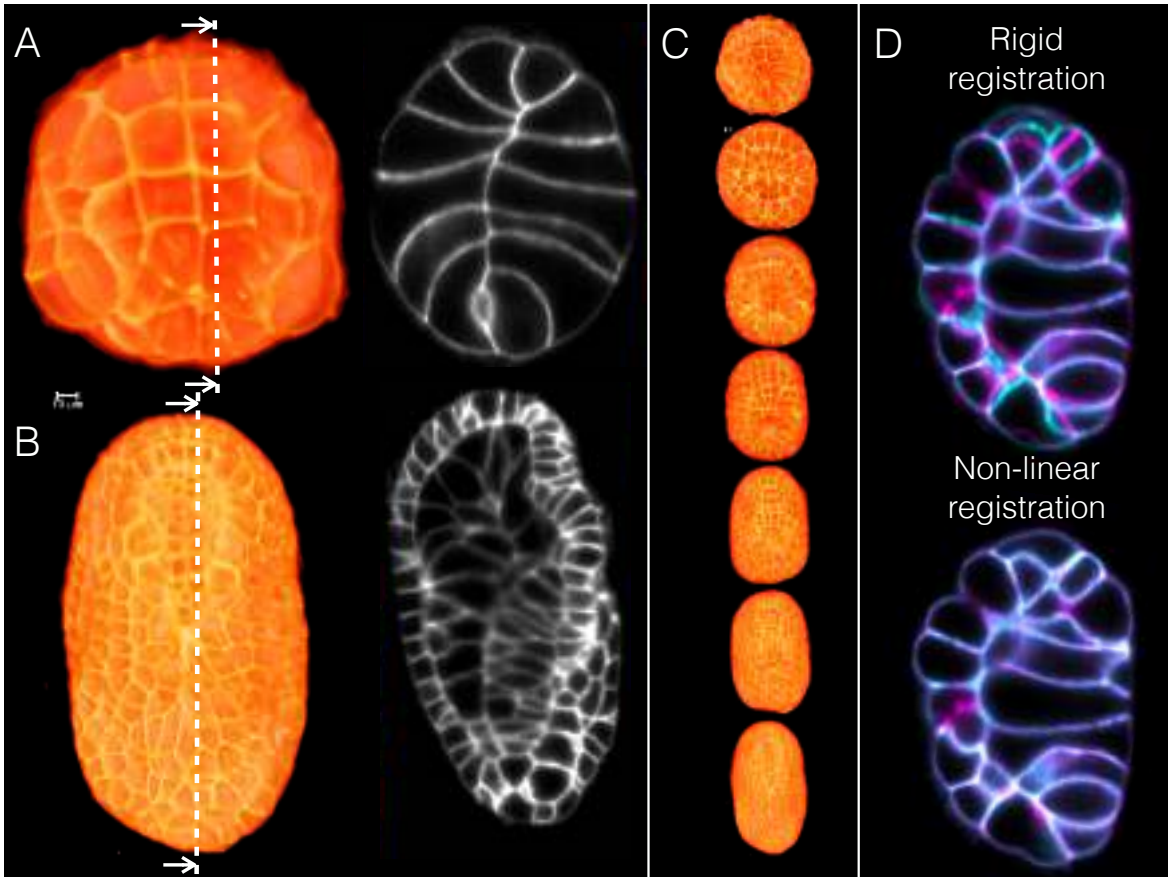


Figure 2.1 – ***Result of imgaging protocol on a developmental sequence of Phallusia mammillata from 64-cell to early tailbud stage.*** *A-C) Overview of image quality. A) First imaged time point, at the 64-cell stage. Left, vegetal view of texture-based volume rendering of the image stack, anterior is to the top. Right optical parasagittal section as indicated by hashed line on left image (animal hemisphere to the left, anterior to the top). B) Final time point at the initial tailbud stage. Left, dorsal view of texture-based volume rendering of the image stack, anterior is to the top. Right, optical sagittal section as indicated by hashed line (dorsal side to the right). C) Excerpt of the complete imaged time-series: volume renderings of 3D image stacks separated by 60 minutes of development. The views show the evolution of the vegetal side of the embryo from the 64-cell stage. Anterior is to the top. D) Example of registration of successive timepoints (40 min, cyan and 42 min, magenta) illustrated at the moment of several cell divisions. Top: comparison of the geometry of the two time points without deformation. Bottom: non-linear registration of timepoint 40 onto timepoint 42.*

To overcome these limitations, we developed a single pass algorithm, ASTEC, for Adaptive Segmentation and Tracking of Embryonic Cells. This iterative algorithm propagates segmentations from one time point to the next, thereby simultaneously segmenting and tracking the lineage of membrane-labelled cells (Figure 2.2A,B, and Supp. text 2.3.3). To initiate the procedure, the MARS algorithm is applied at the first time point,

when the embryo only contains a limited number of cells, followed by a manual curation if necessary. Each iteration then uses the segmentation at time t as a guide to segment the embryo at time $t+1$. First, the 3D image at t is deformed to best match the image at $t+1$. The deformed projection of each segmented cell at time t, defines the region occupied by its progeny at t. In a second step, based on a local analysis of noise within each progeny region, ASTEC detects whether the cell has divided between t and $t+1$, and accordingly places either one or two seeds in each progeny region. Third, all seeds are used to initiate a 3D watershed on the image at $t+1$.

Upon completion of all iterations, two outputs are produced: a segmentation of all embryonic cells present at each time point, and for each cell the identity of its progeny at the next time point, from which global cell lineages can be reconstructed. Analysis of these lineages reveals the persistence of a substantial number of oversegmentation errors, which can be traced back up to the cell where the error occurred (See Figure 2.2B, Supp. Fig. 2.15, Supp. text 2.3.3). A final post-processing step is thus applied to automatically correct these errors, based on the topology of the trees and the volumes of sister cells (Figure 2.2B, Supp. Fig. 2.15). The resulting global lineage tree (Supp. Figure 2.17) contains a total of 58454 digital 3D cell "snapshots", describing the behaviour in time of 1304 individual cells generated by 671 cell division events. This dataset provides a quantitative digital representation of a whole developmental program, which we formalized as a 4D dynamic graph.

To assess the quality of the output of ASTEC, we first manually expertised cell at t=152 minutes (Figure 2.3A), when the embryo counts 218 cells, and at the end of the movie at t=360 minutes when the embryo counts 702 segmented cells. 218/218 (100%) and 702/709 (99%) cells were accurately detected, respectively (Figure 2.3B). Manual expertise of the 3D shape of each individual cell at these two time points and of 2D sections through 8953 cells evenly distributed across the whole sequence (see Supp. Text 2.3.4, Supp Fig. 2.19) showed that more than 99% of voxels were assigned to the right cell (see Figure 2.3, Supp. Fig. 2.19). ASTEC is thus able to detect and segment cells with high accuracy.

To analyze the quality of the cell lineages, we first defined a metric between lineage trees, which was used to compute pairwise distances between trees (See supp. Text 2.3.5). Figure 2.3C shows the widespread dispersion of these pairwise distances between precursors, indicating the presence of both very similar and very different lineage tree structures within the embryo. Consistent with the expected bilateral symmetry of the embryo, pairs of bilaterally symmetric lineages had very small distances (mean difference: 0.061; Figure 2.3C). The volumes of bilateral cells were also highly similar (Supp Figure 2.20B). The same metric was used to compare our lineage to an independent, manually-curated, *Phallusia mammillata* cell lineage covering a similar developmental period from an embryo with fluorescently-labeled nuclei (Faure et al., 2015, personal communication) (Supp. figure 2.21). Pairwise comparison of tree lineages deriving from matching cells in both embryos revealed a high similarity (mean difference: 0.132, Figure

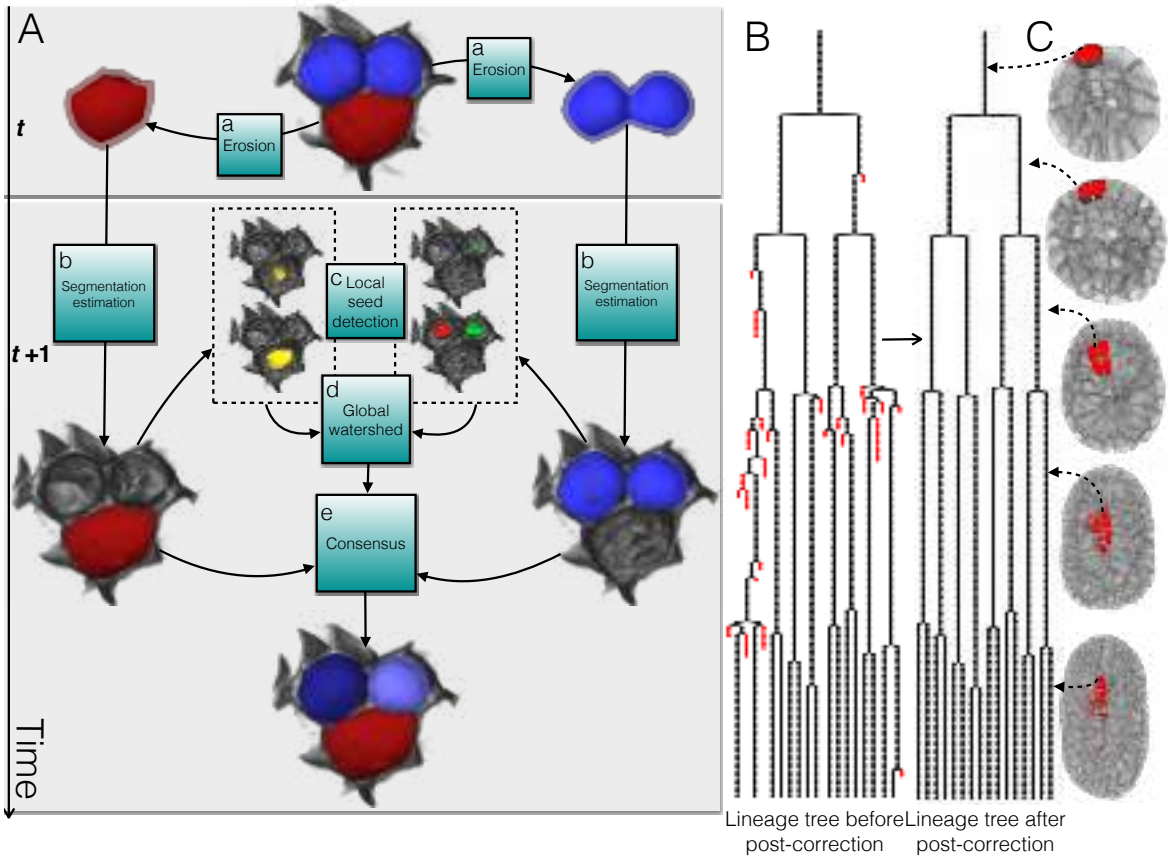


Figure 2.2 – ***The ASTEC pipeline.*** *The segmentation and post-correction algorithm. A) The successive stages of the propagation of segmentation between consecutive time-points. The figure illustrates both the case of a non-dividing cell (red) and of a dividing cell (blue). Each cell is first eroded (a). The deformed erosion is then used to build the projection of the segmentation at time $t+1$ (b). Local seeds are extracted from the projected segmented regions (c). The segmentation estimation is computed using a 3D watershed algorithm based on the intensity image at time $t+1$ and the extracted seeds (d). The final segmentation results from a consensus between the projected and estimated segmentations (e). B) The ASTEC iterative procedure produces a complete cell lineage tree (left tree) that contains remaining errors (red dots). A post-processing correction treatment makes it possible to correct these residual errors (right tree). Illustration of the tracking of the position of the descendency of cell A7.4 at the timepoints indicated by arrows (the cell lineage shown does not correspond to A7.4, see Supp. Fig. 2.18.*

2.20D), except in isolated cases, which could often be traced to issues in the published lineage (Supp. Figure 2.21, Supp. Fig. 2.22). Analysis of the pattern of rounding up of cells around mitosis revealed that the temporal accuracy of detected cell divisions was within 2 minutes of the actual division time (Supp. Figure 2.20C-2.18). ASTEC thus reconstructs cell lineages with a high accuracy.

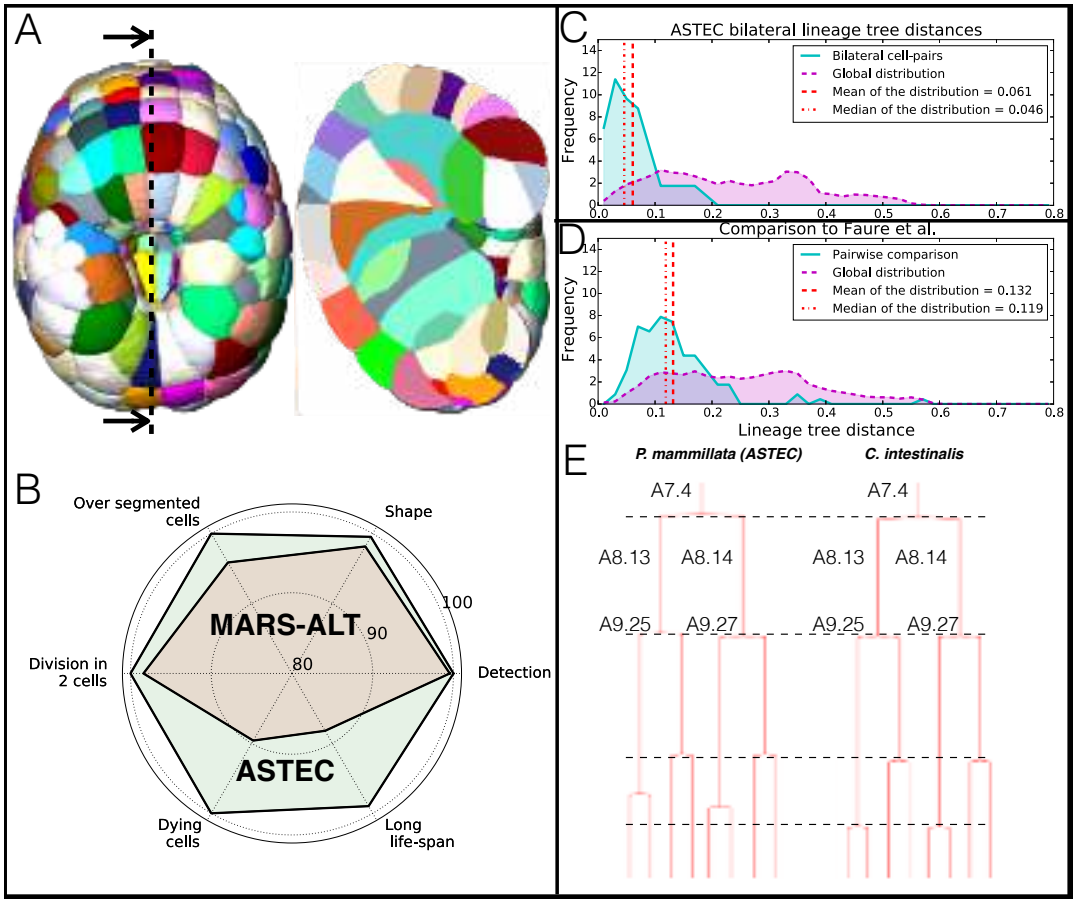


Figure 2.3 – ***Validation of ASTEC output.*** *A) Images of ASTEC segmentation during the gastrula stage (t=152). Left surface view of vegetal side of embryo. Right sagittal view along the hashed lines. The vegetal side is to the right. Anterior is to the top. Colours are randomly assigned. B) Spider graphs showing a comparison of a two pass algorithm (MARS-ALT, pink) and our one pass algorithm (ASTEC, light green). Detection: percentage of accurately detected cells at time 152. Shape: percentage of well allocated voxels in accurately detected cells at time 152. Division in 2 cells: percentage of cells giving rise to at most 2 daughters. Dying cells percentage of cells non dying cells. Correct lifetime: percentage of cells with a life span superior to 30 minutes (up to 30 minutes before end). C) Distribution of the distances between cell lineage trees. Magenta: pairwise comparison between all possible trees originating in the early gastrula stage. Cyan: pairwise comparison restricted to the bilateral cell-pairs of the the early gastrula stage. D) Comparison of cell lineage trees from ASTEC and Faure et al. (personal communication) Magenta: pairwise distribution between all possible trees originating in the early gastrula stage from ASTEC against all possible trees originating in the early gastrula stage from Faure et al. Cyan: pairwise distances between trees originating from the same cell at the same stage in both pipelines. E) The Phallusia (ASTEC), and Ciona ([Nicol and Meinertzhagen, 1991]) neural plate cell lineages are very similar.*

The high similarity of cell lineages in two independent embryos, and between bilateral cells in each embryo, suggests that *Phallusia mammillata* development proceeds with a stereotyped cell lineage until at least the initial tailbud stage. Ascidians are an ancient animal group, which probably emerged during the Cambrian. Comparison of the *Phallusia* (phlebobranchia) lineage with the partial cell lineages determined in *Ciona intestinalis*, another phlebobranchian, and in *Halocynthia roretzi*, a very distantly related stolidobranchian, revealed a high level of evolutionary conservation, most changes corresponding to slight heterochronic shifts in the timing of cell divisions (Figure 2.3E and not shown). We conclude that the stereotypy of the ascidian cell lineages extends to the gastrula and neurula periods, and that these lineages are subjected to very high evolutionary constraints.

Cell fate specification often leads to changes in the pattern of cell divisions [Sulston et al., 1983], thereby affecting the topology of cell lineage trees. Analysis of the different modes of the distribution of pairwise cell lineage tree distances from the 64-cell stage indicated indeed that cell lineage trees within a given tissue (mean distance 0.085) were more similar than cell lineage trees from precursors of distinct tissues (mean distance 0.315) (Figure 2.3A, Supp Figure 2.23). Clustering of lineage trees based on this metric confirmed that the mitotic history of cells was generally diagnostic of the fate of the cells considered (Figure 2.4B).

In ascidian embryos, the majority of early blastomeres become fate restricted to a single embryonic tissue type by the early gastrula stage [Nishida, 1987]. Some of these tissues, however, are subsequently patterned to give rise to several larval or juvenile mesodermal tissues in the case of the Trunk Ventral Cells (TVC), mesenchyme and Trunk Lateral Cells (TLC) [Hirano and Nishida, 1997, 2000; Tokuoka et al., 2005], or to regionalize the complex larval central nervous system [Nicol and Meinertzhagen, 1991; Cole and Meinertzhagen, 2004] or the tail epidermis [Pasini et al., 2006]. To identify the cascade of cell fate specification events occuring during the gastrula stages, we reasoned that if the cell lineages originating from two sister cells significantly differ in their topology or timings of cell division, these cells may have distinct fates. Figure 2.4C shows the distribution of all cell lineage tree distances between 81 pairs of sister cells generated between the 64-cell and mid gastrula stages. 19/23 known fate specification events led to sister cells with cell lineage distances larger than 0.12. By contrast, only 7/57 cell divisions not known to give rise to differentially fated daughters had large cell lineage distances. Two of these candidate cell specification events were found in the TLC lineage, one in the mesenchyme and three in the tail epidermis lineages (Supp Fig. 2.24). Figure 2.4E illustrates the cascade of cell specification events in the Trunk Lateral Cell lineage, which gives rise to juvenile blood, oral siphon and body muscle [Hirano and Nishida, 1997]. These results establish that cell lineage comparisons efficiently identify cell fate specification events and reveal that only two mesodermal tissues that will have a major contribution to the adult tissues are regionalized during the gastrula stages.

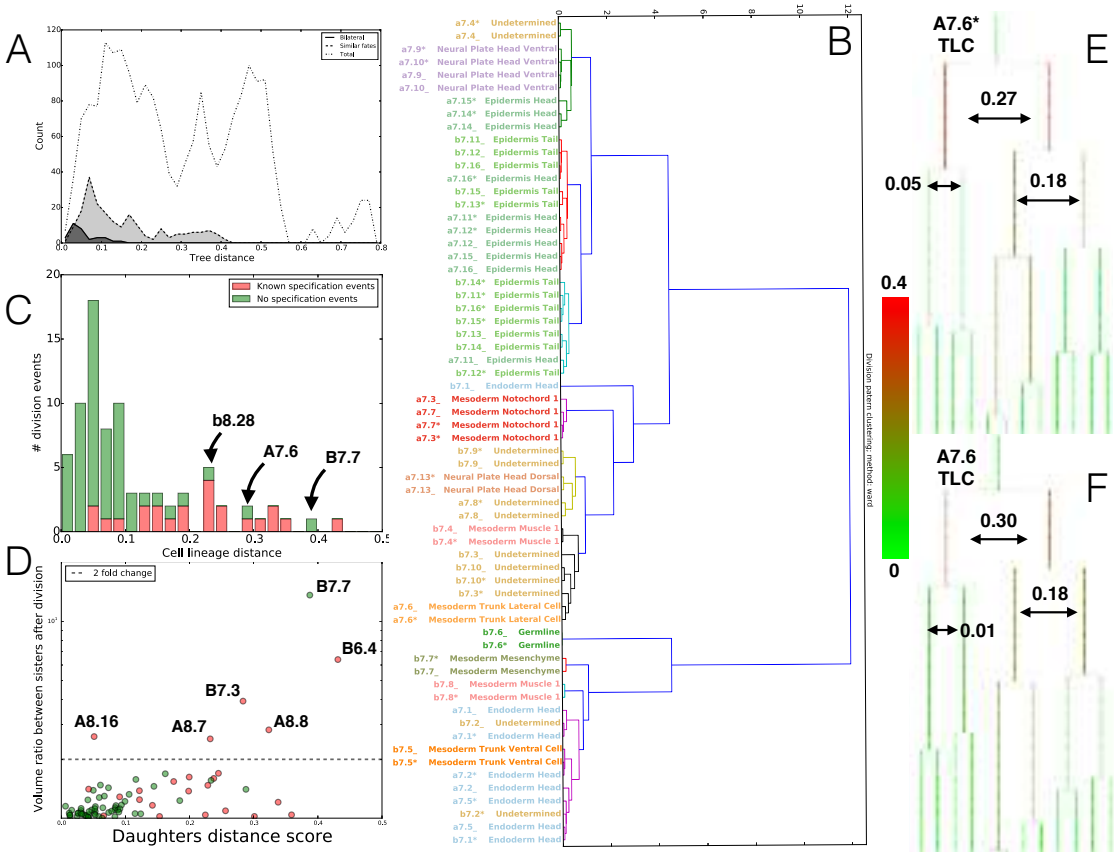


Figure 2.4 – ***Cell lineage tree analysis.*** *A) Distribution of the distances between cell lineage trees. Hashed line: pairwise comparison between all the 64 possible trees originating in the 64-cell stage. Dark grey: pairwise comparison between bilateral cell-pairs at the 64-cell stage. Light grey: pairwise comparison between cell-pairs of similar fates at the 64-cell stage. B) Tree cluster resulting from the hierarchical clustering of the cell lineage trees at the 64-cell stage. C) Distribution of the distances between sisters cell lineage trees. Green: cell divisions thought to occur without cell specification of the daughters. Red: cell divisions known to occur with cell specification events. Blue: cell divisions that might be accompanied with a cell specification event (not proved yet). Vertical hashed line: 0.12 threshold value D) Scatter plot of the distance between sister cell lineage trees and their volume ratio (color code identical to C). Vertical hashed line similar to C), horizontal hashed line: 2 fold ratio threshold. E-F) Example of cascade of novel specification events suggested by ASTEC in the Trunk Lateral Cell lineage. Colour for the nodes: the distance score between the sister cell lineage trees from 0 (green) to 0.4 (red).*

First identified in ascidians [Conklin, 1905], unequal cell cleavages producing daughter cells of different sizes are frequently associated to cell fate specification events [Weisblat, 2007; Knoblich, 2010]. Comparison of the cell volumes of sister cells showed that the cleavage inequality of 6 cells significantly departs from the rest of the distribution up to the late gastrula stage, with the most striking unequality being found in the B7.7

mesenchymal cell lineage (Figure 2.4F). These strong cleavage unequalities were tightly associated to candidate cell fate specification events

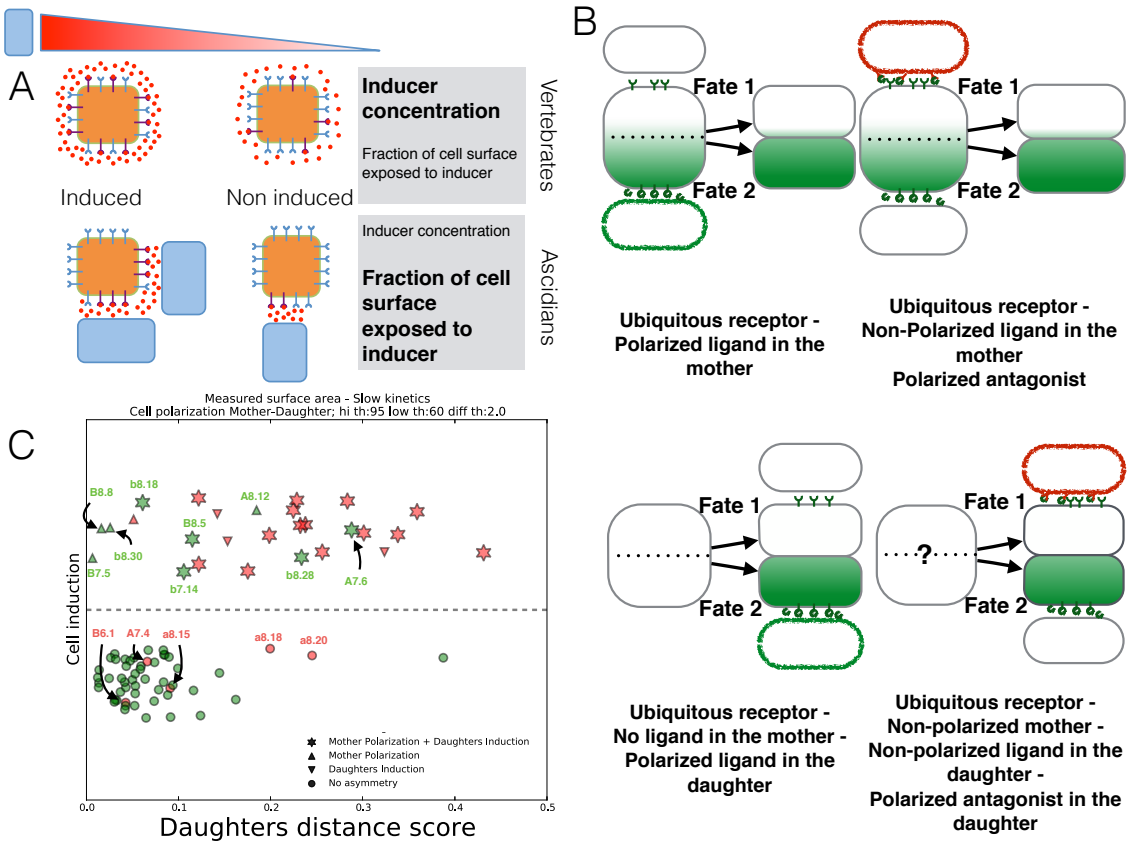


Figure 2.5 – ***Induction modelling.*** *A) Schematic comparison of the induction processes between vertebrates and ascidians B) Representation of the possible juxtacrine inductions. Top: polarization of the cortex of the mother cell. Bottom: Induction of the daughter cells. C) Scatterplot of the outputs of the induction model. Each dot represents a bilateral cell-pair and its associated cell division event. In red, cell-pairs with known differentiation events. In green, cell-pairs unlikely to undergo differentiation events.*

Cell inductions play a dominant role in animal cell fate specification events [Lemaire, 2009], but may be controlled in different ways depending on embryo geometry. In early vertebrate embryos, early embryonic inductions occur within large fields of similar cells and are often controlled by the concentration of diffusible inducer surrounding individual cells [Gurdon et al., 1999]. The situation appears different during early ascidian embryogenesis. For example, the bipotential a5.3 blastomere gives rise to the a6.6 daughter, fated to head epidermis, and to the a6.5 daughter, fated to anterior neural tissue. During this fate decision, cells secreting the FGF9/16/20 neural inducer establish a larger area of contact with a6.5 than with a6.6, and this acts as a strong determinant of the outcome of the induction [Tassy et al., 2006]. During primary notochord specification,

local juxtacrine FGF signalling polarizes a bipotential mother cell, and the daughter inheriting the region of the mother cell that was exposed to FGF signalling adopts the notochord fate [Minokawa et al., 2001]. A similar process is at work during early mesenchyme induction [Kim et al., 2007]. These examples suggest that ascidian embryonic inducers act in a juxtacrine manner and that the success of an induction depends more on the area of cell contacts established between emitting and responding cells, than on the precise global concentration of inducer each cell is exposed to, as in vertebrate embryos (Figure 2.5A).

To test this idea, we built a simple computational model (see Methods 2.3.6). Based on the precise measurements of the contact areas established by each embryonic cell with cells expressing signalling ligands/antagonists, the model looks for pathways that either polarize a bipotential mother cell, or differentially induce each of its daughters. This approach is made possible by the existence of an atlas of signalling gene expression with cellular resolution, which indicates that only five major signalling pathways (FGF/ephrin, Wnt, Bmp, Nodal and Notch) show differential expression of extracellular ligands or antagonists during the cleavage and early gastrula stages (Supp. Table 2.2) [Imai et al., 2004, 2006]. The model considers four different situations (Figure 2.5B) and has three free parameters, set at the same values for all signalling pathways, and which specify: 1) the ratio of signalling intensities necessary to either polarize a mother or differentially induce daughters; 2) A lower threshold of signalling necessary to obtain an induction 3) An upper threshold of signalling intensity received by both sides of a mother or by both daughters above which no polarization or differential induction can occur (see Methods for the precise definition and calculation of these parameters).

We then explored parameter space to identify combinations of parameters that generated predictions most consistent with a set of 20 known inductive events involving 14 cells and 9 cell specification events for which the molecular driving mechanism is unknown (Supp. Tables 2.2-2.3). These simple rules identified a set of parameters (see Methods) that correctly predicted inducers for 18 of the 23 training cell fate specification events (Figure 2.5C, red) and identified the expected ligand in 19/20 cases of known inductions. Furthermore, the model predicted a potential induction in only 6 out of 57 cells in which no fate specification event was expected. Supp. Figure 2.25 shows the sensitivity of the results to parameter variations. Rerunning the parameter optimization after introducing various levels of noise in the surfaces of contact between cells, however, systematically gave poorer results than when the measured surfaces were used (Supp. Figure 2.26), highlighting that the specific stereotyped geometrical organization of ascidian blastomere is optimized for inductive processes to occur. Taken together, these results indicate that, the topology of the early embryo and of expressing cells is sufficient to explain most cell fate specification events taking place up to the mid-gastrula stage, even under the very strong hypothesis that the amount of signalling ligands/antagonists secreted by each expressing cell is not limiting.

Ascidians present a fascinating paradox. Species as distantly related as the phlebobranchian *Ciona intestinalis* and the stolidobranchian *Halocynthia roretzi*, which have diverged several hundred million years ago, have kept remarkably similar embryological morphologies and nearly identical early cell lineages [Lemaire et al., 2008]. Yet, at the molecular level, ascidians are fast evolvers both between and within species [Tsagkogeorga et al., 2009, 2010, 2012], and even some of their core regulatory logics may differ during embryogenesis [Stolfi et al., 2014]. Our analysis of cell fate specification and inductive processes may reconcile these two apparently antagonistic properties. In vertebrate embryos, cell inductions pattern large fields of equally competent cells. Under these conditions, the whole surface of competent cells is exposed to the inducer, whose level of activity needs to be precisely set [Gurdon et al., 1999]. This may impose local constraints on the evolution of the coding and non-coding sequences of inducer genes. In ascidian embryos, inductions act in a juxtacrine manner to either polarize a mother cell or differentially induce two equally competent daughter cells. Our computational simulations show that the portion of the surface of induced cells exposed to the inducer is a major determinant of the outcome of inductions, and that inductions can be accurately predicted without need to take into consideration the precise concentration of ligand emitted by embryonic cells. The evolutionary constraints on the genes coding for these factors may thus have been relaxed. We thus propose that, in ascidians, the stereotypy of embryogenesis may in part explain the accelerated molecular evolution. Interestingly, nematodes, which also develop with invariant cell lineages and short range cell inductions, also show accelerated molecular evolution [Stein et al., 2003]. The functional correlation between stereotyped embryogenesis and accelerated molecular evolution we propose may thus extend beyond ascidians.

2.2.4 Acknowledgements

This work was funded by core support from CNRS to PL, by Inria (core support and IPL morphogenetics) to CG and GM, by the Geneshape project (ANR-SYSC-018-02) to PL and CG and by the Dig-Em project (ANR-14-CE11-0013-01) to PL, CG and GM. CG and PL are members of the Institut de Biologie Computationelle of Montpellier (IBC). Work in LH's lab was funded by EMBL. LG was supported by a doctoral contract from the CBS2 doctoral school of the University of Montpellier, by the Fondation pour la Recherche Médicale (FRM) (FDT20140931061), and by the Morphoscope2 Equipex project. UMF was supported by the Geneshape project, and by the FRM (SPF20120523969). We thank the IT support team of CRBM, and C. Dantec for their help, and the members of the Lemaire and Godin groups for their comments and advise throughout this project.

References

Cole, A. G. and Meinertzhagen, I. A. (2004). The central nervous system of the ascidian larva: mitotic history of cells forming the neural tube in late embryonic ciona intestinalis. Developmental Biology, 271(2):239 – 262. 58

Conklin, E. G. (1905.). The organization and cell-lineage of the ascidian egg / by Edwin G. Conklin. Philadelphia :[Academy of Natural Sciences],. 59

Du, Z., Santella, A., He, F., Tiongson, M., and Bao, Z. (2014). De novo inference of systems-level mechanistic models of development from live-imaging-based phenotype analysis. Cell, 156(1–2):359 – 372. 53

Fernandez, R., Das, P., Mirabet, V., Moscardi, E., Traas, J., Verdeil, J.-L., Malandain, G., and Godin, C. (2010). Imaging plant growth in 4d: robust tissue reconstruction and lineaging at cell resolution. Nat Meth, 7(7):547–553. 53, 68, 69

Guignard, L., Godin, C., Fiuza, U.-M., Hufnagel, L., Lemaire, P., and Malandain, G. (2014). Spatio-temporal registration of embryo images. In ISBI - International Symposium on Biomedical Imaging, Pekin, Chine. IEEE. 67, 70

Gurdon, J. B., Standley, H., Dyson, S., Butler, K., Langon, T., Ryan, K., Stennard, F., Shimizu, K., and Zorn, A. (1999). Single cells can sense their position in a morphogen gradient. Development, 126(23):5309–5317. 60, 62

Hirano, T. and Nishida, H. (1997). Developmental fates of larval tissues after metamorphosis in ascidian halocynthia roretzi. i. origin of mesodermal tissues of the juvenile. Dev Biol, 192(2):199–210. 58

Hirano, T. and Nishida, H. (2000). Developmental fates of larval tissues after metamorphosis in the ascidian, halocynthia roretzi. Development Genes and Evolution, 210(2):55–63. 58

Hotta, K., Mitsuhara, K., Takahashi, H., Inaba, K., Oka, K., Gojobori, T., and Ikeo, K. (2007). A web-based interactive developmental table for the ascidian ciona intestinalis, including 3d real-image embryo reconstructions: I. from fertilized egg to hatching larva. Dev Dyn, 236(7):1790–1805. 53

Hudson, C., Lotito, S., and Yasuo, H. (2007). Sequential and combinatorial inputs from nodal, delta2/notch and fgf/mek/erk signalling pathways establish a grid-like organisation of distinct cell identities in the ascidian neural plate. Development, 134(19):3527–3537. 81

Hudson, C. and Yasuo, H. (2005). Patterning across the ascidian neural plate by lateral nodal signalling sources. Development, 132(6):1199–1210. 81

Hudson, C. and Yasuo, H. (2006). A signalling relay involving nodal and delta ligands acts during secondary notochord induction in ciona embryos. Development, 133(15):2855–2864. 81

Imai, K. S., Hino, K., Yagi, K., Satoh, N., and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development, 131(16):4047–4058. 61

Imai, K. S., Levine, M., Satoh, N., and Satou, Y. (2006). Regulatory blueprint for a chordate embryo. Science, 312(5777):1183–1187. 61

Imai, K. S., Satoh, N., and Satou, Y. (2002). Early embryonic expression of fgf4/6/9 gene and its role in the induction of mesenchyme and notochord in ciona savignyi embryos. Development, 129(7):1729–1738. 81

Keller, P. (2013). Imaging Morphogenesis: Technological Advances and Biological Insights. Science, 340(6137):1234168+. 53

Khan, Z., Wang, Y.-C., Wieschaus, E. F., and Kaschube, M. (2014). Quantitative 4d analyses of epithelial folding during drosophila gastrulation. Development, 141(14):2895–2900. 53

Kim, G. J., Kumano, G., and Nishida, H. (2007). Cell fate polarization in ascidian mesenchyme/muscle precursors by directed fgf signaling and role for an additional ectodermal fgf antagonizing signal in notochord/nerve cord precursors. Development, 134(8):1509–1518. 61

Knoblich, J. A. (2010). Asymmetric cell division: recent developments and their implications for tumour biology. Nat Rev Mol Cell Biol, 11(12):849–860. 59

Krzic, U., Gunther, S., Saunders, T. E., Streichan, S. J., and Hufnagel, L. (2012). Multiview light-sheet microscope for rapid in toto imaging. Nat Meth, 9(7):730–733. 53, 67

Lemaire, P. (2009). Unfolding a chordate developmental program, one cell at a time: invariant cell lineages, short-range inductions and evolutionary plasticity in ascidians. Dev Biol, 332(1):48–60. 60

Lemaire, P., Smith, W. C., and Nishida, H. (2008). Ascidians and the plasticity of the chordate developmental program. Curr Biol, 18(14):R620–31. 62

Marquez-Neila, P., Baumela, L., and Alvarez, L. (2014). A morphological approach to curvature-based evolution of curves and surfaces. IEEE Trans Pattern Anal Mach Intell, 36(1):2–17. 73

Mikula, K., Peyrieras, N., Remesikova, M., and Stasova, O. (2011). Segmentation of 3d cell membrane images by pde methods and its applications. Comput Biol Med, 41(6):326–339. 53

Minokawa, T., Yagi, K., Makabe, K. W., and Nishida, H. (2001). Binary specification of nerve cord and notochord cell fates in ascidian embryos. Development, 128(11):2007–2017. 61

Moore, J. L., Du, Z., and Bao, Z. (2013). Systematic quantification of developmental phenotypes at single-cell resolution during embryogenesis. Development, 140(15):3266–3274. 53

Mosaliganti, K. R., Noche, R. R., Xiong, F., Swinburne, I. A., and Megason, S. G. (2012). *ACME: Automated Cell Morphology Extractor* for comprehensive reconstruction of cell membranes. PLoS Comput Biol, 8(12):e1002780. 53

Nicol, D. and Meinertzhagen, I. A. (1991). Cell counts and maps in the larval central nervous system of the ascidian ciona intestinalis (l.). J Comp Neurol, 309(4):415–429. 57, 58

Nishida, H. (1987). Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. Developmental Biology, 121(2):526 – 541. 58

Ohta, N. and Satou, Y. (2013). Multiple signaling pathways coordinate to induce a threshold response in a chordate embryo. PLoS Genet, 9(10):e1003818. 81

Pascal Ferraro and Christophe Godin (2000). A distance measure between plant architectures. Ann. For. Sci., 57(5):445–461. 75

Pasini, A., Amiel, A., Rothbacher, U., Roure, A., Lemaire, P., and Darras, S. (2006). Formation of the ascidian epidermal sensory neurons: insights into the origin of the chordate peripheral nervous system. PLoS Biol, 4(7):e225. 58

Picco, V., Hudson, C., and Yasuo, H. (2007). Ephrin-eph signalling drives the asymmetric division of notochord/neural precursors in ciona embryos. Development, 134(8):1491–1497. 81

Shi, W. and Levine, M. (2008). Ephrin signaling establishes asymmetric cell fates in an endomesoderm lineage of the ciona embryo. Development, 135(5):931–940. 81

Sommer, C., Straehle, C., Kothe, U., and Hamprecht, F. (2011). Ilastik: Interactive learning and segmentation toolkit. In Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, pages 230–233. 53

Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D. H. A., Fulton, L. A., Fulton, R. E., Griffiths-Jones, S., Harris, T. W., Hillier, L. W., Kamath, R., Kuwabara, P. E., Mardis, E. R., Marra, M. A., Miner, T. L., Minx, P., Mullikin, J. C., Plumb, R. W., Rogers, J., Schein, J. E., Sohrmann, M., Spieth, J., Stajich, J. E., Wei, C., Willey, D., Wilson, R. K., Durbin, R., and Waterston, R. H. (2003). The genome sequence of *Caenorhabditis briggsae* a platform for comparative genomics. PLoS Biol, 1(2):e45. 62

Stolfi, A., Lowe, E. K., Racioppi, C., Ristoratore, F., Brown, C. T., Swalla, B. J., and Christiaen, L. (2014). Divergent mechanisms regulate conserved cardiopharyngeal development and gene expression in distantly related ascidians. Elife, 3:e03728. 62

Sulston, J. E., Schierenberg, E., White, J. G., and Thomson, J. N. (1983). The embryonic cell lineage of the nematode caenorhabditis elegans. Dev Biol, 100(1):64–119. 58

Tassy, O., Daian, F., Hudson, C., Bertrand, V., and Lemaire, P. (2006). A quantitative approach to the study of cell shapes and interactions during early chordate embryogenesis. Curr Biol, 16(4):345–358. 60

Tokuoka, M., Satoh, N., and Satou, Y. (2005). A bhlh transcription factor gene, twist-like 1, is essential for the formation of mesodermal tissues of ciona juveniles. Dev Biol, 288(2):387–396. 58

Tsagkogeorga, G., Cahais, V., and Galtier, N. (2012). The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate ciona intestinalis. Genome Biol Evol, 4(8):740–749. 62

Tsagkogeorga, G., Turon, X., Galtier, N., Douzery, E. J. P., and Delsuc, F. (2010). Accelerated evolutionary rate of housekeeping genes in tunicates. J Mol Evol, 71(2):153–167. 62

Tsagkogeorga, G., Turon, X., Hopcroft, R., Tilak, M.-K., Feldstein, T., Shenkar, N., Loya, Y., Huchon, D., Douzery, E., and Delsuc, F. (2009). An updated 18s rrna phylogeny of tunicates based on mixture and secondary structure models. BMC Evolutionary Biology, 9(1):187. 62

Wagner, E. and Levine, M. (2012). Fgf signaling establishes the anterior border of the ciona neural tube. Development, 139(13):2351–2359. 81

Weisblat, D. A. (2007). Asymmetric cell divisions in the early embryo of the leech helobdella robusta. Prog Mol Subcell Biol, 45:79–95. 59

Yasuo, H. and Hudson, C. (2007). Fgf8/17/18 functions together with fgf9/16/20 during formation of the notochord in ciona embryos. Dev Biol, 302(1):92–103. 81

2.3 Materials and methods

2.3.1 Imaging of Phallusia mammillata embryos

Individual membranes of live Phallusia mammillata embryos were imaged using a light-sheet microscope (MuVi-SPIM, EMBL, Heidelberg; [Krzic et al., 2012]. Membranes were labeled by microinjection of mRNA encoding PH-GFP (45pg per oocyte) synthesized using a pRN3-PH-GFP construct (a kind gift from Alex McDougall, Observatoire Océanologique de Villefranche-sur-Mer, France) as template and an mMessage mMachine T3 Ambion transcription kit following the manufacturer's instructions. PHGFP was excited with a 488nm laser (LuxX 488-60, Omicron) with simultaneous two-sided illumination. The emitted light was collected by a 25x Nikon water dipping objective lens (NA 1.1) combined with a tube lens with a focal length of 300mm leading to a 37.5 fold image magnification. The emitted light was filtered through a band-path BrightLine 525/30 filter (Semrock) and collected by a Hamamatsu Flash 4 SCMOS camera. At each time point, two perpendicular 3D image stacks, were acquired by both cameras. This resulted in four views (0, 90, 180, 270 degrees) of the specimen with a lateral resolution of $0.173\mu m \times 0.173\mu m$ and $1\mu m$ section spacing. Whole embryo stacks were acquired with a frequency of every two minutes. The embryos were imaged in artificial seawater at a temperature of 18C and mounted without embedding on top of a 0.8% GelRite (SIGMA, G1910) support.

2.3.2 Pre-treatment of the intensity images and multi-angle fusion

All the intensity images from the microscopes are 3D volume images of 1600x1700x210 voxels, with a voxel size of 0.17x0.17x1 μm. The images are first automatically cropped and downsized to a resolution of 0.3x0.3x1 μm.

Once the 4 images ($\{I_t^a\}_{a\in[1,4]}$) from the 4 angles of a given time are cropped and downsized, they are fused to build a 3D isotropic image I_t of this given time point improving the global quality (See Supp. Fig. 2.12). The fusion is done in 2 steps. i) The 4 images are first register onto the same referential. ii) The 4 registered images are fused together to create I_t.

To register the 4 images, the referential is first arbitrarily chosen as the referential of the first image I_t^1. Then the 3 affine transformations $\mathcal{T}_{1\leftarrow a}$ that register the frames of I_t^a onto the frame of I_t^1 are computed using the blockmatching algorithm described in [Guignard et al., 2014] (see Supp. Fig. 2.11 and section 3.2).

Then the 3 registered images plus the reference image are averaged. For this average, each voxel is pondered according to its distance to the camera. The contribution of a voxel to the averaged fused image is computed as a hill function of the distance to the camera and decrease with its distance to the camera.

2.3.3 ASTEC pipeline description

2.3.3.1 Definitions

$I_t \colon \mathbb{R}^3 \to \mathcal{I} \subset \mathbb{N}$ is the intensity image at time t (generally, $\mathcal{I} = [0, 2^8 - 1]$ or $\mathcal{I} = [0, 2^{16} - 1]$).
$S_t \colon \mathbb{R}^3 \to C_t \subset \mathbb{N}$ the segmented image at time t where C_t is a finite set of labels identifying in a unique manner each cell snapshots of the embryo at time t, the label 1 being reserved to the "exterior cell".
$\{I_t\}_{t\in[t_0,t_f]}$ is a sequence of intensity images.
$\{S_t\}_{t\in[t_0,t_f]}$ is a sequence of segmented images. To ensure consistency of the labels throughout the sequence it is ensured that:

$$\forall i, j \in [t_0, t_f]^2, i \neq j \Rightarrow C_i \cap C_j = \varnothing \tag{2.1}$$

This last property implies that for a given sequence $\{S_t\}_{t\in[t_0,t_f]}$, each cell snapshot has a unique identifier. We can also define the operator time that maps a unique time (and consequently a segmentation) to a cell snapshot label c:

$$\text{time}(c) = t \Leftrightarrow c \in C_t \tag{2.2}$$

2.3.3.2 Segmentation propagation pipeline

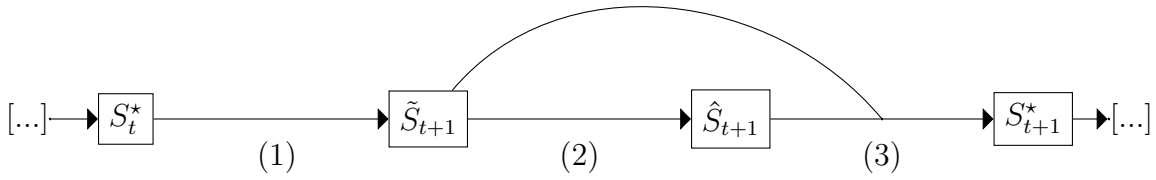


Figure 2.6 – ASTEC propagation pipeline:
(1): Segmentation projection,
(2): Segmentation estimation,
(3): Segmentation consistency checking

ASTEC performs both cell segmentation and tracking iteratively by propagation of segmentations from left to right on the image sequence $\{I_t\}_{t\in[1,N]}$.

Initialization To initiate the process, a segmentation S_{t_0} of the image I_{t_0} is first computed using **MARS** segmentation Fernandez et al. [2010] algorithm and manually corrected to produce the segmented image $S_{t_0}^\star$.

Iteration Then, assuming by induction that a segmentation $S_t^\star$ has been obtained at time t, the algorithm propagate $S_t^\star$ into $t+1$ to build the segmentation at time $t+1$. This is done in 3 steps (see fig. 2.6.):

1. **Segmentation projection**: projection of the segmentation from time t onto time $t+1$. This step produces a segmented image $\tilde{S}_{t+1}$. However, this fails to capture cell division events.

2. **Segmentation estimation**: to correct potentially missed divisions, an estimation of the segmentation at time $t+1$ knowing $\tilde{S}_{t+1}$ is performed. This step produces the segmented image $\hat{S}_{t+1}$.

3. **Segmentation consistency checking**: Since the two previous segmentations can give different results, a checking of the consistency of the segmented cells of $\hat{S}_{t+1}$ and $\tilde{S}_{t+1}$ is ultimately done to ensure the best possible segmentation. This last step produces $S^{\star}_{t+1}$, the final segmentation.

Post-processing Finally, a **Post-correction** algorithm is applied to get rid of the remaining errors. This algorithm is based on the analysis of the lineage tree consistency.

We detail thereafter these different steps.

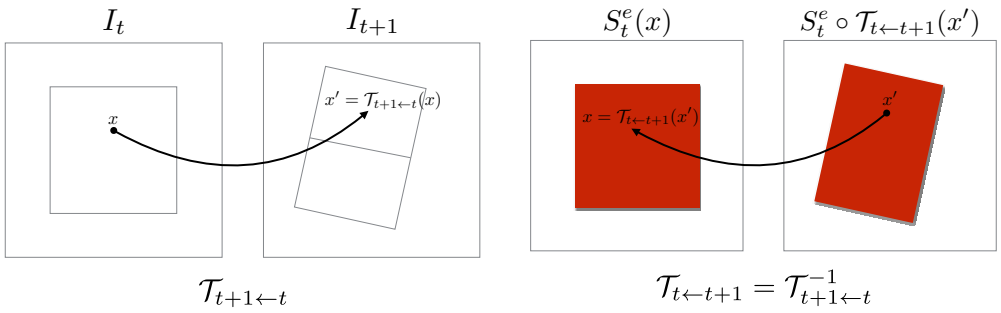
2.3.3.3 Initialization

To perform the segmentation of I_{t_0} we use the MARS algorithm corresponding to the pipeline described in Fernandez et al. [2010]. In this segmentation algorithm, the seeds are determined by the h-minima operator that finds the set of local minima regions in the smoothed image $I^{\sigma_1}_{t_0}$ that are separated by a minimum height of h. Computing h-minima of I_{t_0} consists in first subtracting h from I_{t_0} and then perform iterative grey-level dilation while remaining "under" the image I_{t_0}. Second, this image, which is I_{t_0} where peaks of height h have been erased, is subtracted from I_{t_0} which yields an image of peaks. Last, to retrieve the peaks and their spatial extension in a binary image, an hysteresis thresholding (high threshold of h, low threshold of 1) is performed. Finally, connected components of this binary image, which will be the seeds for the subsequent watershed operation are labeled. This image $Seeds^{h}_{t_0} = \text{reg-min}(I_{t_0}, h, \sigma_1)\colon \mathbb{R}^3 \rightarrow C_{t_0} \cup \{0\}$ where the reg-min operator performs the previous sequence of operations (smoothing, h-minima, hysteresis threshold and connected component decomposition). Voxels that do not belong to a seed are labeled 0. The watershed is applied to $Seeds^{h}_{t_0}$ and $I^{\sigma_2}_{t_0}$ and gives the segmented image S_{t_0}=$\mathbf{WS}(Seeds^{h}_{t_0}, I_{t_0}, \sigma_2)$.

Since this segmentation S_{t_0} will be the initiation of the global segmentation algorithm, it needs to be expertised to remove potential errors. To ease the correction, the h parameter is voluntarily chosen to favour over-segmentation and avoid under-segmentation for this step (small h). Over-segmented cells are manually fused to create the final segmentation of I_{t_0}: $S^{\star}_{t_0}$.

2.3.3.4 Segmentation projection

Assuming by induction that the segmentation $S^{\star}_{t}$ of I_t is given. The algorithm first projects $S^{\star}_{t}$ onto the frame of I_{t+1}. This process is split in 2 steps i) computation of the non-linear transformation that registers I_t onto I_{t+1}, ii) transformation of $S^{\star}_{t}$ onto the



(a) Computation of the non-linear vector field that register I_t into I_{t+1}. *(b) Computation of the projected eroded cells $S^e_{t+1\leftarrow t}$*

Figure 2.7 – Segmentation projection

frame of I_{t+1}. The non-linear deformation field ($\mathcal{T}_{t\leftarrow t+1}$), that allows to register images from the frame of I_t onto the frame of I_{t+1}, is computed using the block-matching algorithm described in Guignard et al. [2014] (see section 3.2, Figure 2.7a).

Each cell $c \in C_t$ of $S_t^\star$ is individually eroded, these eroded cells define regions $\{R_c^e\}_{c\in C_t}$. All the eroded regions R_c^e are then merged together to form S_t^e.
S_t^e is then registered onto the frame of $t+1$: $S^e_{t+1\leftarrow t} = S_t^e \circ \mathcal{T}_{t\leftarrow t+1}$ (Figure 2.7b). $S^e_{t+1\leftarrow t}$ is finally used as image of seeds for the watershed applied on I_{t+1}. The result of this watershed is the propagated segmentation $S_t^\star$ onto the time $t+1$:

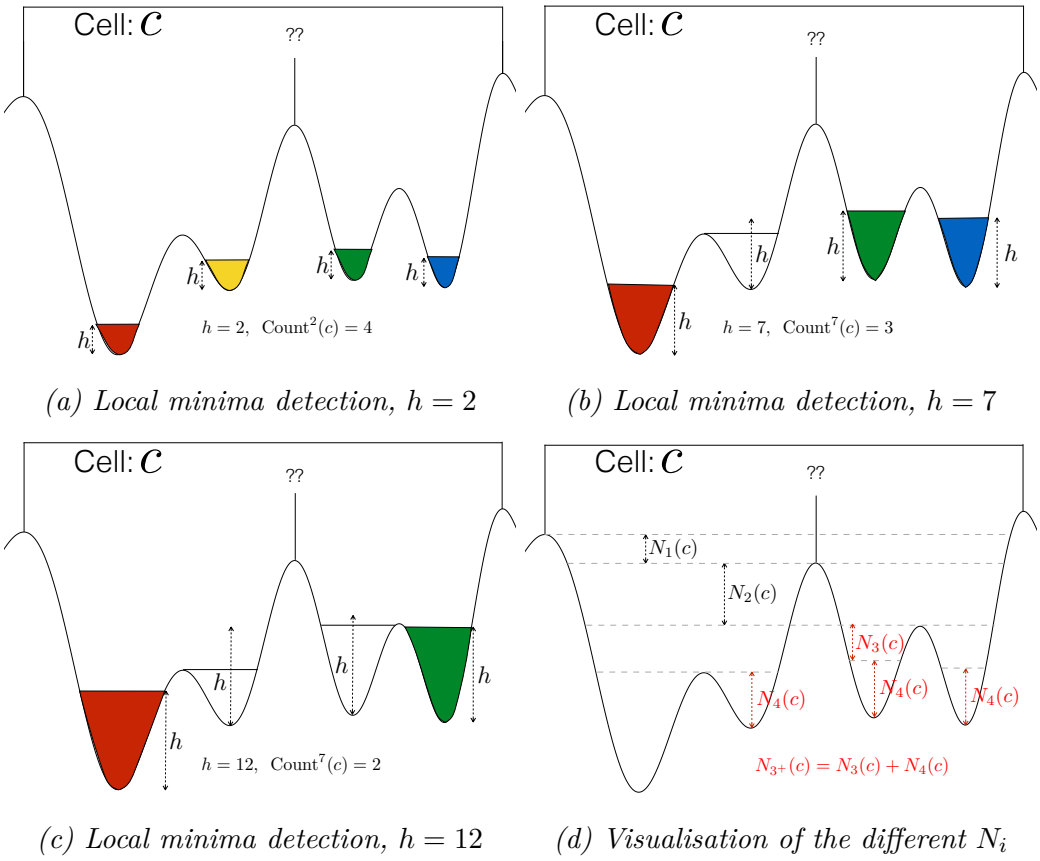
$$\tilde{S}_{t+1} \quad = \quad \mathbf{WS}(S^e_{t+1\leftarrow t}, I_{t+1}, \sigma_2) \tag{2.3}$$

2.3.3.5 Segmentation estimation

Segmentations resulting from the projection of time t onto $t+1$ have usually two major defects. First, as expected, cells that underwent division are under-segmented (the division is missed). Secondly, since the seeds used to build $\tilde{S}_{t+1}$ are issued from the transformation of eroded cells, they are not actual local minima. In this case, the watershed can under-perform and produce wrong shapes (the error is however at most of the size of the erosion).

To avoid these issues, h-minima will be recomputed locally, i.e. for each cell issued from $\tilde{S}_{t+1}$. Several h from an interval $[h_{min}, h_{max}]$ will be tested, and the point is now to determine the optimal h in $[h_{min}, h_{max}]$ for each cell. First, h of small value are likely to extract h-minima in a noisy background, so h have to be chosen above the amplitude of the noise (assumed to be additive) to be sure to avoid this drawback. Second, when a value of h yields two seeds (indicating a cell division), we have to make sure that the corresponding cell wall has a sufficient signal amplitude to be an effective wall.

1. To build Seeds_{t+1}, the set of seeds images $\{\text{Seeds}^h_{t+1} = \text{reg-min}(I_{t+1}, h, \sigma_1)\}_{h\in H\subset\mathbb{N}}$ is computed on the intensity image I_{t+1}. For each $c \in C_t$, the region occupied by c in $\tilde{S}_{t+1}$: $R_c = \{x \in \mathbb{R}^3 \mid \tilde{S}_{t+1}(x) = c\}$ is computed. Then, the number of



(a) Local minima detection, $h = 2$ *(b) Local minima detection, $h = 7$*

(c) Local minima detection, $h = 12$ *(d) Visualisation of the different N_i*

Figure 2.8 – h-minimum

seeds found in each region for each value of the h parameter of the seed detection Count$^h(c)$ is computed. Each region R_c can contain either 1 (if the cell c has not divided), 2 seeds (if the cell c has divided between t and $t+1$) or more, the latter corresponding to a value too small of h since more than 2 divisions can not occur in the sequences under investigation. Therefore, any h that yields more than two seeds in a cell c is below the noise amplitude. Thus a minimal amplitude of the noise is the maximum value of h that splits c into more than two regions. To decide whether the h yielding exactly two seeds correspond to noise, we compare the range of h yielding two seeds to this minimal noise amplitude.

To address this question the following metrics are computed. Let $N_{2^+}(c)$ and $N_{2^-}(c)$ denote respectively the maximal and minimal values of h that yields exactly two seeds, and by $N_2(c)$ the range of the interval $[N_{2^-}(c), N_{2^+}(c)]$ i.e. $N_2(c) = N_{2^+}(c) - N_{2^-}(c) + 1$. First, the signal amplitude that splits c into exactly two cells, $N_{2^+}(c)$. This value is the maximum h that splits c into two cells or more. Then the difference of amplitude between the signal that is known to be noise and the signal that splits c into exactly two cells, $N_2(c)$, is computed. This is the difference between the maximum and the minimum value of h that splits c into exactly two cells (Figures 2.8 and 2.9). $N_{2^+}(c)$ can be assumed as real signal (and therefore as membrane signal) if it is i) high enough and ii) significantly higher to

the noise. i) and ii) are respected if the product $s(c) = N_{2^+}(c).N_2(c) > \tau$ (fig 2.9). This criteria allows to decide whether there are two or one seeds in the projection of a cell c coming from S_t onto I_{t+1}, i.e. to decide whether there is a cell division or not. It has to be pointed out that this procedure allows to choose locally an optimal h for each cell projection.

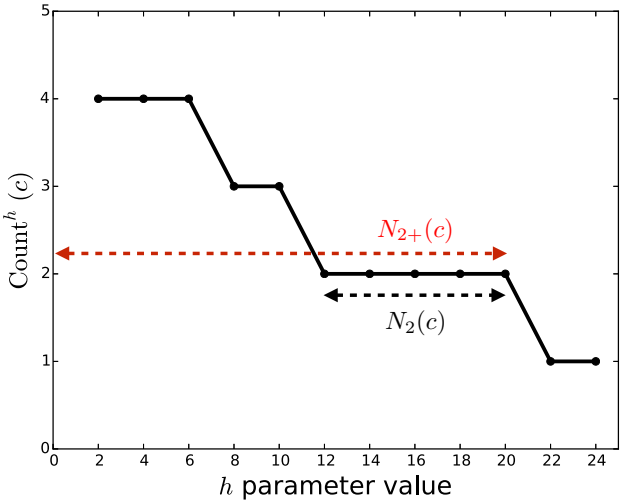


Figure 2.9 – Number of seeds found for a cell c for different values of the parameter h for the seed detection.

In any case, if multiple values h are valid for the targeted number of seeds, the highest one is chosen. It can happen that the number of seeds found is always higher than 2. In this case, either 3 seeds can be found and then they are used to segment c. Then the smallest resulting region from the three seeds is fused to the one it shares the most surface of contact with. In the case where at least 4 seeds were found in c the eroded cell of c from $S^e_{t+1\leftarrow t}$ is kept as the seed for the region c. Following the preceding rules a h and its corresponding set of seeds can be associated to each cell c.

The image of seeds, $\text{Seeds}_{t+1} : \mathbb{R}^3 \rightarrow C_{t+1} \cup \{0\}$, computed by the above operations is used as the image of seeds together with the intensity image I_{t+1} as the input of the watershed. This produces the estimation of the segmentation of I_{t+1}:

$$\hat{S}_{t+1} \quad : \quad \mathbb{R}^3 \rightarrow C_{t+1} \tag{2.4}$$

$$\hat{S}_{t+1} \quad = \quad \mathbf{WS}(\text{Seeds}_{t+1}, I_{t+1}, \sigma_2) \tag{2.5}$$

2. The tracking of the cells from t to $t+1$ is built together with this estimation operation. The cell snapshot c at time t is linked to the set $T(c)$ of its corresponding

cell snapshots at $t+1$:

$$T\colon C_t \rightarrow \mathcal{C} \tag{2.6}$$
$$c \mapsto T(c) = \{c' \in C_{t+1} \mid \exists x \in R_c, \mathrm{Seeds}_{t+1}(x) = c'\} \tag{2.7}$$

2.3.3.6 Segmentation consistency checking

On one hand, as said previously, $\tilde{S}_{t+1}$ is highly accurate in terms of cell shape (the error is at most the size of the erosion). But this error is likely to occur since the seeds used for the projection are not actual local minima. On the other hand, the estimated segmentation $\hat{S}_{t+1}$ can have bigger errors since the watershed has more freedom (more little seeds) but these errors are less likely to occur since actual local minima are used as seeds. Shape errors in $\hat{S}_{t+1}$ usually result in loss of material "eaten" by the outside, in the case of images where the outer membranes are faint (like in the case of *Phallusia*).

To take advantage of both methods volume consistency is checked between $\hat{S}_{t+1}$ and $\tilde{S}_{t+1}$ and corrected if needed to build the final segmentation $S^{\star}_{t+1}$. This consistency checking is done in two consecutive steps, (1) a first checking spots the gross inconsistencies of volume and tries to correct them. This step is followed by (2) a fine checking adjusting the potential remaining mistakes.

(1) Cells in $\hat{S}_{t+1}$ and in $\tilde{S}_{t+1}$ are linked by T and thus can be compared. If a cell in $\tilde{S}_{t+1}$ is at least 50% bigger than its corresponding cells, it is checked for correctness. The correction then consists in increasing the number of seeds in order to cover more space and avoid matter loss. If the cell was considered not divided by the previous steps, then it is divided into two cells by the correction procedure if possible. If not the cell snapshot is voluntarily over-segmented to maximize the covered surface by the seeds and minimize the possibility of volume loss. The over-segmented cells are then fused.

(2) Ultimately, a last checking is done to recover from errors due to lost matter to the exterior due to too faint membranes. The volume ratios are re-processed between $\tilde{S}_{t+1}$ and the new segmentation S'_{t+1}. The cells in S'_{t+1} that are at least 10% smaller than their equivalent in $\tilde{S}_{t+1}$ are tagged for correction. Let c be a cell snapshot to correct. First, if the c was divided by the algorithm, this division is cancelled and the two sister are fused. Then an active contour algorithm (morphosnake algorithm described in Marquez-Neila et al. [2014] and implemented by P. Marquez Neila) is applied using the dilated shape of c as the initial contour and the gradient norm transformation of I_{t+1} as the intensity profile. The algorithm is applied up to stability (at $\pm\delta$ voxels) or after n iterations. After this procedure, the outer region(s) in S'_{t+1} that is/are included in the resulting shape of the morphosnake algorithm are attributed to the cell that was corrected. This procedure creates a new temporary segmentation S''_{t+1}.

To avoid over-expansion of outer cells due to mistakes in the morphosnake algorithm, an opening operation is applied to the whole embryo as a unique object to remove the potential outer protrusions in the cells corrected by this last step. This gives the final segmentation $S^{\star}_{t+1}$ along with the tracking of the cells from t to $t+1$ $T\colon C_t \rightarrow C_{t+1}$.

2.3.3.7 Post-correction

Ultimately, once the segmentation time-series $\{S_t^\star\}_{t\in[t_0,t_f]}$ is created by the algorithm, four type of errors can be identified: splitting a cell into two or more ($\mathbf{E_s}$), merging two or more cells into one ($\mathbf{E_{me}}$), missing a cell ($\mathbf{E_{mi}}$) and adding a cell that does not exist ($\mathbf{E_a}$).

The propagation algorithm design implies that type $\mathbf{E_{mi}}$ errors are uncommon since the cells are propagated from time to time (it can only happen for outer cells with faint outer membranes). As for $\mathbf{E_{mi}}$ errors, we empirically assessed that they do not occur often (we never identify this type of error) and decided not to take care of. Type $\mathbf{E_a}$ implies that a new cell would appear on the outside of the embryo. Since the cells are propagated from one time to the next, this kind of errors are also unlikely to happen. As for $\mathbf{E_{mi}}$ errors, we empirically assessed that they do not occur often (we never identify this type of error) and decided not to take care of. $\mathbf{E_{me}}$ errors are missed divisions and type $\mathbf{E_s}$ errors are either divisions occurring too early or non-existent divisions. To deal with these two types of errors, we favour type $\mathbf{E_s}$ over $\mathbf{E_{me}}$ by setting a low value for the parameter τ. That favours the choice of two seeds over one for each cell. We then fuse the cells that are issue of type $\mathbf{E_s}$ errors by applying a post-correction algorithm.

This post-correction algorithm takes advantage of the propagation properties of the sequential segmentation and biological knowledge on the studied organism:

1. The propagation only allows to split a cell into two from one time to the next, therefore E_s errors are only **one** cell over-segmented into **two**.

2. E_s errors are propagated which implies that they cannot be expressed for only one time point.

3. If a cell c is over-segmented into two cells c_1 and c_2, then the sum of the volumes of c_1 and c_2 should be equal to the volume of c.

4. If a cell c is over-segmented into two cells c_1 and c_2, then the fake membrane that split it into two is supported by noise. This noise is unstable from one time to the next and therefore the volume of the cells c_1 and c_2 should not be constant.

5. The volume of a correctly segmented cell remains constant.

Altogether, these properties together imply that the volumes of two cells c_1, c_2 resulting of the over-segmentation of a cell c have their volumes evolving in an anti-correlated manner.

The post-correction algorithm first gets rid of trivial over segmentations. The cells that have their volume decreasing and that end up disappearing before attaining a life-span of 50 minutes are automatically fused to their corresponding sister. Then, using the property of volume anti-correlation, the sister cells that share anti-correlated volume evolution (Pearson correlation under -0.9) are fused together. Since each cellular lineage

tree is independent they are treated individually and sequentially. After this treatment, all the over-segmented cells that are the result of the propagation of ASTEC are fused.

Parametrization For this study on *Phallusia mammillata* embryos, we used a gaussian filter with a $\sigma_1 = .6\mu m$. For the segmentation of the first time point of our dataset we used the following parameters: $\sigma_1 = .6\mu m$, $\sigma_2 = .15\mu m$ and $h = 4$. The erosions on the cells of S_t were done using 3D 6 connected structuring element. These erosions were done 10 times per cell. The range of h for this dataset was $H = \{2n \mid n \in [1 \dots 9]\})$ and the τ value used to discriminate the division were $\tau = 100$.

To accelerate the computation of $\{\text{Seeds}_{t+1}^{h}\}$, it is interesting to notice that:

$$\forall (i,j) \in \mathbb{N}, \forall h_i \in H_i(c), \forall h_j \in H_j(c) \quad : \quad i < j \Leftrightarrow h_i > h_j \tag{2.8}$$

That allows us to reduce the search space for the local minima recursively using the output of Seeds_{t+1}^{h} to compute Seeds_{t+1}^{h-1}. Moreover, the cell erosions, the morphosnakes processes were computed in parallel at each time point.

2.3.4 Manual curation of segmented embryos.

2D and 3D manual curation of segmented embryos was made using the commercial software AMIRA. Segmented objects (cells) and the fused membrane fluorescence data were overlaid for the data curation. The brush tool was used to add or remove voxels where the automatic segmentation did not match the cell contours defined by the fused fluorescence data.

For the 3D manual curation each cell was individually corrected, plane by plane, along 3 different views.

For the 2D manual curation all cells were individually analysed and all detected cell shape mistakes corrected. Cells where without a 3D perspective one could not be sure of the presence of a segmentation mistake were not corrected and not taken into account on the cell shape quality analysis of the automatic segmentation method here described.

2.3.5 Cell lineage tree distance

To compute distance between cell lineage trees we used the distance measure between plant architectures described in Pascal Ferraro and Christophe Godin [2000]. The distance between two given trees T_1 and T_2 is the minimum cost necessary to transform T_1 (resp. T_2) in T_2 (resp. T_1). The cost is defined by the sum of the atomic operations costs to transform T_1 in T_2. The atomic operations are matching, deletion and insertion. In this study, since the goal is to compare the division pattern the matching cost is set to 0 and the deletion and insertion costs are set to 1 (Figure 2.10). In practice, to be able to compare cell lineage trees of different sizes, this edition score is normalised to the average number of cell snapshots in the lineage trees that are compared.

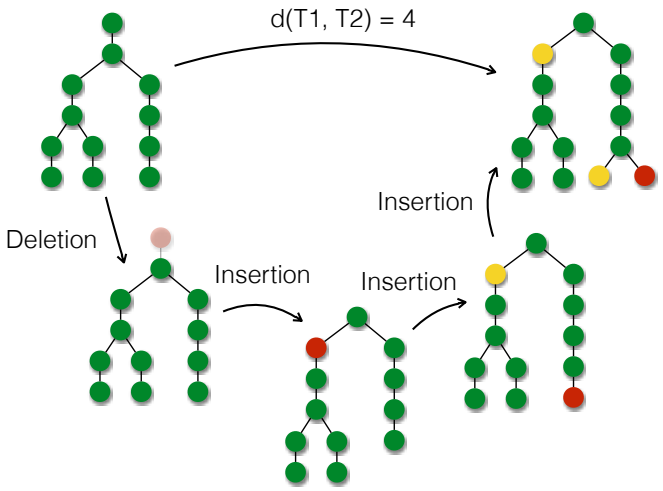


Figure 2.10 – Example of lineage tree distance computation. The red cell snapshots are currently changed, the orange ones were changed previously.

2.3.6 Model of differential induction

We addressed the question of understanding how a cell can give birth to a progeny of two daughter cells that give rise subsequently to different cell fates. Based on preliminary partial results obtained on Ciona, we make the hypothesis that cell fate is induced by juxtacrine signalling, and that physical cell-cell contacts are key determinants in this process.

To test this hypothesis, we used the geometric information acquired on our digitized embryo and built a model of cell-cell signalling interactions taking into account the precise area of cell-cell contacts. The model aims at identifying cells that are differentially induced by a set of pre-specified secreted signalling ligands or antagonists. Two cases are distinguished (see figure 5B).

- Case 1: Differential sister cell induction. Two sister cells are differentially induced during their lifespan.
- Case 2: Mother cell polarization. The cortical regions of the mother cell that will be inherited by the two sisters receive polarized signals during the lifespan of the mother.

2.3.6.1 Case 1: Differential sister cell induction.

Notations : For a cell i, we denote $N(i)$ the set of its cell neighbours (i.e. cells that have a non null contact surface with i). The surface area of a cell i is denoted A_i and the area of contact between two neighbouring cells i and j is A_{ij}.
Let us denote $\mathcal{P}$ and $\mathcal{A}$ the sets of paracrine and autocrine ligands respectively.
Let g_j be the concentration of a ligand G synthesized in cell j and call $s_i(G)$ the intensity

of the signal triggered in cell i by G (number of molecules of internalized signal). In the case of a dual paracrine/autocrine signalling, we assume that :

$$s_i(G) = \sum_{j \in N(i)} \alpha A_{ij} g_j + \alpha A_i g_i \tag{2.9}$$

where the first term is the paracrine component of the signalling and the second term is the autocrine component. α is a constant parameter of the dimension of a length m. In case of pure paracrine (resp. pure autocrine) signalling, the first (resp. the second) term would disappear from the equation.

The concentration of signalling ligands/antagonist cannot be experimentally measured, in our system. We therefore make the assumption that the concentration of ligand/antagonist intensity emitted by all cells is the same. Therefore, in the model the signalling intensity $s_i(G)$ becomes proportional to the area of contact with ligand/antagonist expressing cells:

$$s_i(G) = \alpha\Big(\sum_{\substack{j \in N(i), \\ j \text{ expresses } G}} A_{ij} + A_i \Big) \tag{2.10}$$

In cell i, the model then assumes that a number of ligands/antagonists are able to activate a specific pathway Pw. Let us denote $L^+(Pw)$ (resp. $A^-(Pw)$) the set of ligands/antagonists able to activate (resp. to inhibit) Pw. Then the total activation signal $S_i^+(Pw)$ received by cell i for pathway Pw can be written as:

$$S_i^+(Pw) = \sum_{G \in L^+(Pw)} s_i(G) \tag{2.11}$$

Similarly, the total inhibiting signal received by cell i to silence pathway Pw is:

$$S_i^-(Pw) = \sum_{H \in A^-(Pw)} s_i(H) \tag{2.12}$$

We assume that these upstream signals act on the downstream activity of pathway Pw depending on their intensity. We distinguish 3 intensity levels corresponding respectively to the absence or presence of trace levels (level '-'), intermediate level (level '+') or high levels (level '++') of signalling intensity received by the pathway Pw. At trace levels, the ligands/antagonist will not impact the downstream activity of Pw. At intermediate levels, the downstream activity of the pathway may or may not be triggered. At high levels, the downstream activity of the pathway is in all cases triggered.

Two threshold parameters ti and TI separate these three signalling levels. They are estimated from the distributions of signalling intensities received by neighbours of cells expressing each ligand/antagonist G of Pw. For each ligand, Gti and GTI correspond to positions in this distribution noted by percentages (e.g. $Gti = 50\%$ and $GTI = 90\%$ means that 1) a cell that receives a signalling intensity inferior to that received by half

of the neighbours of G-expressing cells would be considered to receive trace signals; 2) a cell that receives a signalling intensity higher than 50% of neighbours, but lower than the top 10% of neighbours receives an intermediate signal; 3) a cell that receives a signalling intensity corresponding to that experienced by the top 10% of neighbours receives a high level of signalling). These values are integrated at the level of the pathway (for a ligand) :

$$ti = max_{G \in L^+(Pw)}\{Gti\}.$$

For two sister cells, i and j, we then compare the relative level of signalling intensities received by the sisters, $S_i(Pw)$ and $S_j(Pw)$, triggering pathway Pw in each cell. We say that there exists a *differential signalling intensity* of $S_i(Pw)$ and $S_j(Pw)$ received by these cells if the three following conditions are all satisfied:

- at least one of the signals is significantly intense, i.e. $S_i(Pw) > ti$ or $S_j(Pw) > ti$.
- one of the signals is significantly higher than the other, i.e. $S_i(Pw)/S_j(Pw) > r$ or $S_j(Pw)/S_i(Pw) > r$, r being a minimal differential threshold parameter of signalling intensity (a parameter in the model).
- at least one of $S_i(Pw)$ and $S_j(Pw)$ is below the high signalling intensity level (i.e. less than TI).

With these definitions, we can formalize the cases in which a pathway Pw can be induced in our differential induction. For two neighbouring cells, i and j, we consider whether the activation signals are differentially expressed or not in the two cells and similarly with the inhibition signals. This results in 4 cases that are depicted in table 2.1 and described hereafter.

1 both activation and inhibition signals are not differential. In this case, there is no differential activation of pathway Pw, which is therefore not a candidate for the differential induction of the two sisters.

2 if the activation signal is differential, but not the antagonist signal, then there is a differential induction only if the level of the inhibitory signal is not significant (level '-'). In this case the agonist drives the differential induction.

3 if the activation signal is not differential, but the antagonist signal is, then there is a differential induction only if the level of the activation signal is intermediate (level '+') but not high (level '++'). In this case, the antagonist signal drives the differential induction.

4 if both the activation and the inhibition signal are differential, and the sister receiving the strongest agonist signals also receives the weakest antagonist signal, then there is a differential induction. In this case ligands and antagonist cooperate to differentially induce the two sisters. Otherwise (both activation signals and inhibition signals are highest in the same sister), then we consider that there is no differential induction.

2.3.6.2 Case 2: Mother cell polarization.

The differential inductions follow the same principle and uses the same thresholds as in the case of sister induction, except that the signalling intensities are calculated at the level of the cortical regions of the mother cell that will be inherited by the two sisters after its division.

To identify the part of the cortex that will be inherited by each sister, we calculate the surfaces of contacts of each sister with its neighbours. We then consider that the part of the cortex of the mother that will give rise to sister A has the same surface of contact with cells expressing ligand G as A with its G-expressing neighbours. For example, in the notochord lineage A6.2 is polarized to give rise to A7.4 (nerve cord) and A7.3 (notochord), and is polarised by Ephrin Ad signalling coming from the a-line. The surface of A6.2 that will be inherited by A7.4 is that that contacts the antecedents of the a-line neighbours of A7.4. We thus consider that the surface of contact of A6.2 with the a-line that will be inherited by A7.4 is the sum of the contacts between A7.4 and its ephrin-expressing a-line neighbours.

2.3.6.3 Estimating ligand/antagonist spatio-temporal availability

Description of in situ hybridization profiles with cellular resolution for all ligands and inhibitors of signalling pathways were obtained from the ANISEED database.

To account for the translation, processing and secretion of the signalling ligand/antagonists, we introduced for the FGF and Wnt pathways a 40 minutes (or one developmental stage) delay between onset of RNA expression and protein availability. This delay was compatible with known inductions in ascidians. We then considered that the secreted ligand/inhibitor protein signalled for less than 30 minutes after its production (Notch, Nodal, Bmp) or for a time comprised between 30 and 60 minutes after production (FGF, Wnt).

2.4 Tables

$S^+(Pw)$ \ $S^-(Pw)$	not differential	differential
not differential	No	Yes if S^+ significant
differential	Yes if S^- not significant	Yes if opposite significance

*Table 2.1 – **Differential induction rules**: S^+ and S^- correspond to intensities of activation and inhibitory signals in neighboring cells i and j. Depending on whether these activation/inhibition signals are differentially expressed in these cells, our model specifies the conditions for a differential induction of one of the cells (i.e. the other one remaining not induced)*

Mother cell	Fate daughter D_1	Fate daughter D_2	Found by model
A6.2	Notochord	TNC + VG + PSV + Neck	YES
A6.3	Head Endoderm	TLC	YES
A6.4	Notochord	TNC + VG + Muscle	YES
B6.1	Head Endoderm	Endodermal Strand + Head Endoderm	NO
B6.2	Mesenchyme + Notochord	Muscle	YES
B6.4	Mesenchyme	Muscle	YES
B6.3	TVC	Germline	YES
a6.7	SV + Trunk Epidermis	Trunk Epidermis	YES
b6.5	Tail Epidermis + Muscle	TNC + VG + PSV + Tail Epidermis	YES
A7.4	PSV + TNC	TNC + VG + PSV + Neck	NO
A7.8	TNC + VG	Muscle + TNC	YES
B7.3	Mesenchyme	Notochord	YES
a7.9	ASV	Palps + NH	YES
a7.10	ASV	Palps + NH	YES
a7.13	ASV	Trunk Epidermis	YES
b7.9	Muscle + Endodermal Strand + TNC	Tail Epidermis	YES
b7.10	TNC + VG + PSV	Epidermis	YES
A8.7	TNC + PSV	PSV	YES
A8.8	TNC + VG	PSV + Neck + VG	YES
A8.15	TNC + VG	VG	NO
A8.16	Muscle	TNC	YES
a8.18	Palps + NH	Palps + NH	NO
a8.20	Palps + NH	Palps + NH	NO

Table 2.2 – ***Known fate decision events.*** *All known fate decision events used to train the model. For this table the mother AX.n gives rise to the two daughters D_1: $A(X+1).(2n-1)$ and D_2: $A(X+1).(2n)$. ASV: Anterior Sensory Vesicle, NH: Neurohypophysis Primordium, CEN: Dorsal Caudal Epidermal Neurone, PNS: Peripheral Nervous System, TNC: Tail Nerve Cord, VG: Visceral Ganglion, PSV Posterior Sensory Vesicle, ATEN: Apical Trunk Epidermal Neurones, TVC: Trunk Ventral Cells, TLC: Trunk Lateral Cells*

Mother cell	Fate daughter 1	Fate daughter 2	Induction timing	Pathway	Found by model	Ref
A6.2	Post ventral NP (A7.4)	Notochord (A7.3)	Mother	FGF/ERK	YES	(1)
			Daughter 2	FGF/ERK	YES	(2)
			Daughter 1	Nodal	YES	(2)
A6.4	Lateral NP/2^{nd} muscle (A7.8)	Notochord (A7.7)	Mother	FGF/ERK	YES	(1)
			Daughter 2	FGF/ERK	YES	(2)
			Daughter 2	Nodal	YES	(3)
A6.3	TLC (A7.6)	Head Endoderm (A7.5)	Mother	FGF/ERK	YES	(4)
			Daughter 1	Nodal	YES	(5)
B6.4	Primary Muscle (B7.8)	Mesenchyme (B7.7)	Mother (Daugthers)	FGF/ERK	YES	(6)
a6.7	Head epidermis (a7.14)	Lateral NP (a7.13)	Daughter 2	BMP	YES	(7)
A7.4	NP (col 1) (A8.8)	NP (col 2) (A8.7)	Daughter 1	Notch	NO	(8)
A7.8	Lat NP (A8.15)	2^{nd} muscle/Lat NP (A8.16)	Mother	Notch	YES	(8)
B7.3	Notochord (B8.6)	Mesenchyme (B8.5)	Mother	Notch	YES	(5)
			Daughters	Notch	YES	(5)
a7.9	NP (row III/IV)	NP (row V/VI)	Daughters	ERK[1]	YES	(9)
a7.10	NP (row III/IV)	NP (row V/VI)	Daughters	ERK[1]	YES	(9)
a7.13	a8.25	a8.26	Daughters	ERK/FGF	YES	(9)
a8.19	a9.37	a9.38	Daughters	ERK/FGF	YES	(9)
a8.17	a9.33	a9.34	Daughters	ERK/FGF	YES	(9)
a8.25	a9.49	a9.50	Daughters	ERK/FGF	YES	(9)

Table 2.3 – ***Fate inductions where the ligands are known in ascidians between the 32 and the early gastrula.*** [1]*: ligands unknown. References: (1): [Picco et al., 2007], (2): [Yasuo and Hudson, 2007], (3): [Hudson and Yasuo, 2005], (4): [Shi and Levine, 2008], (5): [Hudson and Yasuo, 2006], (6): [Imai et al., 2002], (7): [Ohta and Satou, 2013], (8) [Hudson et al., 2007], (9): [Wagner and Levine, 2012]*

2.5 Supplementary figures

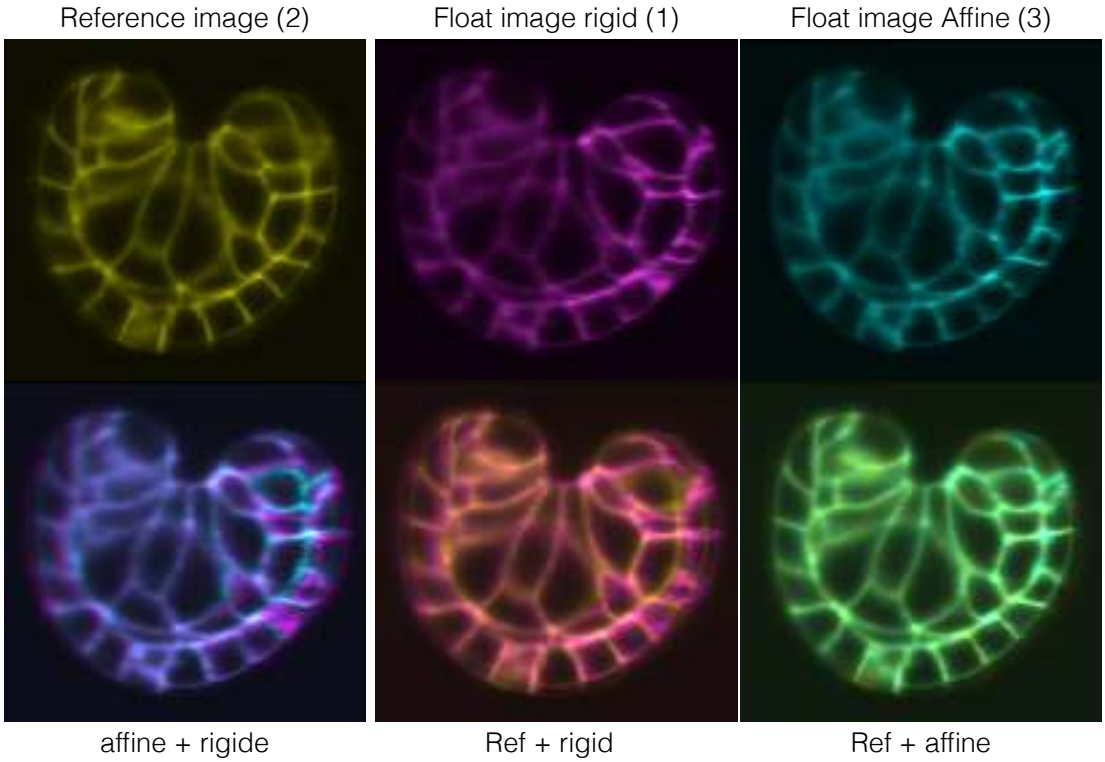


Figure 2.11 – ***Rigid versus affine registration of the different angles of the acquisition of an embryo at a given time from two angles.*** *Two cross sections of the acquisition of the embryo at the time 152 minutes. A referential image is used to register the 3 other floating images, either with a rigid or an affine registration (top). The bottom row shows the differences between rigid and affine registration and that the affine registration is necessary to correctly register the two angles in a similar frame.*

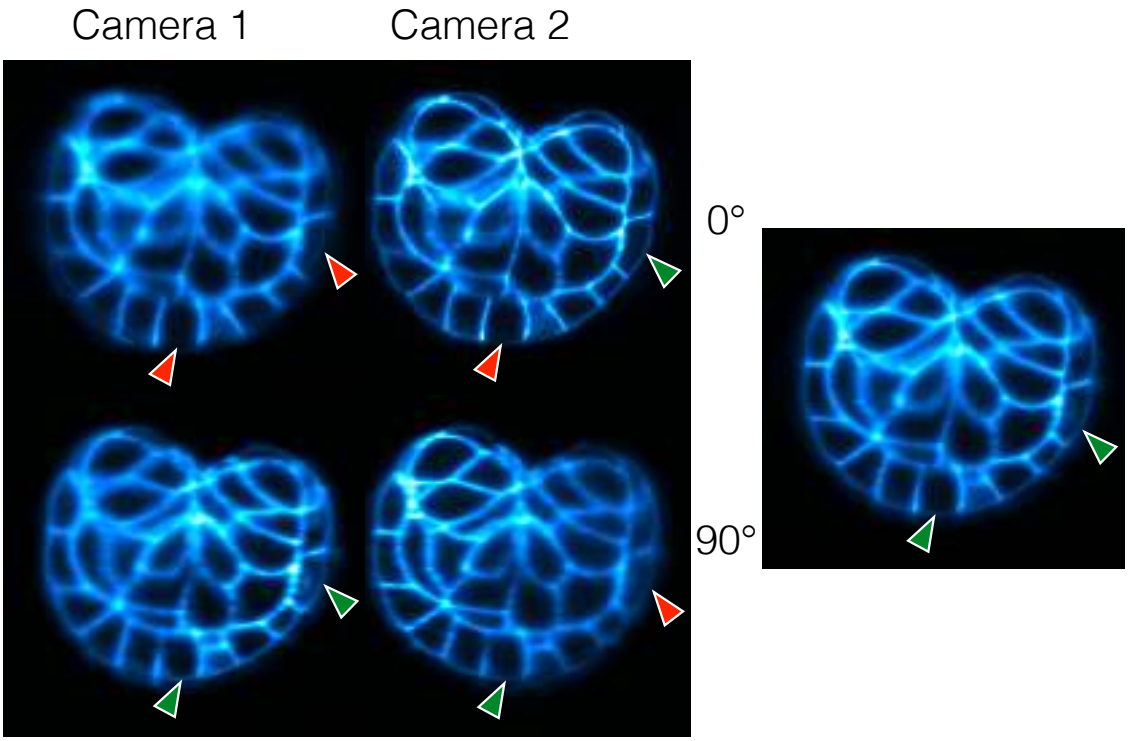


Figure 2.12 – ***Complementary contribution of individual views to the fused image.*** *Left: Matching optical sections through the same embryo, but acquired from 4 different angles of views. Note the high similarity of the geometry of the embryo in all images. Right: Resulting fused image. Arrow heads point to membranes of interest. Green arrow heads: high quality signal. Red arrowheads: faint or absent signal.*

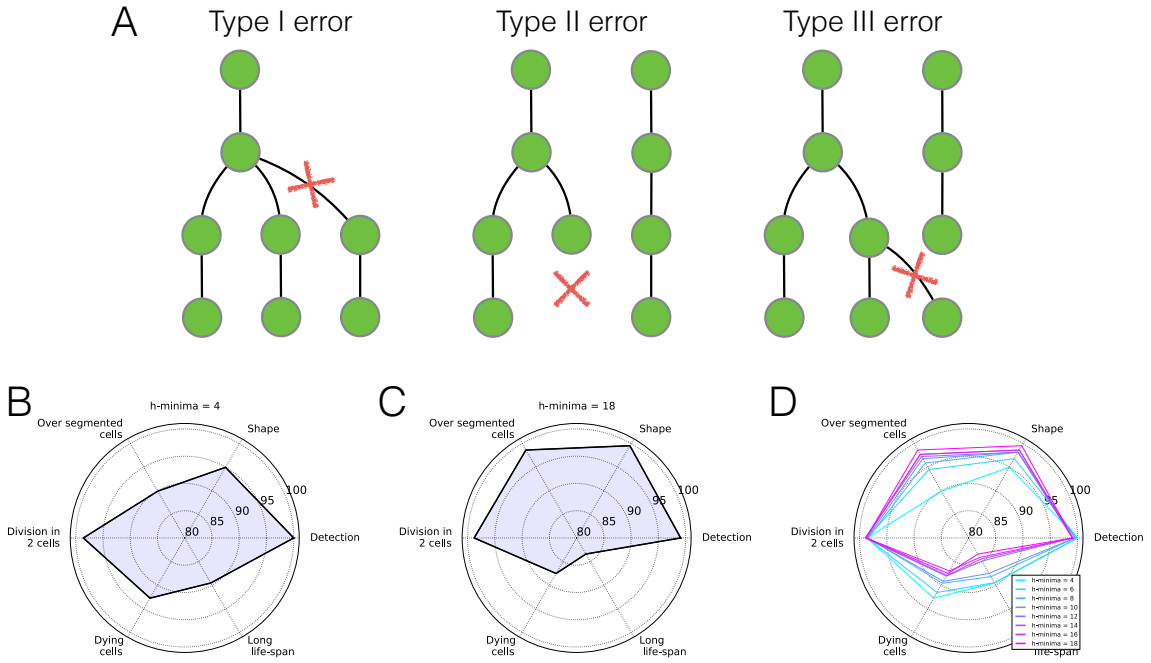


Figure 2.13 – ***Quality of the segmentations and tracking obtained using MARS-ALT with different*** *h* ***values, after post-correction.*** *A) Types of cell lineage errors. B-C) Spider graphs presenting the analysis of the segmentation and tracking quality for h-min values of 4 (B) and 18 (C). Detection: percentage of true cells detected by the segmentation algorithm; Shape: percentage of well-allocated voxels in accurately detected cells; Cell division: percentage of cells giving rise to at most 2 daughters (a measure of type I errors); Uninterrupted progeny: percentage of cells that either divide to produce two daughter cells, or live until the end of the film (a measure of type II errors); Long lifespan: percentage of cells with a lifespan* $\geq$ *30 minutes (a measure of type III errors). D) Spider graph on which the analyses of all h-min values are superposed, to allow their comparison. Note that no single h-min value optimizes all scores: smaller h-min values give better uninterrupted progeny scores, higher h-min values give better shape scores.*

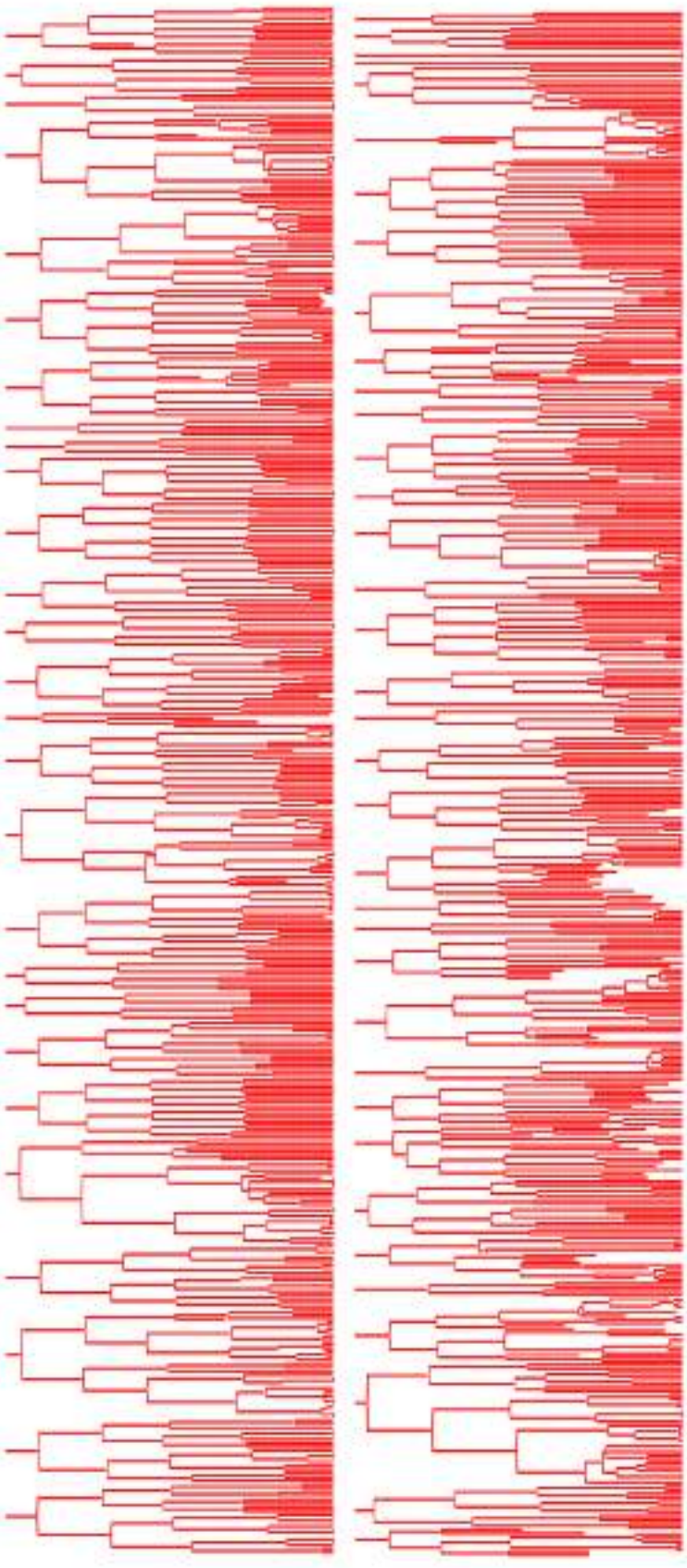


Figure 2.14 – Cell lineage tree resulting of MARS-ALT using the optimal parametrization.

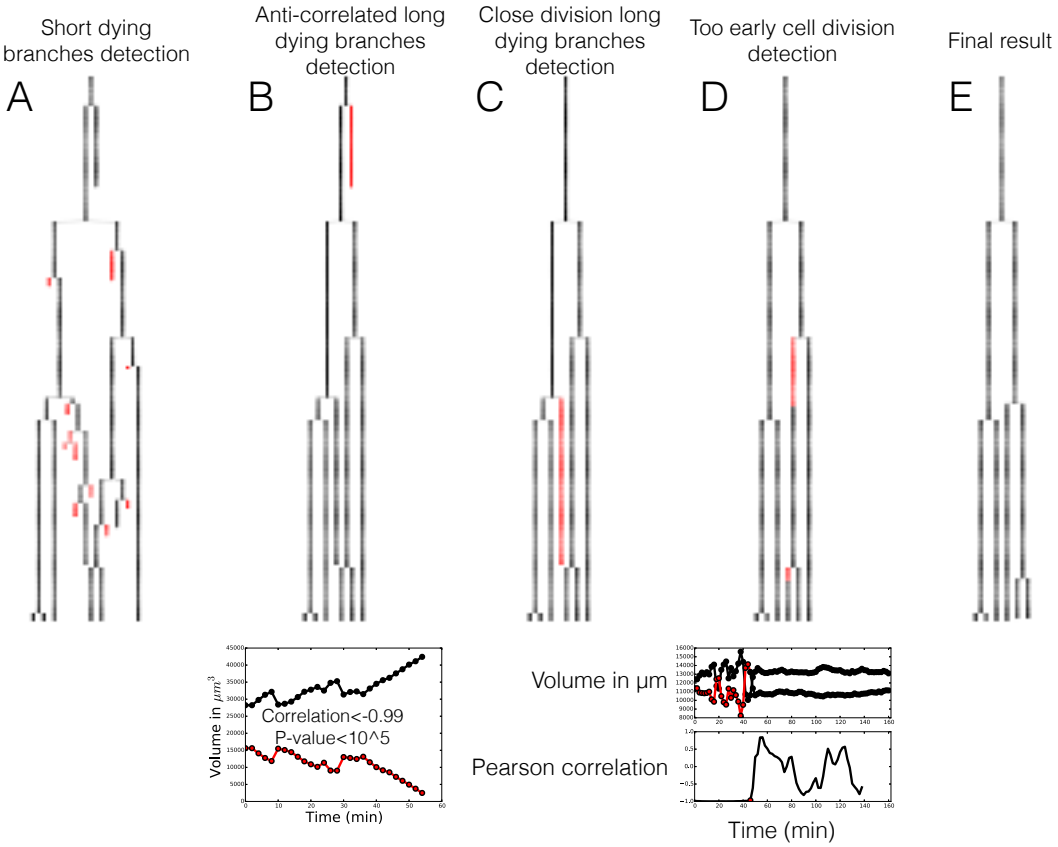


Figure 2.15 – A) Short interrupted branches (< 10 time points, red) are first fused to their sister branch. B) Long dying branches (≥ 10 time points) with anticorrelated cell volume to their sister branch are fused to heir sister. C) Cells with close timing division are checked. D) The precise timing of each cell division is checked for anticorrelation between sister cells, in which case, cell division is post-poned until the end of the anti-correlation period. E) Final corrected cell lineage.

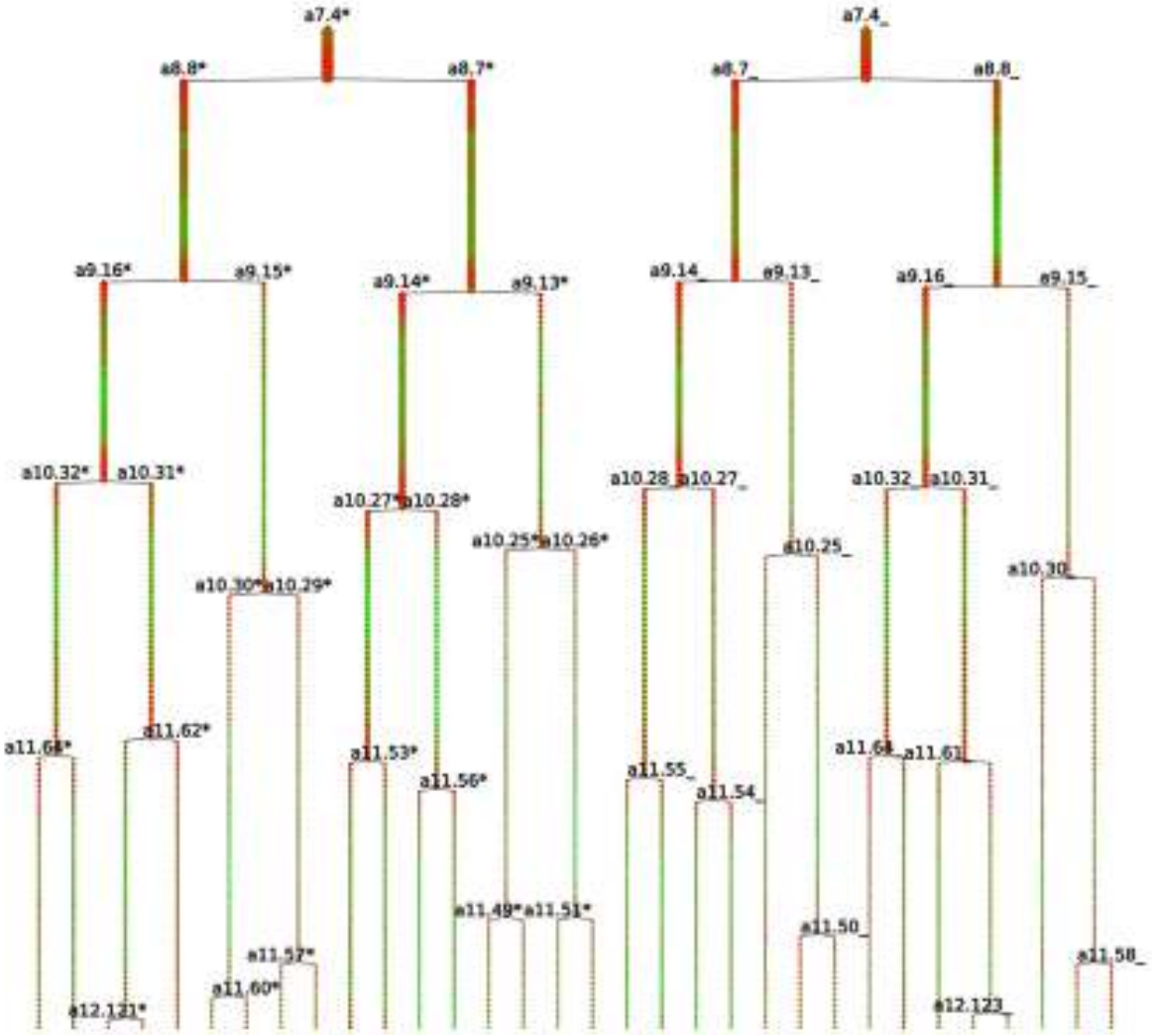


Figure 2.16 – Example of lineage tree from ASTEC - A7.4. The color represent the compactness of the cell, from green not compact to red compact. Note that the cells get more compact around the divisions. The size of the branches corresponds to the volume of the cell. Note that the divisions of A8.7 and A8.8 are assymetric in volume and that this assymetry is conserved bilaterally.

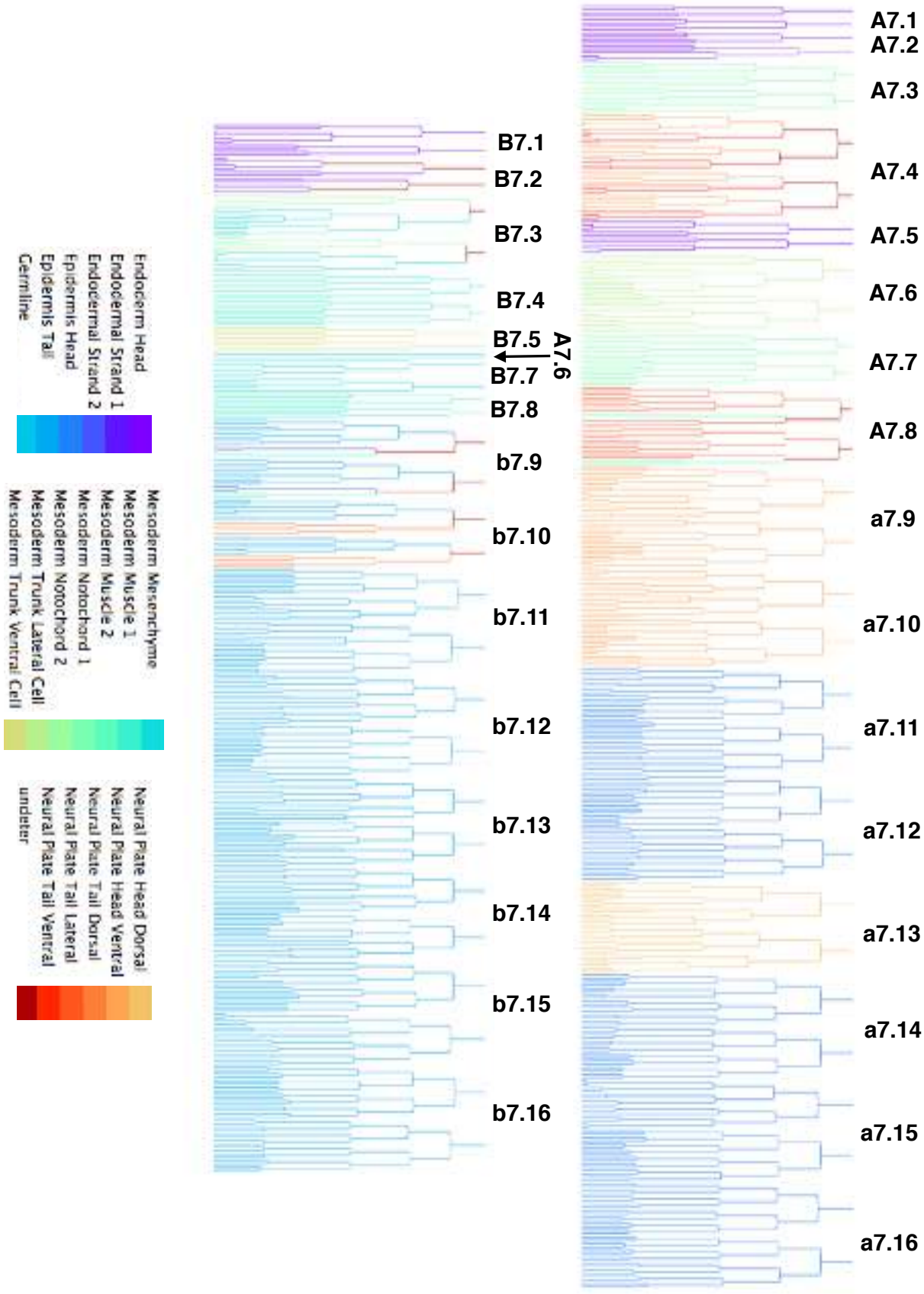


Figure 2.17 – ASTEC cell lineage tree with fate map.

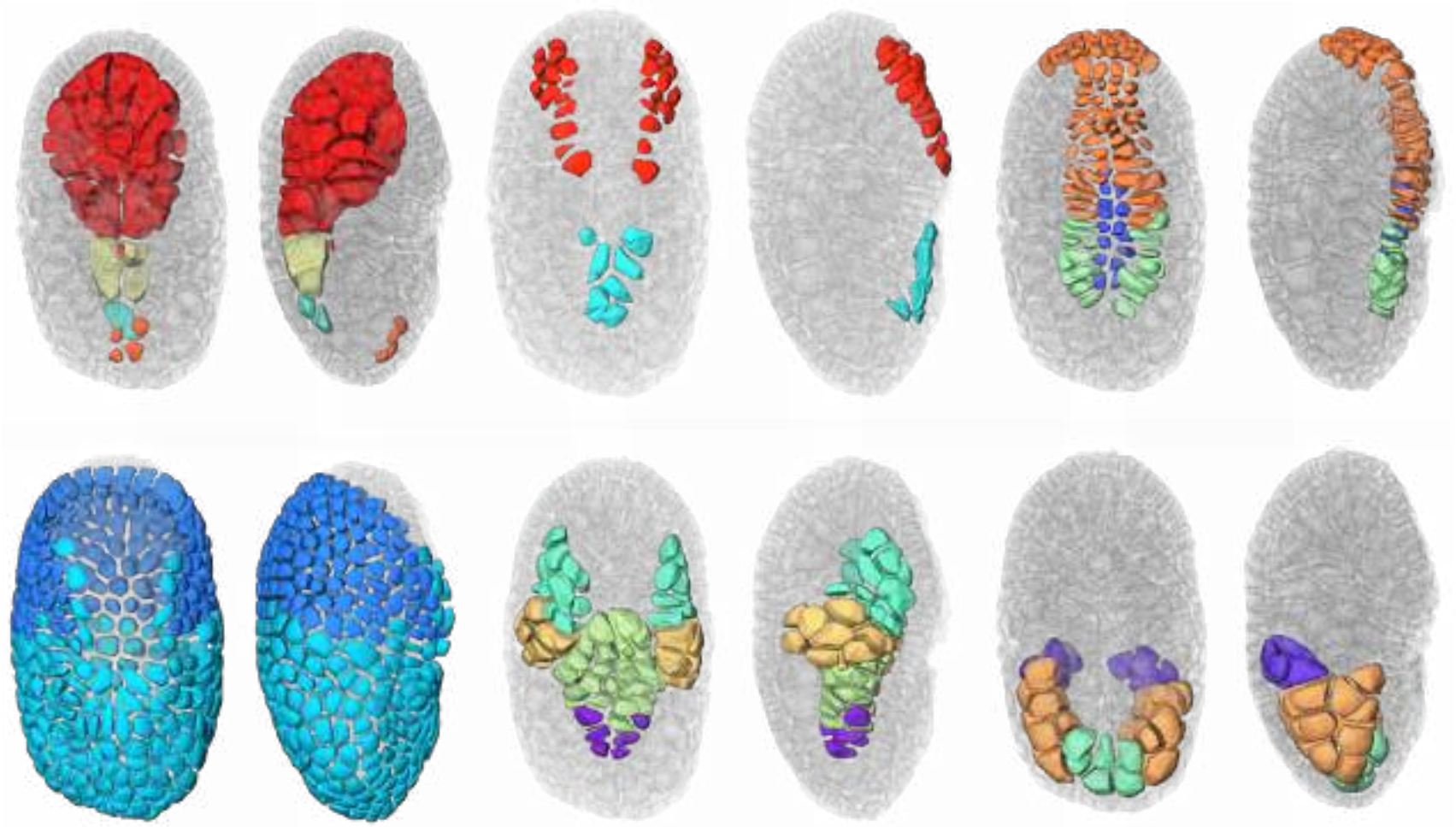


Figure 2.18 – 3D projection of the segmented embryo at the early tailbud stage. The cells are colored by tissue, the color code are as in Figure 2.17. The cells are slightly eroded to allow their distinction. The other cells of the embryo are in transparent grey. The doral and lateral sides are shown.

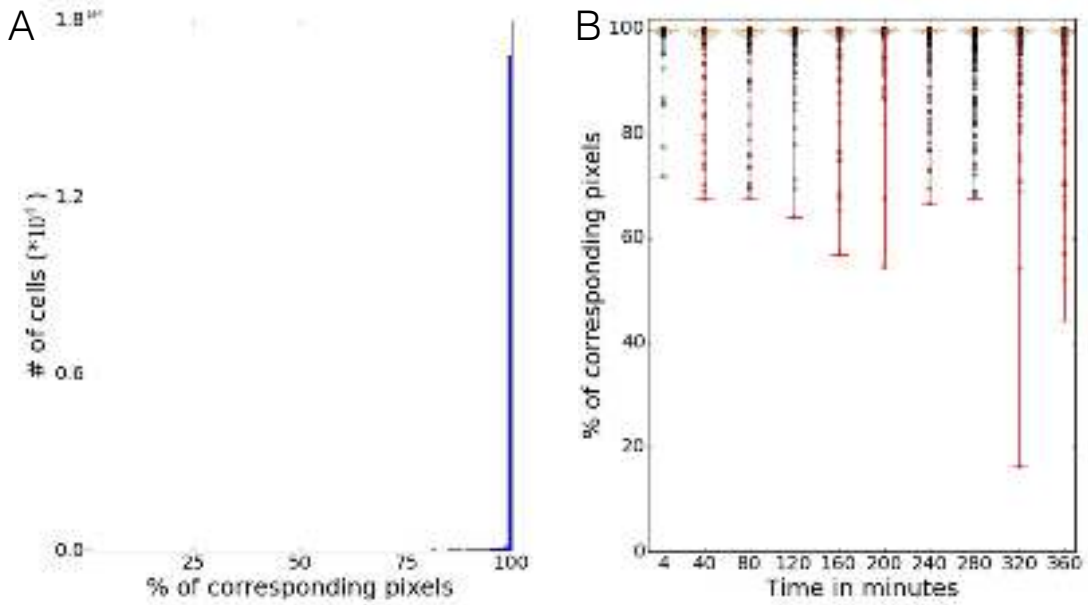


Figure 2.19 – ***Quantification of the quality of the shape of ASTEC-segmented cells.*** *A) Distribution of the percentages of pixels allocated to the same cell in automated and manually corrected segmentations (measured on a set of 150 2D sections through the embryo covering the whole duration of the movie). 100 bins are shown. B) Violin plot of the distribution of correctly allocated pixels, calculated on a set of 15 2D sections for each indicated time point.*

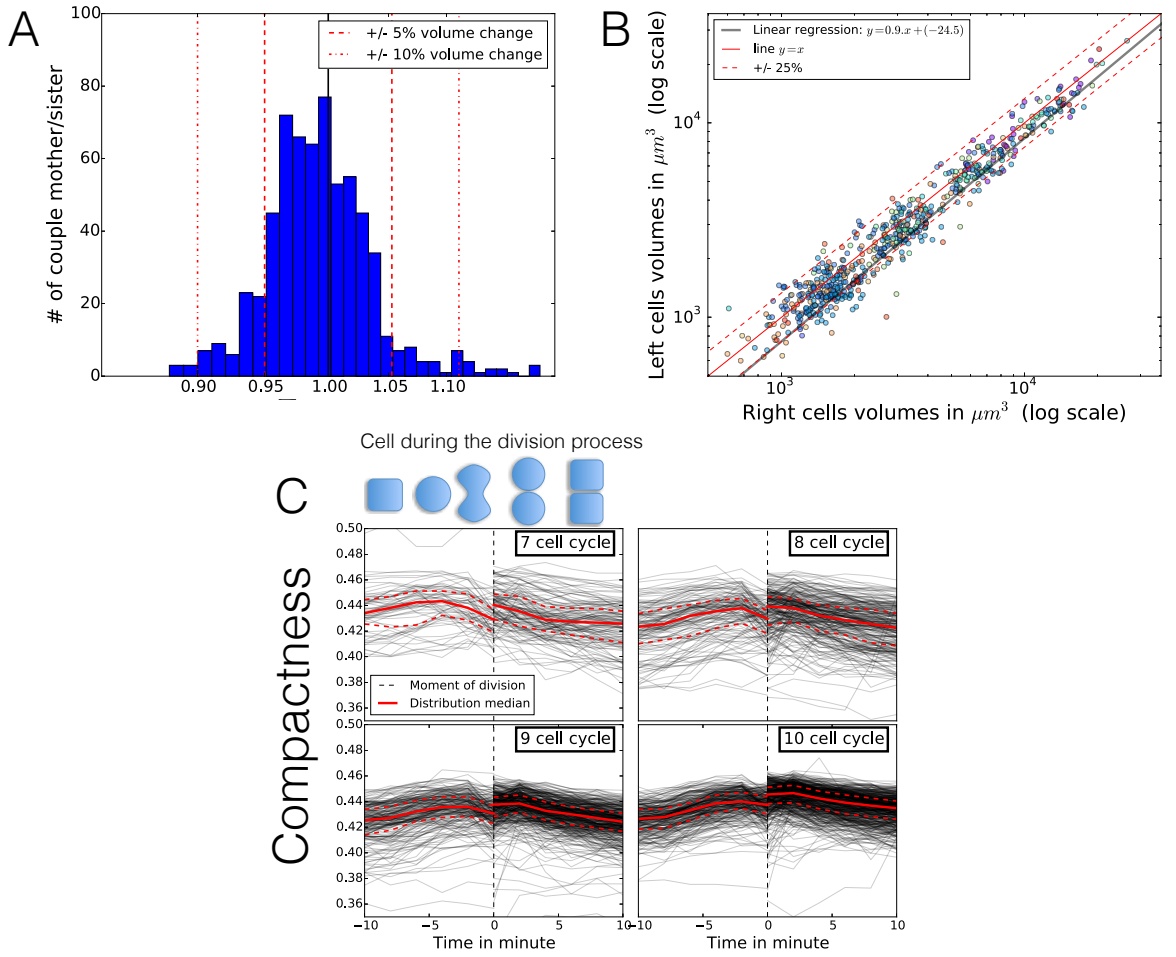


Figure 2.20 – ***Biological validation of the ASTEC segmentation.*** *A) The volume of the two daughter cells equals that of their mother. Distribution of the ratio of the sum of the volumes of the two daughter cells over that of their mother just before division. B) The volumes of matching right and left bilateral cells are highly correlated. Note that the right cells are systematically larger than the left cells, which may reflect a slight geometrical asymmetry of the first zygotic cell division. C) Evolution of the compactness of the cells around cell division events with underlying explanatory diagrammes. Cell divisions at the end of the indicated cell cycle were temporally aligned onto the vertical black hashed line. The red line represents the median of the distribution, the red hashed lines the bottom and top 25% of the distribution.*

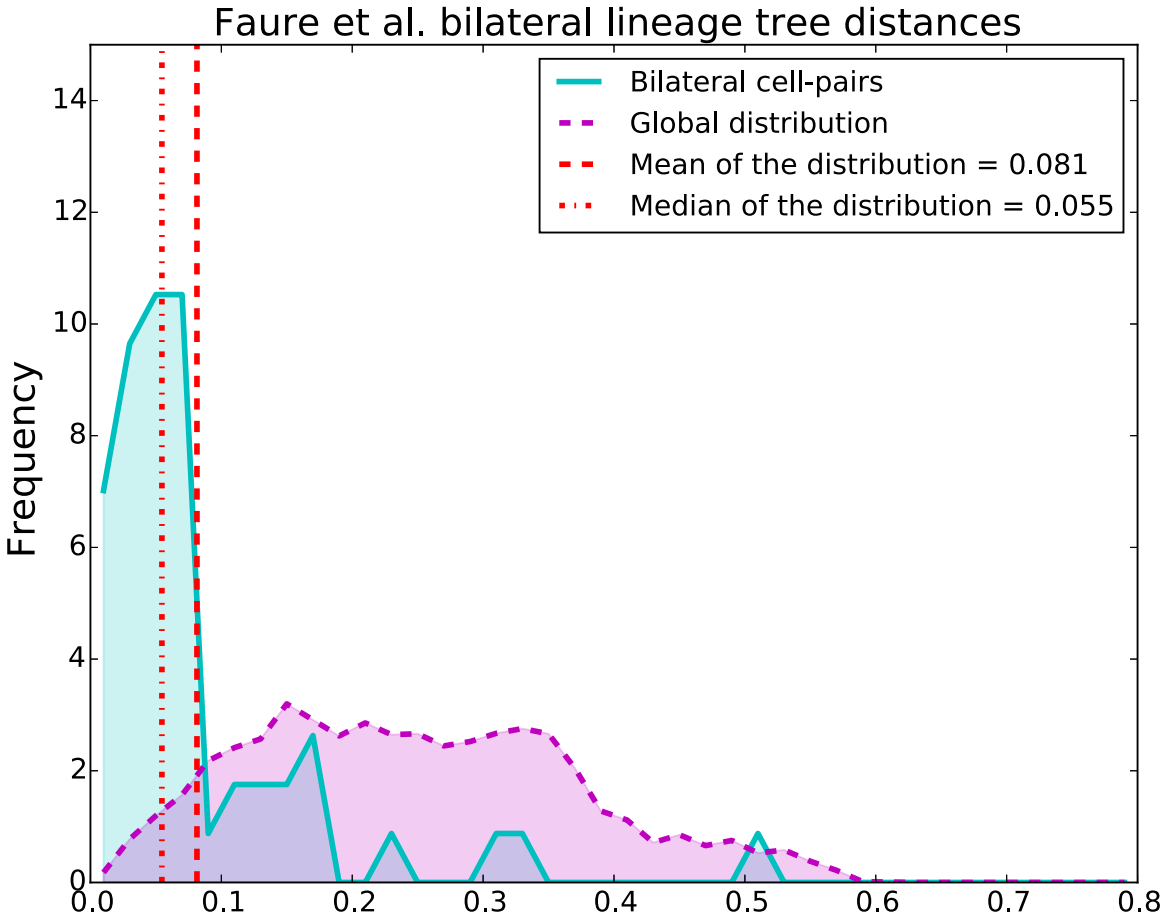


Figure 2.21 – ***Pairwise comparison of the cell lineages of bilateral cell pairs in an independently reconstructed Phallusia mammillata cell lineage.*** *Distribution of the distances between cell lineage trees from Faure et al. Magenta: pairwise comparison between all possible trees originating in the early gastrula stage. Cyan: pairwise comparison restricted to the bilateral cell-pairs of the early gastrula stage.*

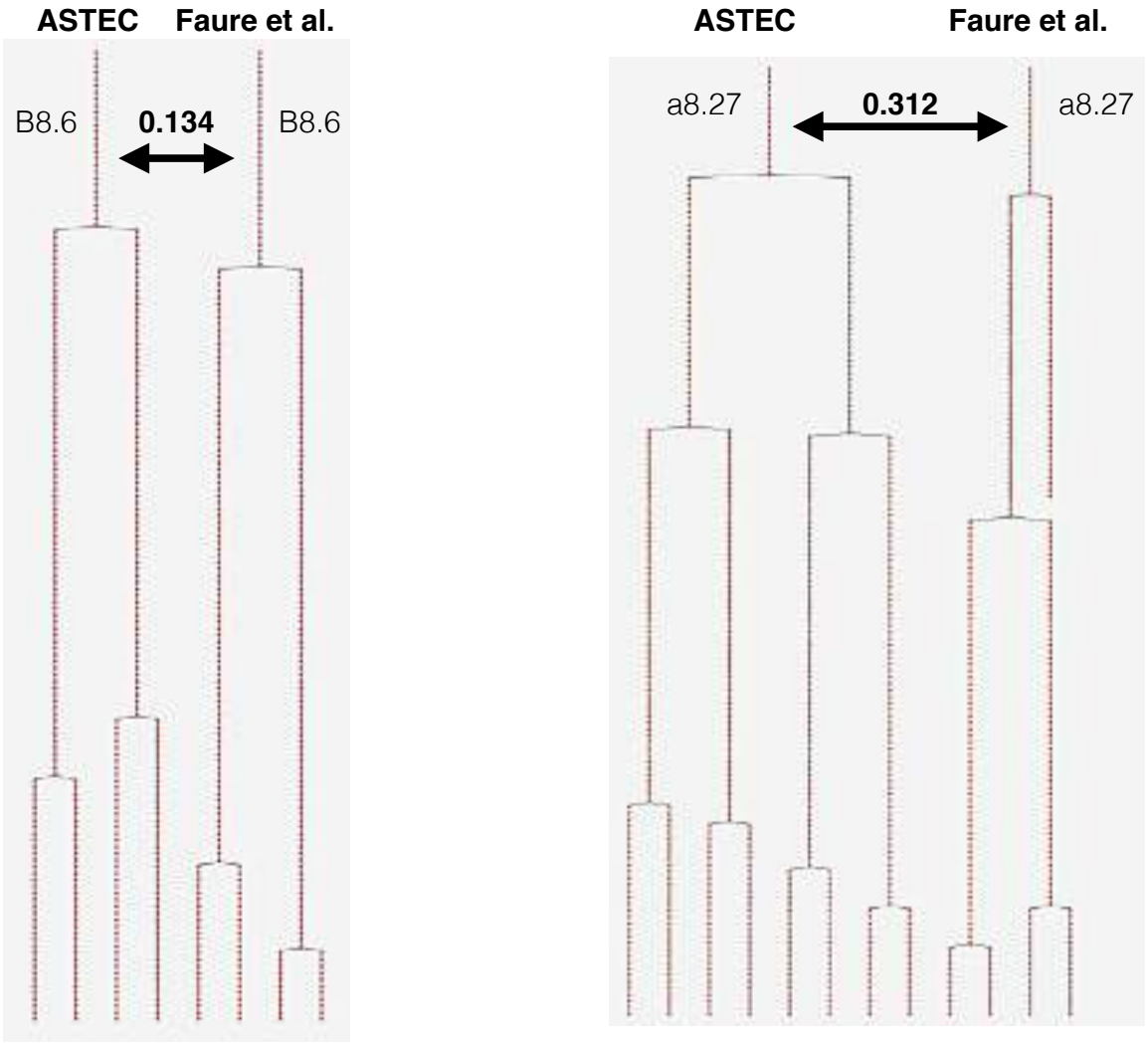


Figure 2.22 – ***Comparison of cell lineage trees between ASTEC and Faure et al.*** *Two couple of cell lineage trees are represented with their respective distance given by the tree compraison.*

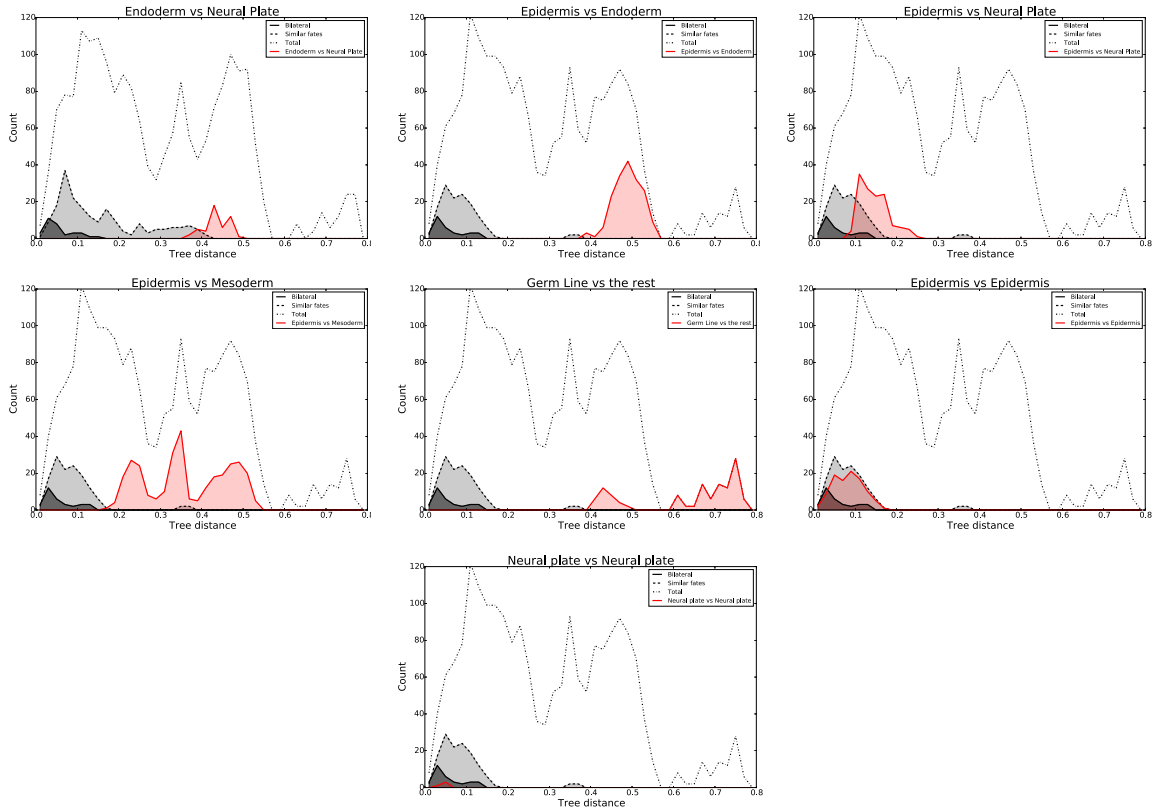


Figure 2.23 – Distribution of the distances between cell lineage trees. Hashed line: pairwise comparison between all the 64 possible trees originating in the 64-cell stage. Plain line: pairwise comparison restricted to the bilateral cell-pairs of the 64 cell-stage. Red line: pairwise comparison restricted to the mentioned cell fates of the 64 cell-stage.

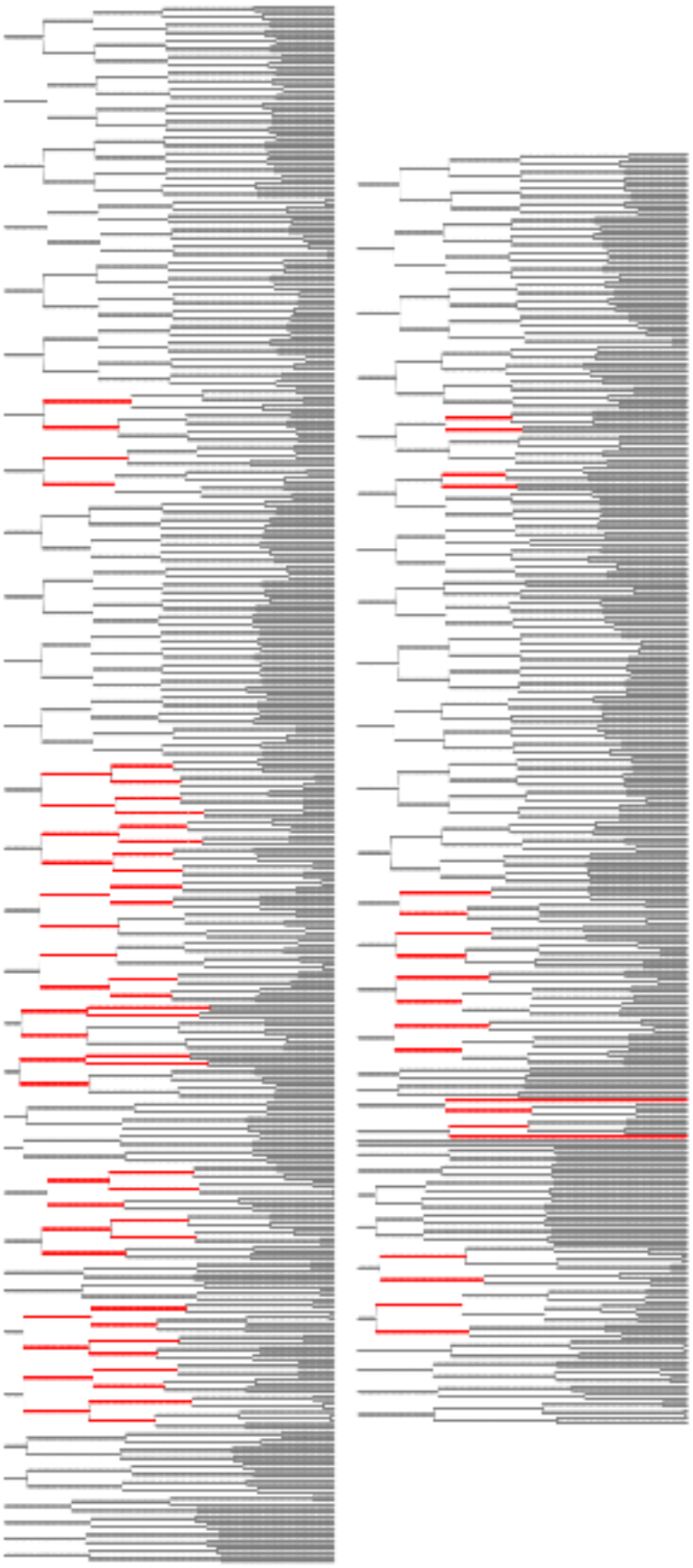


Figure 2.24 – Cell lineage trees from the 64 cell stage. In red: the events of cell fate decision suggested by the distance of sisters division patterns.

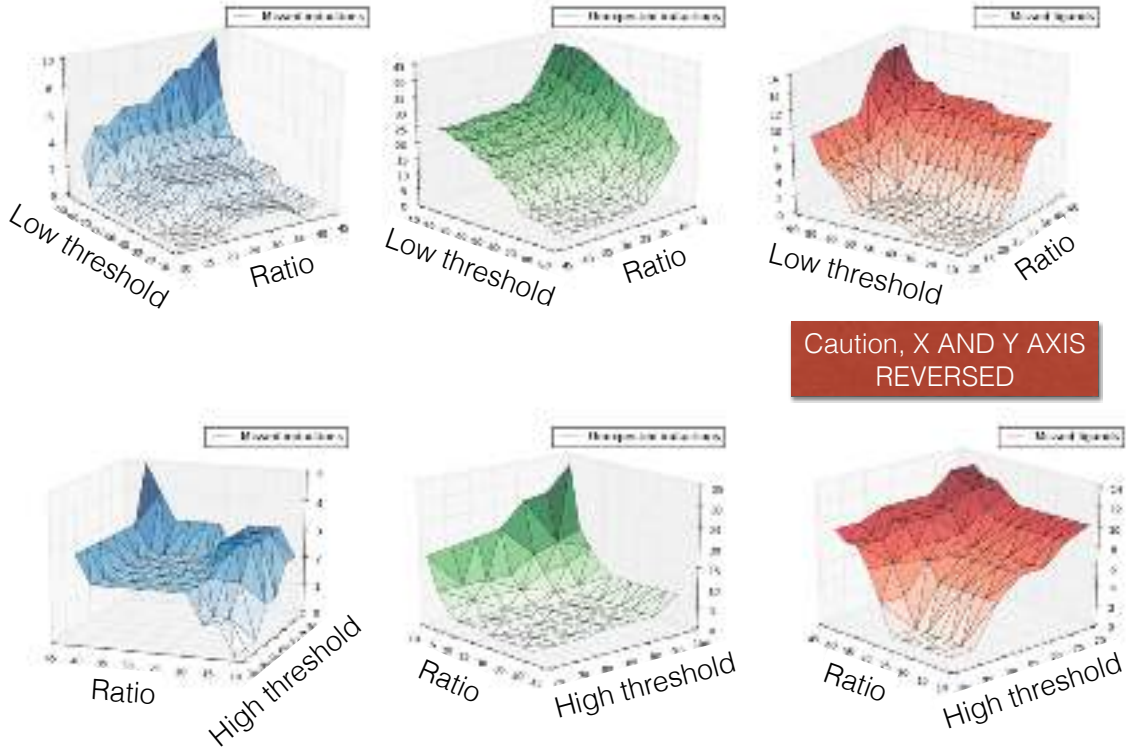


Figure 2.25 – ***Model parameter sensitivity study.*** *The model is trained to minimize the number of missed known inductions, missed knwon induction-ligands pairs and the number of cells found with inductions that were known not to specify at this time. The graphics represent the evolution of these numbers to the variation of the 3 parameters of the model (High/Low thresholds and the ratio threshold). Note that the set of parameters resulting in the minimization of the 3 previously mentioned values is small since the number of cells found with inductions that were known not to specify at this time increase when the number of missed known inductions decrease.*

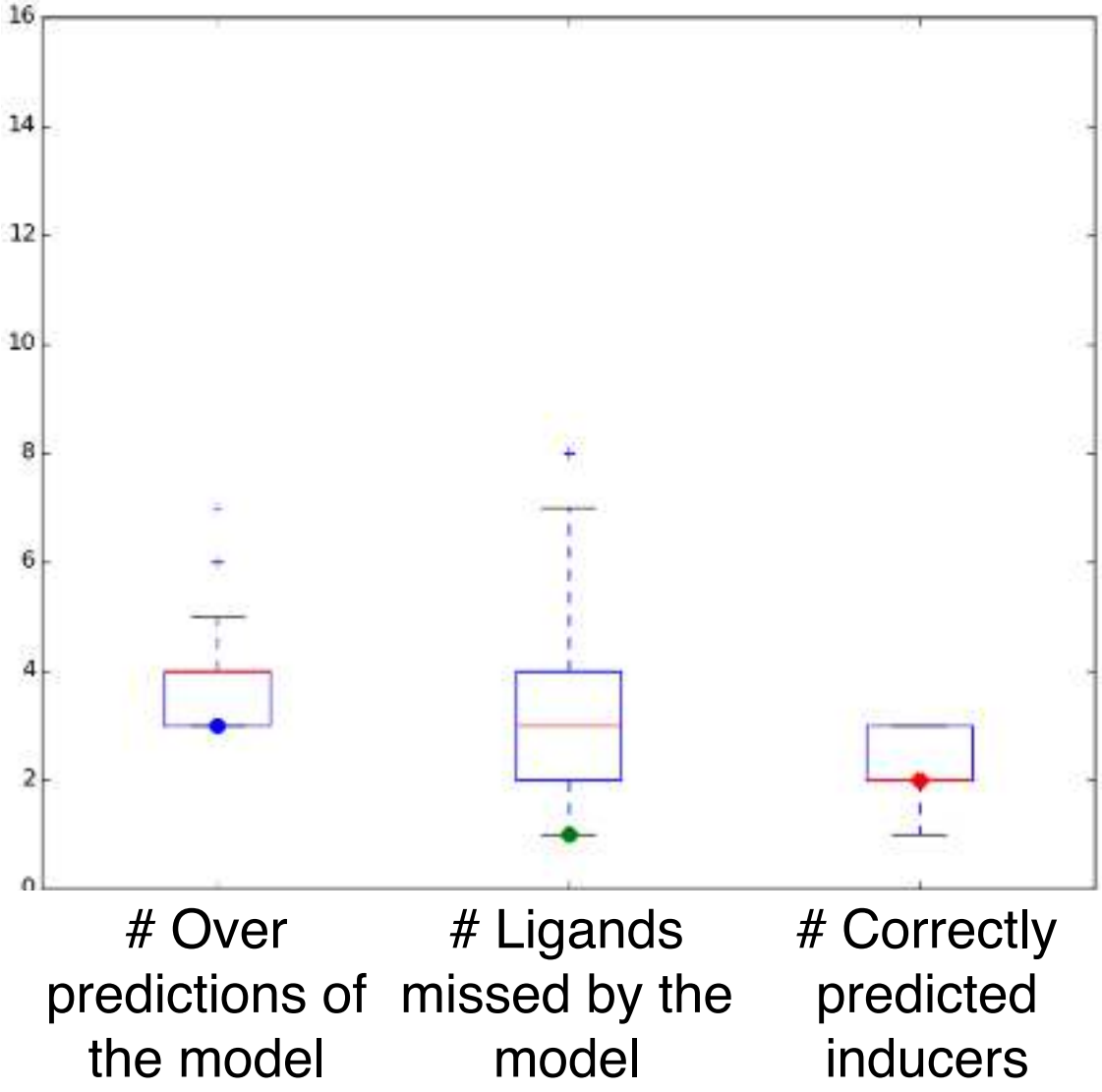


Figure 2.26 – ***Low robustness of the model to the noise.*** *The distributions shown represent the different results of the previously mentioned metrics to minimize with random gaussian additive noise added to the membranes. The order of magnitude of the noise added is similar to the noise found between bilateral cells. When additive noise is added to the surface of contacts an increase of the metric to minimize is usually observed. Only in few cases (less than 25%), the number of missed ligands is decreased. Thus, the induction events seem to be highly sensitive to the noise in the cell-cell area of contacts.*

Chapter 3

3D+t Sequence Registration

Contents

3.1 Introduction

The previous work exposed a tool to segment and track cells, ASTEC. It allowed to exhibit a marker of the different specification events, the division patterns. The model of the neighbouring cell interaction showed that the surface of contact between cells probably have a major impact in these specification events. The importance of the surface of contacts could explain in part the reason of the previously observed stereotyped development. To be able to understand better this stereotypy and the resulting implications, it is necessary to first quantify it and therefore record and segment more of these embryos.

To quantify the stereotypy (or variability) of development between and within species, segmented embryos have to be compared. The first step for this comparison is to align and register the 4D time-series against each others, in time and space.

This chapter contains two articles published respectively in the proceedings of ISBI, Beijing, China. IEEE. 2014 and in the proceedings of ISBI, New-York, United States of America. IEEE. 2015

3.1.1 Article presentation

In section 3.2, I present a method to perform spatio-temporal registration of time-series based on intensity images. This registration allows the comparison of the time-series. Once the intensity time-series are registered, it is possible to use the segmentation of one time-series to help the segmentation of the second.

Authors and contribution. In this section, I did the development of the spatio-temporal registration pipeline and wrote the manuscript. U.-M. Fiuza did all the embryo preparation and image acquisitions. L. Hufnagel helped with the image acquisition protocol and provided the MuVi-SPIM. C. Godin and P. Lemaire supervised parts of the work. G. Malandain supervised the work and wrote the manuscript. All authors contributed to the manuscript.

A method to spatially register two segmented images of different embryos from the same stage of development is presented in section 3.3. It is first assumed that it is trivial from the segmented data to temporally register two embryos (using the cell number for example). Then using the plane of symmetry of these embryos, an spatial registration is done. This method yields a spatial registration and help to build the 4D average embryo of *Phallusia mammillata* and enable embryos comparison.

Author and contribution. In this section, G. Michelin did the development of the spatial registration pipeline and wrote the manuscript. I helped develop the pipeline and performed the segmentations. U.-M. Fiuza did all the embryo preparation and image

acquisition. G. Malandain supervised the work and wrote the manuscript. All authors contributed to the manuscript.

3.2 Spatio-temporal registration of embryo images

Guignard L.[1,2], Godin C.[2], Fiuza U.-M[1], Hufnagel L.[3], Lemaire P.[2], Malandain G.[4]

[1] CRBM, UMR5237, CNRS-U. Montpellier, 1919 Route de Mende Montpellier Cedex 5, France
[2] Inria project-team Virtual Plants, joint with CIRAD and INRA, Campus St Priest - BAT 5, 860 rue de St Priest, F-34095 Montpellier Cedex 5, France
[3] EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany
[4] Inria project-team Morpheme, 2000 route des lucioles, les algorithmes - CS 40121 Sophia Antipolis cedex F-06903, France
Contact: {leo.guignard,gregoire.malandain}@inria.fr

Abstract Current imaging techniques can capture temporal sequences of 3D images with very high time resolution over several hours. Comparing sequences covering the same time period opens the way to the study of developmental variability. Stitching together sequences captured from different embryos may help producing a sequence covering the whole development of the animal of interest. For this, it is necessary to align two sequences in both time and space.

We present here a method to align two 3D+t time series, based on the detection and pairing of 3D+t landmarks. These landmarks, which correspond to periods of fast morphogenetic change, are deduced from the analysis of the non-linear transformations that allow to co-register pairs of consecutive 3D images in each sequence.

3.2.1 Introduction

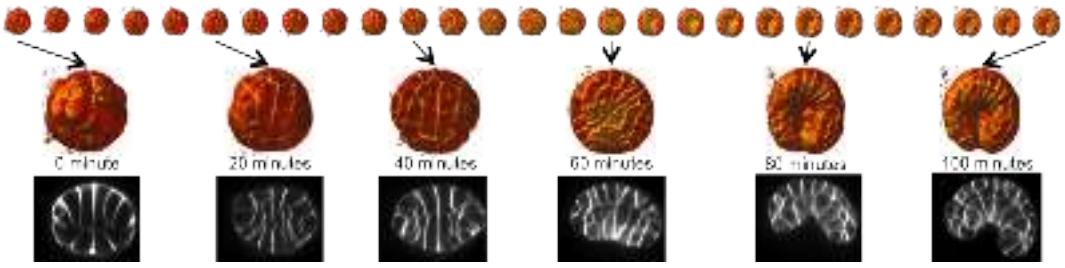


Figure 3.1 – Phallusia mammillata embryo from 32 *cells stage to* 112 *cells stage, 100 minutes later. The first row shows the 3D renderings every 4 minutes, while the second and third rows display respectively some magnified views and corresponding cross-sections.*

Animal and plant morphogenesis is a highly dynamic process spanning several temporal and spatial scales. One challenge is to describe cell and/or embryo shapes through development, to analyze their dynamics and variability within and between species and

to correlate this information with gene activity [Munro et al., 2006; Traas and Hamant, 2009]. For this, it is necessary to compare the development of several embryos in order to identify similarities and variations in morphogenetic events.

Recent progress in light microscopy allows the capture of 3D images of entire live embryos at a sub-cellular level of resolution with a high rate of acquisition and without interfering with development [Keller, 2013]. These protocols produce 3D+t images depicting the development. Comparing or fusing such sequences captured from different embryos is a challenging task, but of crucial importance. While it is generally not possible to image an embryo throughout its whole development, one can generate developmental sequences from different individuals, with partial temporal overlap. Stitching together such time series can be used to to produce a consensual representation of the whole development, and would also allow to identify variations in this programs.

There are however two main challenges. Firstly, it is experimentally difficult or impossible to orient the embryos in a standardized manner during imaging. Moreover embryos often progressively drift in the field of view during long time-lapse sessions. It is thus necessary to realign spatially each sequence so that the embryos are in the same orientation, and thus can be easily compared. Secondly, one needs to temporally align these time series over their periods of overlap. Because of the very large size of such datasets (up to several hundred 3D images per sequence), alignment methods have to be as automated as possible.

We propose here a landmark-based registration method for two 3D+t time series. These landmarks are extracted from the analysis of the sequences. More precisely, in a given time series, every pair of consecutive 3D images are co-registered with a non-linear transformation, and the largest regional deformations are extracted from the whole sequences, yielding 3D+t landmarks. Alignment of two independent sequences is then achieved by finding the optimal pairings between two sets of 3D+t landmarks.

3.2.2 Data description

We used the simple marine invertebrate chordate *Phallusia mammillata* to test and develop our algorithms. Embryonic development of this species is highly similar across individuals, because the timing and orientation of cell divisions are highly stereotyped. As a consequence all cells can be unambiguously named and recognized up to the 112-cell stage, when the first major morphogenetic process, gastrulation is initiated.

We imaged live embryos in using a lightsheet microscope (MuViSPIM, EMBL, Heidelberg [Krzic et al., 2012]). To image membranes, embryos were soaked in a lipophilic dye (FM4-64). The dye was excited with a 594nm laser (Cobolt AB) and the emitted light was collected by two opposing 25x Nikon water dipping objective lens (NA 1.1) and after passing thought a long-pass filter (BLP01-594R-25, Semrock), was collected with a Hamamatsu Flash 4 SCMOS camera. At each time point, two perpendicular 3D image stacks, each with 1200×1200 pixels per plane and a total of 200 planes were acquired by both cameras. This resulted in a total of four views (0, 90, 180, 270 degrees) of the specimen each with a lateral resolution of $0.26\mu m \times 0.26\mu m$ and $1\mu m$ plane spacing. This procedure was repeated every two minutes. To obtain an 3D image with isotropic resolution, the 4 image stacks were fused into a single 3D dataset as described

in [Fernandez et al., 2010].

For the purpose of this paper we will use time-series from two different embryos, each counting 50 time points (1h40 min). Both time series start at the the 32-cell stage and end after the 112-cell stage. The imaging temperatures were different (20°C and 18°C) so that both embryos exhibit slightly different developmental kinetics. Fig 3.1 depicts 2D sections of some 3D images at different time-points extracted from one sequence as well as the corresponding 3D renderings.

3.2.3 Intra-sequence registration

3.2.4 Registration method

Registration is a frequently addressed issue in the literature [Zitová and Flusser, 2003; Sotiras et al., 2013]. Although the choice of a particular registration method may not be critical, it is important that the chosen method would be robust with respect to topological changes. For that reason, we chose to use a block matching scheme [Ourselin et al., 2000] to compute either linear or non-linear transformations. Such a scheme is comparable to the ICP method [Besl and McKay, 1992] except that iconic primitives are matched instead of points.

Registration aims at the computation of the transformation $T_{F\leftarrow R}$ that will allow to resample a floating image I_F onto a reference image I_R. The transformation $T_{F\leftarrow R}$ is iteratively computed by integrating incremental transformations δT^i, ie $T^{i+1}_{F\leftarrow R} = \delta T^i \circ T^i_{F\leftarrow R}$.

At iteration i, blocks (or sub-images) b_R of the reference image I_R are compared to blocks b_F of the floating image I_F, the best block pairing (the one that yields the best iconic measure, here the normalized correlation) yield a point pairing, (c_R, c_F), by associating the block centers. The incremental transformation is then estimated by

$$\delta T^i = \arg\min_{\delta T} \sum \left\| c_F - \delta T \circ T^i_{F\leftarrow R} c_R \right\|^2 \tag{3.1}$$

Linear transformations are computed with a Least Trimmed Squares [Rousseeuw and Leroy, 1987a] method that allows to discard outlier pairings.

The non-linear transformation is represented by a dense vector field. The transformation update follows then the following steps: 1. the sparse pairing field (c_R, c_F) is transformed into a dense field by Gaussian interpolation (this Gaussian interpolation also acts as a fluid regularization); 2. outlier pairings are discarded; 3. the remaining pairings are transformed into a dense field by Gaussian interpolation; 4. this dense field is composed with the transformation found at the previous iteration; 5. the resulting transformation is smoothed by a second Gaussian filter that acts as an elastic regularization.

3.2.4.1 Pairwise registration

Consider a temporal sequence of 3D images, $\{I_t\}$ for $t = 1 \dots N$. We aim at co-registering a pair of images captured at two different time points (i.e. I_t and $I_{t+\delta t}$).

Since a δt of 2 minutes does not capture deformation of sufficient magnitudes choose $\delta t = 4$ minutes. We want to be sure to capture sufficient deformations to identify cell division events.

We first co-register every pair of successive images of the time series with a rigid transformation representing (and then allowing to compensate for) the small displacements of the embryo that may occur during the acquisition. Hence, we get a series of rigid transformations $R_{t \leftarrow t+\delta t}$ allowing to resample I_t onto $I_{t+\delta t}$.

To compensate for the displacements along the whole sequence, a reference image is chosen (here I_N), and all the transformations $R_{t \leftarrow N}$ are calculated by transformation composition, which allows to resample all I_t images into I_N frame

$$I_{N \leftarrow t} = I_t \circ R_{t \leftarrow N} = I_t \circ R_{t \leftarrow t+\delta t} \circ \ldots \circ R_{N-\delta t \leftarrow N} \tag{3.2}$$

From now on, it is assumed that the sequence is compensated for the small rigid displacements (ie $I_{N \leftarrow t}$ is now denoted I_t).

The remaining deformations that exists between two successive images can be captured thanks to a non-linear registration step. We then compute the non-linear transformations $T_{t+\delta t \leftarrow t}$ thereby resampling $I_{t+\delta t}$ in the same frame as I_t.

The non-linear transformation T is encoded by a vector field $\mathbf{v}$ such that $T(m) = m + \mathbf{v}(m)$, the modulus of the vector $\mathbf{v}(m)$ indicating the local amount of variation at point m. Figure 3.2 depicts such a vector field calculated from the registration of 3D images at two successive time-points: it demonstrates that the deformations are regionally localized, suggesting that these areas are of interest from a morphogenesis point of view.

3.2.4.2 Registration assessment

To assess the registration quality, we used a distance between the two images based on a comparison of membranes location. For this, we take advantage of the dye that makes the membranes hyper-intense and extract them by thresholding. More precisely, 10% on average of the image represents membranes, so the membrane set $\mathcal{M}(I)$ of image I is defined by the binary image resulting of the thresholding of I by the intensity representing the 90^{th} percentile of the cumulative histogram. Although not perfect, such a membrane definition is sufficient for a rough registration assessment.

Let I_F and I_R be respectively a floating and a reference image, $T_{F \leftarrow R}$ the non-linear transformation from I_R to I_F, $I_{F \rightarrow R}$ the floating image resampled onto the reference one ($I_{F \rightarrow R} = I_F \circ T_{F \leftarrow R}$), and $\mathcal{M}_R$ and $\mathcal{M}_{F \rightarrow R}$ the membrane sets of respectively I_R and $I_{F \rightarrow R}$. We define a registration score S by measuring the average distance from one membrane set to the other thanks to

$$S\left(I_{F \rightarrow R}, I_R\right) = \frac{1}{|\mathcal{M}_R|} \sum_{m \in \mathcal{M}_R} \min_{m' \in \mathcal{M}_{F \rightarrow R}} \|mm'\| \tag{3.3}$$

Figure 3.3 shows the evolution of this score from rigid registration, ie $S\left(I_{t+\delta t}, I_t\right)$, where the small rigid displacements have already been compensated for, to non-linear registration, ie $S\left(I_{t+\delta t \rightarrow t}, I_t\right)$.

Figure 3.2 – A cellular division occurs between 16 minutes (rendering at left) and 20 minutes (rendering at right) for the embryo of Fig. 3.1. The non-linear registration that co-registers both images exhibits some high deformations, represented by large vectors (at the middle; only vector with a norm greater than 7.5μm are displayed).

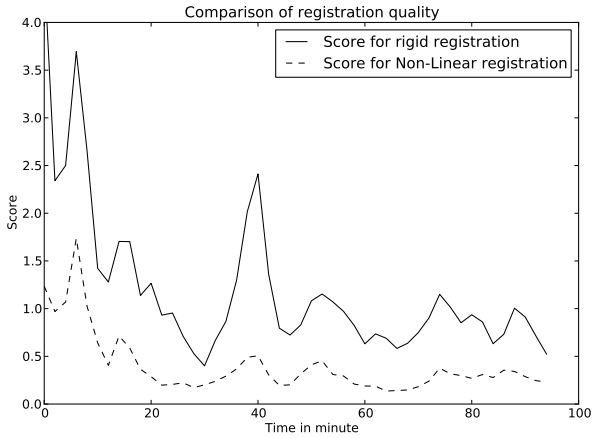


Figure 3.3 – Registration score over time for both rigid and non-linear registration.

3.2.5 Alignement of independent 3D+t time series

3.2.5.1 Definition of 3D+t landmarks

The development of the embryo is punctuated by regional deformations at precise steps. For example, during the cell division process the dividing cells round up, which causes local deformation. Also, during gastrulation part of the embryo invaginates, a process that creates regional deformations. We use these spatially and temporally localized events as 3D+t landmarks to guide the co-registration of two morphogenesis sequences.

We then propose to use the deformations extrema as 3D+t landmarks. Let us denote the set of landmarks to be built by $\mathcal{L} = \{L_i\}$. A 3D+t landmark L_i is defined by a time t_i and a spatial position x_i. Let us consider a temporal sequence of 3D images, $\{I_t\}$ where the small rigid displacements have been compensated for, together with the non-linear transformations $T_{t+\delta t \leftarrow t}$.

We start by detecting the temporal extrema of the deformations. We compute the

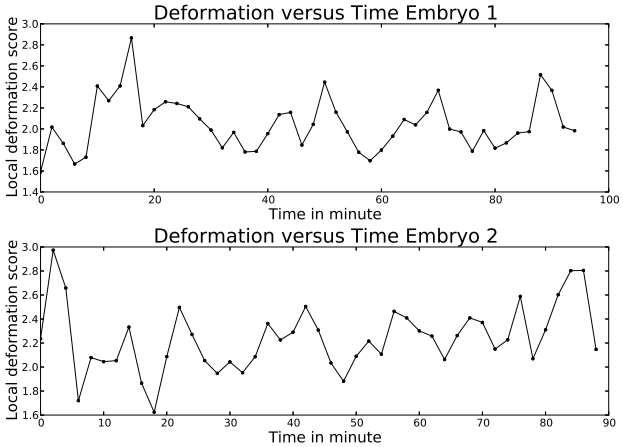


Figure 3.4 – Score of local deformation over time for the two embryos to be co-registered.

cumulative histogram of the vector modulus $\|\mathbf{v}(m)\|$ (recall that the non-linear transformation is encoded by a vector field) for the membrane points $m \in \mathcal{M}(I_t)$, calculate its 90^{th} percentile, and propose as deformation score the ratio between the mean value of the vector modulus computed on the 10% membrane points having the largest vector modulus (set denoted by $\mathcal{M}_{10\uparrow}$), and the mean value of the vector modulus computed on the 90% membrane points having the smallest vector modulus (set denoted by $\mathcal{M}_{90\downarrow}$).

$$D(T_{t+\delta t \leftarrow t}) = \frac{\sum_{m \in \mathcal{M}_{10\uparrow}(I_t)} \|v(m)\|}{|\mathcal{M}_{10\uparrow}(I_t)|} \left(\frac{\sum_{m \in \mathcal{M}_{90\downarrow}(I_t)} \|v(m)\|}{|\mathcal{M}_{90\downarrow}(I_t)|} \right)^{-1} \tag{3.4}$$

This deformation score exhibits some extrema with respect to time (Fig. 3.4), which correspond to the time positions of our 3D+t landmarks, denoted by $t_i = t(L_i)$. From now on, we assume that the landmarks are temporally ordered, ie $i < i' \Leftrightarrow t_i < t_{i'}$.

For each of these particular time points, we now compute the connected components of $\mathcal{M}_{10\uparrow}$, retain the largest of them (this somehow corresponds at extracting the largest, or extremal, deformation), and compute the center of mass of this largest connected component: this yields the spatial positions of our 3D+t landmarks, denoted by $x_i = x(L_i)$.

3.2.5.2 Sequences co-registration

Let us consider two morphogenesis time series $\{I_t\}$ and $\{I'_{t'}\}$ and their associated landmarks $\mathcal{L} = \{L_i\}$ and $\mathcal{L}' = \{L'_j\}$. To register the two time series, we build and test pairing hypothesis between $\mathcal{L}$ and $\mathcal{L}'$, this pairing hypothesis is included in the powerset $\mathcal{P}(\mathcal{L} \times \mathcal{L}')$.

A pairing hypothesis $H = \{(i_H, j_H)\}$ consists of a set of couples representing the landmark pairing $\{(L_{i_H}, L_{j_H})\}$. Pairing hypothesis are not all tested, and we retain only those obeying: 1) at least 5 landmarks have to be effectively paired; 2) pairings are

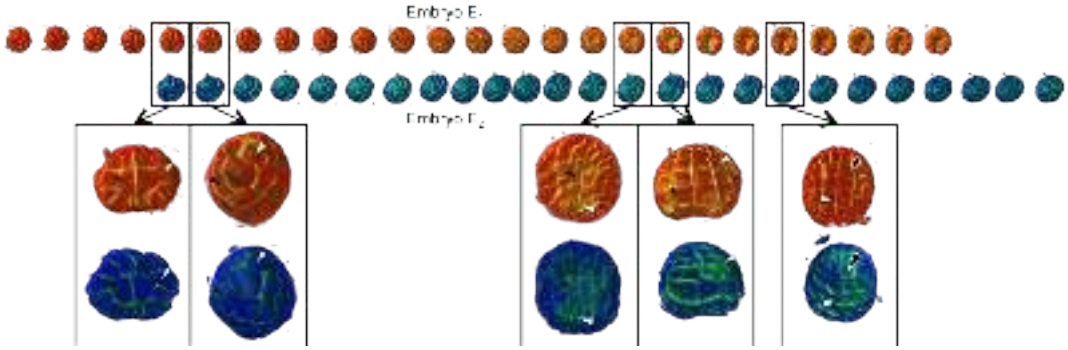


Figure 3.5 – Spatio-temporal registration of two time-series of embryo. Enlarged renderings indicate the registered times $\{(t_{i_{\hat{H}}}, t'_{j_{\hat{H}}})\}$. Notice that the temporal registration is not linear since the interval length between two registered time is different from one embryo to the next.

temporally consistent, ie for two pairings (i_H, j_H) and (i'_H, j'_H), if $i_H < i'_H$ then $j_H < j'_H$ (recall that the landmark indexes are temporally ordered). Given a pairing hypothesis, the transformation registering the paired landmarks can be easily estimated in the least squares sense :

$$T_H = \arg\min_T \sum_{\{(i_H, j_H)\}} \left\| T(x_{i_H}) - x'_{j_H} \right\|^2 \tag{3.5}$$

The mean of the residuals measures the hypothesis registration quality, allowing to define the best pairing hypothesis, $\hat{H}$, by

$$\hat{H} = \arg\min_H \frac{1}{|H|} \sum_{\{(i_H, j_H)\}} \left\| T_H(x_{i_H}) - x'_{j_H} \right\|^2 \tag{3.6}$$

yielding not only a spatial co-registration of both sequences thanks to $T_{\hat{H}}$ but also a temporal co-registration thanks to the couples $\{(t_{i_{\hat{H}}}, t'_{j_{\hat{H}}})\}$. This temporal co-registration is sparse, since only a few time-points are co-registered. A dense temporal co-registration (where all time-points are co-registered) can easily be induced by simple means (e.g. linear interpolation) although more sophisticated schemes are under investigation.

Last, the spatial transformation between the two sequences is refined twice. First, we re-estimate $T_{\hat{H}}$ with a Least Trimmed Squares optimization (which discards the worse pairings out of $\{(i_{\hat{H}}, j_{\hat{H}})\}$). Second, we resample the images $I'_{j_{\hat{H}}}$ using $T_{\hat{H}}$ and co-register the resulting images $I'_{j_{\hat{H}}} \circ T_{\hat{H}}$ with the images $I_{i_{\hat{H}}}$. It yields a number of transformations $T_{(I'_{j_{\hat{H}}} \circ T_{\hat{H}}) \leftarrow I_{i_{\hat{H}}}}$. Averaging them allows to get the spatial transformation $T_{\{I'_{t'}\} \leftarrow \{I_t\}}$ that aligns the sequence $\{I'_{t'}\}$ onto $\{I_t\}$ (see Fig. 3.5).

$$T_{\{I'_{t'}\} \leftarrow \{I_t\}} = T_{\hat{H}} \circ \left(\frac{1}{|\hat{H}|} \sum_{\{(i_{\hat{H}}, j_{\hat{H}})\}} T_{(I'_{j_{\hat{H}}} \circ T_{\hat{H}}) \leftarrow I_{i_{\hat{H}}}} \right) \tag{3.7}$$

A visual inspection of the registered sequences demonstrate both the temporal re-synchronization of both embryos as well as the superimposition of the corresponding cells (that can named in the first embryogenesis stages).

3.2.6 Discussion

We proposed a method to spatially and temporally align two time series of embryos of *Phallusia mammillata*, with a linear spatial transformation and a non-linear temporal one. It relies on the fact that morphogenesis events, as cellular divisions, induce temporally and spatially localized deformations. The detection of the extremal deformations is used to build sets of 4D landmarks. Assuming that the development of different embryos of the same species is similar, landmarks identified in different embryos can be paired to eventually find the 4 dimension transformation. Our work validates this assumption in the case of the stereotyped embryos of the ascidian *Phallusia mammillata*.

By providing a spatio-temporal registration of 3D+t time series of images based on membrane deformation, our method opens the way to the geometrical quantification of morphogenesis at a cellular resolution. Finally, from this method we will be able to build longer time series based on fusion of several overlapping smaller time series.

3.3 Cell pairings for ascidian embryo registration

Michelin G.[1], **Guignard L.**[1,2], **Fiuza U.-M.**[2], **Maladain G.**[1]

[1] Inria project-team Morpheme, 2000 route des lucioles, les algorithmes - CS 40121 Sophia Antipolis cedex F-06903, France
[2] Inria project-team Virtual Plants, joint with CIRAD and INRA, Campus St Priest - BAT 5, 860 rue de St Priest, F-34095 Montpellier Cedex
Contact: {gael.michelin,gregoire.malandain}@inria.fr

Abstract Recent microscopy techniques allow imaging temporal 3D stacks of developing organs or embryos with a cellular level of resolution and with a sufficient acquisition frequency to accurately track cell lineages.

Imaging multiple organs or embryos in different experimental conditions may help decipher the impact of genetic backgrounds and environmental inputs on the developmental program. For this, we need to precisely compare distinct individuals and to compute population statistics. The first step of this procedure is to develop methods to register individuals.

From a previous work of cell segmentation from microscopy images, we here demonstrate how to extract the symmetry plane of embryos at early stages, and how to use this information as a geometrical constraint to both register these embryos and obtain a cell-to-cell mapping.

3.3.1 Introduction

A central aim in developmental biology is to better understand how each tissue of an embryo progressively acquires its functional shape, a process called morphogenesis. Image-based studies therefore represent a method of choice. Current live microscopy techniques allow the acquisition of temporal sequences of 3D images with a spatio-temporal resolution high enough to follow embryo development at sub-cellular scale [Keller, 2013]. An automatic framework to register individual cells from distinct developing embryos would allow quantifying the variability in embryo development at the cellular level, which is a major issue in morphogenesis studies.

The present work describes a complete framework to register two embryos at similar developmental stages and to provide cell-to-cell mapping. The proposed framework takes advantage of a cell segmentation method based on differential and structural information (section 3.3.2) to derive a novel symmetry plane extractor (section 3.3.3), and cell-to-cell mapping registration method (section 3.3.4). Its efficiency is demonstrated on embryo microscopy images of the simple marine invertebrate chordate Phallusia mammillata.

3.3.2 Cell segmentation framework

Embryo image segmentation is performed with the method proposed in [Michelin et al., 2014]. We recall here its main steps (Fig. 3.6):

1. Planar response filter using 1st and 2nd order derivatives, see Fig. 3.6(b),
2. Extrema extraction and binarization, see Fig. 3.6(c),
3. Gap filling by Tensor Voting framework, see Fig. 3.6(d), and
4. Cell detection and segmentation by seeded watershed approach, see Fig. 3.6(e)-(f).

The planar response filter, inspired from [Krissian et al., 2000], uses the property that a membrane has homogeneous intensity locally in the plane tangential to the membrane whereas the intensity varies strongly in its normal direction. Therefore, the Hessian matrix eigenvector, denoted $\mathbf{n}$ in the next section, associated to the largest eigenvalue in absolute value indicates the normal direction to the membrane, while the two other eigenvectors are tangential to the membrane. Using 1st order derivatives of the image, an edge response at a constant distance to the voxel is then integrated on either side in the normal direction to the membrane. This yields the planar response filter.

Then, the maxima of the planar response filter with respect to the normal direction to the membrane, are extracted, so that most of the irrelevant information is removed. The resulting image is finally thresholded in order to segment membrane voxels. This last image will be denoted B in next section.

The unavoidable gaps are filled by the application of the tensor voting framework [Medioni et al., 2000]. The binarised voxels are encoded as 2nd order tensors that describe local shape of the data, using the voxel orientation information obtained from Hessian matrix. The voting process consists, for each voxel, in propagating its local shape in its neighborhood. The accumulation of votes results in a new tensor map in which each tensor encodes structural information. A scalar surfaceness map is then extracted from this tensor map.

The surfaceness map is used as an input of a dedicated process for cell detection and segmentation by seeded watershed algorithm [Fernandez et al., 2010].

In the following sections, we will use the result of membranes binarisation (Fig. 3.6(c)) with associated membrane normal orientations to extract the embryo symmetry plane. The individualized cells (Fig. 3.6(f)) will be used for embryo co-registration.

3.3.3 Symmetry plane extraction

The embryo morphology holds a left-right (L-R) symmetry across early stages until the neurula rotation that takes place during the late neurula stage [Nishide et al., 2012]. The interface between cells of both sides of the L-R symmetry forms a plane-like structure (Fig. 3.7).

The principle of the symmetry plane extraction is threefold:

1. estimation of a set of plane normals through the study of the membrane orientation distribution;

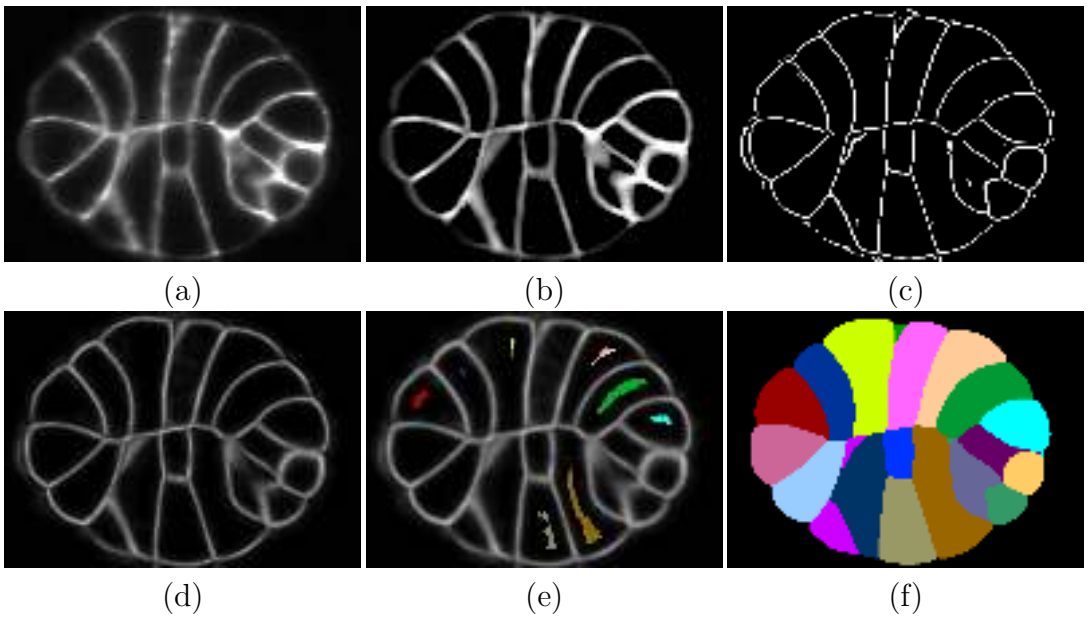


Figure 3.6 – Cell segmentation framework. (a) A section through a 3D embryo image. (b) Planar response filter. (c) Thresholded extrema. (d) Surfaceness map after tensor voting. (e) Seeds detection. (f) Cell segmentation by seeded watershed.

2. for each normal, the optimal plane is computed with an associated figure of merit: only the one with the best figure of merit is retained; and
3. the plane equation is refined using a least squares minimization.

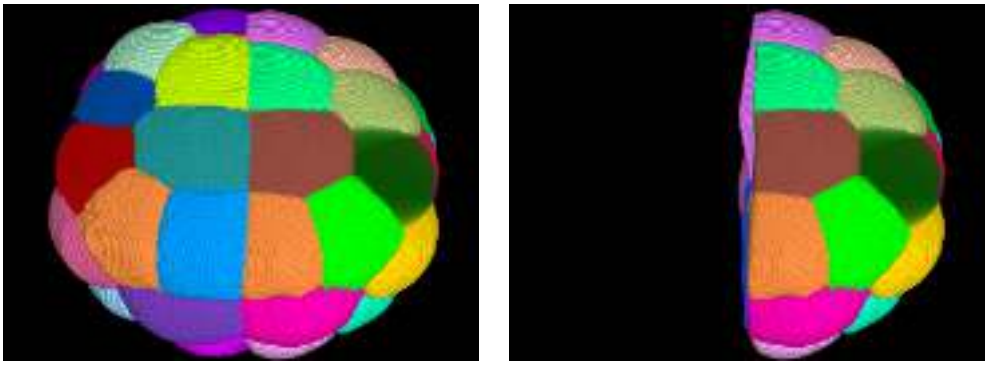
B denotes the thresholded extrema of the segmentation method (see section 3.3.2), while $\mathbf{n}(b)$ is the normal orientation (i.e. the unit eigenvector associated to the Hessian matrix largest eigenvalue) of point $b \in B$. Since B may be noisy, we mask B with the image M of dilated frontiers of the segmented cells. $B \cap M$ is an image in which most of the false membrane detections are removed.

3.3.3.1 Symmetry plane normal estimation

Since many cell membranes participate or are parallel to the symmetry plane, it is hypothesized that a large number of membrane normal directions are aligned with the symmetry plane normal, thus that the distribution of membrane normal directions will exhibit a maximum for the symmetry plane direction.

To compute this distribution, we discretize the unit sphere into the set of vectors $\mathbf{N} = \{\mathbf{n}_i\}$ such that $-\mathbf{n}_i \in \mathbf{N}, \forall \mathbf{n}_i \in \mathbf{N}$ (typically around 2000 vectors), and compute a kernel density estimate of the distribution of the membrane normal directions onto the $\{\mathbf{n}_i\}$ (Fig. 3.8). The distribution value at $\mathbf{n}_i$ is given by

$$D(\mathbf{n}_i) = \frac{1}{|B \cap M|} \times \sum_{b \in B \cap M} \frac{1}{N(b)} w(\mathbf{n}_i, \mathbf{n}(b)), \tag{3.8}$$



(a) (b)

Figure 3.7 – 3D view of embryo cell segmentation. (a) Full embryo. (b) Embryo with cells of its left half hidden. This illustrates the plane-like structure of Left-Right cell interface.

where $|B \cap M|$ denotes the cardinal of $B \cap M$, $N(b)$ is a normalization constant, and the kernel $w(\cdot,\cdot)$ is defined by

$$w(\mathbf{n}_i, \mathbf{n}(b)) = \exp \frac{-\arccos(|\mathbf{n}_i.\mathbf{n}(b)|)^2}{2\sigma^2} \tag{3.9}$$

with σ fixed to 2° since it ensures a maxima extraction fine enough to ensure satisfying candidate axes extraction. Let $\{\bar{\mathbf{n}}_k\}$ be the set of local maxima of the distribution D, $\{\bar{\mathbf{n}}_k\} \subset \{\mathbf{n}_i\}$. By construction, D is symmetrical, and since the search plane orientation is unoriented, we picked maxima in the half unit sphere. Moreover, let $D_{\underline{max}}$ be the maximal value of the distribution D, i.e. $D_{\underline{max}} = \max_{\mathbf{n}_i} D(\mathbf{n}_i)$, we restrict the maxima to those that verify $D(\bar{\mathbf{n}}_k) \geq D_{\underline{max}}/2$.

3.3.3.2 Symmetry plane equation global estimation

A plane P is defined by its normal $\mathbf{n} = (n_x,n_y,n_z)^T$ and a scalar d that sets the plane position along its normal axis. The plane equation is

$$P_{\mathbf{n},d}(x,y,z) = n_x x + n_y y + n_z z + d = 0. \tag{3.10}$$

We hypothesize that the symmetry plane of normal $\bar{\mathbf{n}}_k$ should be at the middle of the embryo, more precisely that the voxels $b \in B \cap M$ that contribute to the local maximum $\bar{\mathbf{n}}_k$ are equally parted by the symmetry plane , which can be formalized by

$$d_k = \arg\min_d \left| \sum_{b \in B^+(\bar{\mathbf{n}}_k, d)} w(\bar{\mathbf{n}}_k, \mathbf{n}(b)) - \sum_{b \in B^-(\bar{\mathbf{n}}_k, d)} w(\bar{\mathbf{n}}_k, \mathbf{n}(b)) \right| \tag{3.11}$$

where $B^-(\bar{\mathbf{n}}_k, d) = \{b \in B \cap M | P_{\bar{\mathbf{n}}_k,d}(b) < 0\}$ and $B^+(\bar{\mathbf{n}}_k, d) = (B \cap M) \setminus B^-(\bar{\mathbf{n}}_k, d)$.

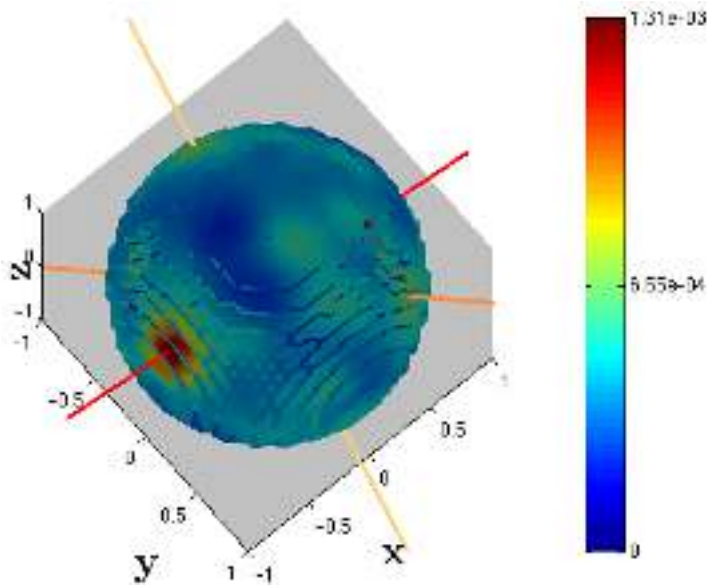


Figure 3.8 – Kernel density estimate of the orientation of binarised voxels. Red axis corresponds to the highest density. The other axes are local maxima greater than half the value of the global maximum.

3.3.3.3 Symmetry plane figure of merit

From the local maxima $\{\bar{\mathbf{n}}_k\}$, we have defined a set of planes $\{(\bar{\mathbf{n}}_k, d_k)\}$. Each plane can now be assessed with respect to the original data. Let S be the embryo segmentation, i.e. the union of the individualized cells, and $S \circ P$ the symmetrical of S with respect to plane P. If P is a symmetry plane, then S and $S \circ P$ are identical. We use the Dice coefficient to compare S and $S \circ P$ and finally retain as the symmetry plane the one that maximize the Dice coefficient,

$$(\hat{\mathbf{n}}, \hat{d}) = \arg \max_{(\bar{\mathbf{n}}_k, d_k)} \text{Dice}(S, S \circ P(\bar{\mathbf{n}}_k, d_k)). \tag{3.12}$$

3.3.3.4 Symmetry plane final estimation

Experiments demonstrated that $P(\hat{\mathbf{n}}, \hat{d})$ is a quite good symmetry plane, but may be slightly different from the L-R cell interface. Indeed all points $b \in B$ contribute to the distribution D and the retained symmetry plane $P(\hat{\mathbf{n}}, \hat{d})$ may be impaired by points b that did not belong to the L-R cell interface.

For a more accurate estimation of the symmetry plane, we perform a last iterative least-squares estimation until plane convergence where points b far away from the symmetry plane or with an orientation different from the plane normal are penalized:

$$(\mathbf{n}, d)^{(i+1)} = \arg \min_{(\mathbf{n}, d)} \sum_{b \in B} \rho_{(\mathbf{n},d)^{(i)}}(b) \| P_{\mathbf{n},d}(b) \|^2, \tag{3.13}$$

with

$$\rho_{(\mathbf{n},d)^{(i)}}(b) = \exp \frac{-\left(P_{(\mathbf{n},d)^{(i)}}(b)\right)^2}{2\sigma_d^2} \exp \frac{-\arccos(|\mathbf{n}^{(i)}.\mathbf{n}(b)|)^2}{2\sigma_a^2}$$

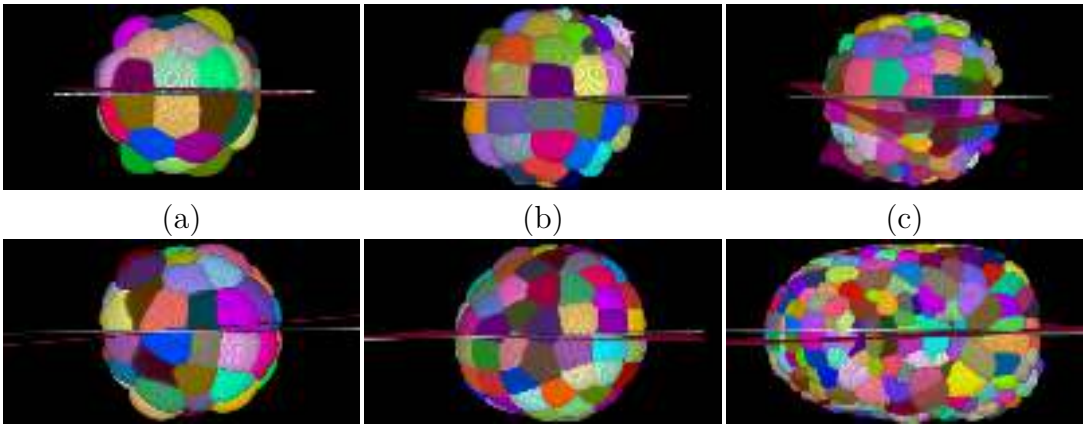


Figure 3.9 – L-R symmetry plane initialization ($P(\hat{\mathbf{n}}, \hat{d})$, in red) and after least squares estimation (in white). (a-c) First embryo, 32-, 76-cells and early gastrula stages. (d-e) Second embryo, 64- and 112-cells stages. (f) Fifth embryo, mid neurula stage.

$(\mathbf{n}, d)^{(0)}$ being initialized by $(\hat{\mathbf{n}}, \hat{d})$. The parameters σ_d and σ_a are the standard deviations weighting respectively the distance to the plane and the angular difference between $\mathbf{n}_b$ and the plane normal. We fix σ_d at half the approximate diameter of an embryo cell (it depends on the embryo stage) and arbitrarily set the parameter σ_a to 5°.

3.3.3.5 Experiments

We worked on Phallusia mammillata embryo images coming from 5 different samples.

For the first embryo, the membranes were marked by a lipophilic dye (FM4-64). The other embryo membranes were genetically marked using PH-GFP [Carroll et al., 2003]. Imaging of all samples was done at each time point from 4 different angles separated by 90 degrees with a MuViSPIM microscope [Krzic et al., 2012]. The 4 raw images are fused into one in order to have a time point image less sensitive to imaging defects and higher and isotropic resolution.

We tested the proposed method on 115 images: 99 images of the first embryo (ranging from 32 cells to 172 cells stages), 9 images of the second embryo (64 cells and 112 cells stage), 5 images of the third embryo (64 to 70 cells stage), one image of the fourth embryo (114 cells stage) and one image of the fifth embryo (mid neurula stage) that counts around 450 cells.

Visual inspection demonstrates that the method worked perfectly on all of the tested images (see Fig. 3.9). The last least squares estimation yields often similar result to the global estimation $P(\hat{\mathbf{n}}, \hat{d})$, but clear differences appear on some images (typically Fig. 3.9(c)-(f)).

With our implementation, the kernel density estimate building is computationally the most expensive step and is processed in about 20 minutes for an image of size $400 \times 400 \times 400$, while the plane initialization and refinement steps take less than 10 seconds.

3.3.4 Embryos registration

At early stages, ascidian development is stereotyped and invariant, i.e., there exists a cell-to-cell mapping between different embryos at the same stage. Such mapping offers a powerful means for population studies with single cell resolution but is challenging to establish. Indeed, at early stages, the embryo resembles a sphere tiled with cells, and pairing cells requires experimented researchers. Even after gastrulation, where a concavity (the blastopore) appears that is a natural landmark, recognizing pairs is made difficult by the increasing number of cells. We propose a method to register embryos (here with an affine transformation), which enables to build a further cell-to-cell mapping.

In this section, E and F are two embryo images at the same developing stage, with respectively the following notations:

- (P) and (Q) are their computed symmetry planes,
- $\mathbf{n}_P$, $\mathbf{n}_Q$ are normal vectors of (P), (Q),
- G, H are the centers of mass of the entire embryo cell segmentation,
- g, h are the projections of G and H on (P) and (Q),
- $C = \{c_i\}_{i\in[1,n]}$, $D = \{d_i\}_{i\in[1,m]}$ are the centers of mass of the segmented cells. Please note that it is not required to have an error-free segmentation.

3.3.4.1 Embryo registration

To register two embryos, the transformation that optimizes a cell-to-cell mapping is computed. Since the true mapping is unknown, several of them are compared. Obviously, testing all cell-to-cell pairings between two embryos is computationally intractable. However, it is reasonable to assume that registering embryos implies that their symmetry planes superimpose. Using this geometrical constraint allows to reduce greatly the number of cell pairings to be tested.

Let $\mathbf{I}^{(+)}$ and $\mathbf{I}^{(-)}$ be the rigid transformations that align h on g and $\mathbf{n}_Q$ on either $\mathbf{n}_P$ or $-\mathbf{n}_P$. This provides a reasonable alignment of both embryos up to one last degree of freedom, namely the angle θ of a rotation $\mathbf{R}(\theta)$ of center h and axis $\mathbf{n}_Q$. Let $\mathbf{T}^{(+)}(\theta) = I^{(+)} \circ \mathbf{R}(\theta)$, $\mathbf{T}^{(-)}(\theta) = I^{(-)} \circ \mathbf{R}(\theta)$. $[0, 2\pi]$ is discretized into the set of angles $\{\theta_i\}$ (we took 128 angles), yielding a set of transformations $\{\mathbf{T}^{(+)}(\theta_i)\} \cup \{\mathbf{T}^{(-)}(\theta_i)\}$.

For each transformation $\mathbf{T}$ of this set, we built a set of pairs of cell barycenters $\{(c_j, d_j)\}_{\mathbf{T}}$ such that they are the closest to each other, meaning that

$$\|c_j\mathbf{T}(d_j)\| = \min_{d_i\in D} \|c_j\mathbf{T}(d_i)\| = \min_{c_i\in C} \|c_i\mathbf{T}(d_j)\|. \tag{3.14}$$

Some cell barycenters c_i or d_i may remain unpaired. From the pairings $\{(c_j, d_j)\}_{\mathbf{T}}$, we estimated a transformation (we chose affine transformations) that minimizes the average of the square residuals:

$$\hat{\mathbf{T}}(\mathbf{T}) = \arg\min_{\mathbf{t}} r_{\mathbf{T}}(\mathbf{t}) \text{ with } r_{\mathbf{T}}(\mathbf{t}) = \frac{1}{|\{(c_j, d_j)\}_{\mathbf{T}}|} \sum_j \|c_j\mathbf{t}(d_j)\|^2.$$

Instead of a least squares estimation, and to discard erroneous pairings (e.g. due to segmentation errors), we preferred a least trimmed squares estimation [Rousseeuw and Leroy, 1987b], so that the pairings with the largest residuals are excluded from the transformation calculation. We chose to discard 20% of the pairings, i.e. the $k = \lfloor 0.2 \times |\{(c_j, d_j)\}_{\mathbf{T}}| \rfloor$ worst pairings, since it has been estimated that the segmentations have less than 10% of errors.

Last, the transformation $\mathbf{T}_{\underline{opt}}$ that best registers the two embryos is the one that has the minimum average of the square residuals:

$$\mathbf{T}_{\underline{opt}} = \arg \min_{\hat{\mathbf{T}}(\mathbf{T}), \mathbf{T} \in \{\mathbf{T}^{(+)}(\theta_i)\} \cup \{\mathbf{T}^{(-)}(\theta_i)\}} r_{\mathbf{T}}(\hat{\mathbf{T}}(\mathbf{T})). \tag{3.15}$$

3.3.4.2 Cell-to-cell mapping

The transformation $\mathbf{T}_{\underline{opt}}$ allowed to build the final cell-to-cell mapping $\{(c_j, d_j)\}_{\mathbf{T}_{opt}}$ between the two embryos, using the same symmetrical constraint as Eq. 3.14

$$\|c_j \mathbf{T}_{\underline{opt}}(d_j)\| = \min_{d_i \in D} \|c_j \mathbf{T}_{\underline{opt}}(d_i)\| = \min_{c_i \in C} \|c_i \mathbf{T}_{\underline{opt}}(d_j)\|. \tag{3.16}$$

3.3.4.3 Experiments

We used images from first to fourth embryos (described in section 3.3.3.5) for our experiments. We extracted from the set of images of the first, second and third embryos an image corresponding to the 64 cells stage and from the first, second and fourth embryos another image corresponding to the 112 cells stage, so that for both 64 cells stage and 112 cells stage, we could experiment the registration method on three image pairs coming from three distinct embryos.

Since we are interested in automated processing of high throughput acquisitions, we do not perform any correction of the cell segmentation provided by the framework of section 3.3.2. This test enables to evaluate the method's robustness facing segmentation errors. At the 64 cells stage, the first embryo has 71 segmented regions (instead of 64) due to 7 over-segmentation errors while the second one has 66 regions due to 2 over-segmentation errors and the third one has 69 regions due to 5 over-segmentation errors. At the 112 cells stage, the first embryo has 133 regions (instead of 112) due to 21 over-segmentation errors while the second one has 113 regions due to 1 under-segmentation error and 2 over-segmentation errors and the fourth one has 114 regions correctly segmented.

- For the 64 cells stage, the final cell-to-cell mapping built 64, 63 and 60 pairs between respectively the first and the second, the first and the third, and the second and the third embryo.

- For the 112 cells stage, the final cell-to-cell mapping built 105, 102 and 100 pairs between respectively the first and the second, the first and the fourth, and the second and the fourth embryo.

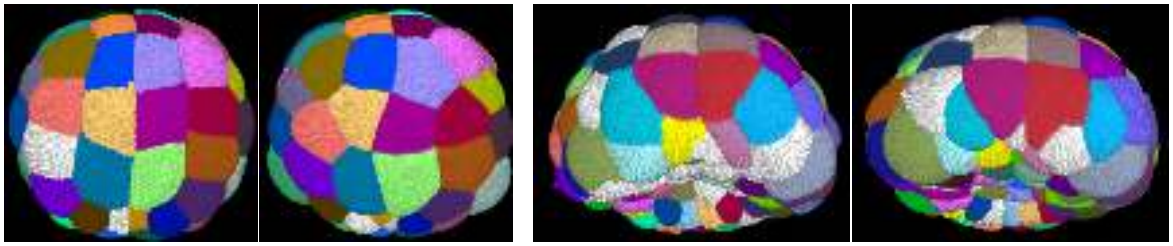


Figure 3.10 – Final region correspondence maps between the first and the second embryos. Left: 64 cells stage embryos. Right: 112 cells stage embryos. The white regions are unpaired ones.

As demonstrated by Fig. 3.10, and confirmed by careful visual inspection, almost all the pairings are correct. Mapping errors may be due to segmentation errors and/or limitations of affine transformation that did not allow a perfect superimposition of both embryos. Given that no other information has been incorporated (as adjacency between cells, etc.), there is room for further improvements that are however beyond the scope of this paper.

3.3.5 Conclusion and Future work

We proposed a novel method for early embryo Left-Right symmetry detection and showed that it enables to register embryos at a single cell level. The ability of robustly identifying corresponding cells in different embryos is a breakthrough for embryology, since it offers the means to first, conduct population statistical analysis at a cell level, and second, to register an atlas (i.e. a perfectly segmented template) onto any embryo which will thus allow to correct unavoidable segmentation errors.

Future research directions consist in improving the cell-to-cell mapping by incorporating additional information (lineages, cell adjacencies, etc.) and in building an average developing embryo from a population.

References

Besl, P. and McKay, N. (1992). A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell, 14(2):239–256. 103

Carroll, M., Levasseur, M., Wood, C., Whitaker, M., Jones, K., and McDougall, A. (2003). Exploring the mechanism of action of the sperm-triggered calcium-wave pacemaker in ascidian zygotes. Journal of Cell Science, 116(24):4997–5004. 114

Fernandez, R., Das, P., Mirabet, V., Moscardi, E., Traas, J., Verdeil, J., Malandain, G., and Godin, C. (2010). Imaging plant growth in 4-D: robust tissue reconstruction and lineaging at cell resolution. Nat Meth, 7:547–553. 103, 110

Keller, P. (2013). Imaging Morphogenesis: Technological Advances and Biological Insights. Science, 340(6137):1234168+. 102, 109

Krissian, K., Malandain, G., Ayache, N., Vaillant, R., and Trousset, Y. (2000). Model-based detection of tubular structures in 3d images. Comput Vis Image Underst, 80(2):130–171. 110

Krzic, U., Gunther, S., Saunders, T., Streichan, S., and Hufnagel, L. (2012). Multiview light-sheet microscope for rapid in toto imaging. Nat Methods, 9(7):730–3. 102, 114

Medioni, G., Lee, M., and Tang, C. (2000). Computational Framework for Segmentation and Grouping. Elsevier Science Inc., New York, NY, USA. 110

Michelin, G., Guignard, L., Fiuza, U., and Malandain, G. (2014). Embryo Cell Membranes Reconstruction by Tensor Voting. In ISBI, Beijing, China. IEEE. 110

Munro, E., Robin, F., and Lemaire, P. (2006). Cellular morphogenesis in ascidians: how to shape a simple tadpole. Curr Opin Genet Dev, 16(4):399–405. 102

Nishide, K., Mugitani, M., Kumano, G., and Nishida, H. (2012). Neurula rotation determines left-right asymmetry in ascidian tadpole larvae. Development, 139(8):1467–75. 110

Ourselin, S., Roche, A., Prima, S., and Ayache, N. (2000). Block matching: A general framework to improve robustness of rigid registration of medical images. In MICCAI'00, volume 1935 of LNCS, pages 557–566. Springer. 103

Rousseeuw, P. and Leroy, A. (1987a). Robust Regression and Outlier Detection. Wiley & Sons, New-York. 103

Rousseeuw, P. and Leroy, A. (1987b). Robust Regression and Outlier Detection. John Wiley & Sons, New York, NY, USA. 116

Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: a survey. IEEE Trans Med Imaging, 32(7):1153–90. 103

Traas, J. and Hamant, O. (2009). From genes to shape: understanding the control of morphogenesis at the shoot meristem in higher plants using systems biology. C R Biol, 332(11):974–85. 102

Zitová, B. and Flusser, J. (2003). Image registration methods: a survey. Image Vis Comput., 21(11):977–1000. 103

Chapter 4

Discussion

Contents

4.1 Quantifying the development

4.1.1 Segmenting and tracking cells.

As covered in Chapter 1, the digital acquisition of living organisms based on segmentation and tracking of cell lineages are the first steps towards quantifying the organism development. Imaging the development of a living organism is usually done by staining either membranes or nuclei; several methods to accurately segment both types of images have been proposed in the literature (briefly reviewed in chapter 1). We organize these methods in two categories depending on whether they proceed to segmentation and tracking in one or two distinct passes.

These two different methods are more or less appropriate for various types of segmentation needs. Two-step algorithms uncouple the segmentation and tracking processes. Following the segmentation of images at all time points, tracking links the segmented images together. Two pass methods have been shown to be accurate for low throughput image sequences for membranes [Fernandez et al., 2010] or nuclei [Keller et al., 2008; Long et al., 2009]. However, they are not sufficiently efficient when the number of imaging time points to segment in the sequence increases ([Amat et al., 2014] for nuclei and Chapter 2 for membranes). To tackle this issue, Amat et al. [2014] proposed a single step method that propagates the segmentation from one time point to the next, thereby integrating the segmentation and tracking processes. Because one-pass method propagate segmented information from time to time, decisions at a given time point may rely both on the image signal at this time point, but also on previous time-points. This makes it possible to make predictions locally that can then be checked against actual signal (e.g. in our work, the global seedling strategy used prior to the watershed is adapted locally).

For nuclei, this kind of integrated algorithm yields highly accurate results on various embryo and microscope types. To segment high-throughput sequences of membrane images, such one-pass algorithms are also necessary. During my PhD, I developped such a method applied to cell membrane segmentations. Although, some attempts in this direction had been made previously, these implementations had strong limitations. For example in Liu et al. [2014] and colleagues, no cell divisions were allowed. by contrast, the ASTEC pipeline can detect, segment and track cells across numerous cell divisions and over 180 time points.

ASTEC implementation takes advantage of three major biological *a-priori* of the studied system. Firstly, the cells cannot divide more than once between two acquisitions. Secondly, there is no cell death during the observed period of development. And thirdly, the cell volumes remain relatively stable through time. These are not common developmental features across all organisms so these *a-priori* restrain the type of organisms that can be observed using the current implementation of ASTEC. For example, in plant shoot apical meristems where the cells grow, the rule of conserved volume cannot be used. These biological rules, in ASTEC, have been split into separated modules, which

allows us to modify or remove rules if necessary (for instance, allowing two divisions between two time-points instead of one, allowing cell growth, etc.). However, since the algorithm described in this PhD takes full advantage of these *a priori* informations, the results strongly depends on them. To test the possibility to port ASTEC to another organism, I have started to test it on *Arabidopsis thaliana* shoot apical meristems, the results are encouraging but can still be improved. The biggest issue with ASTEC on *A. thaliana* is that the small number of time-points in the sequence does not allow to apply correctly the post segmentation algorithm.

The processing time of ASTEC is currently manageable. It takes about 1 week to process one sequence of 190 time-points. Interestingly, this processing time is independent of the number of cells to segment in the imaged objects. At present, segmentation and tracking thus takes about 10 times longer than imaging, which can be limiting when cohorts of embryos need to be reconstructed, for example to explore natural variability in the developmental process.

The length of processing time is mainly due to the fact that, to identify the seeds that will initiate the final watershed of each segmentation steps, ASTEC explores an extensive set of values for the h-min seeding parameter. This can be limiting when reconstructing cohorts of embryos to perform statistical analysis, for example. In order to decrease this computational time, some parts of the ASTEC pipeline, such as the computation of the local minima or the watershed algorithm, could be run in parallel. To parallelise even more the pipeline, one could consider the fact that each cell lineage is built independently and that they could be processed in parallel. Moreover, some parts of the process could probably be ported to run with graphic process units (GPU) increasing processing speed (but this would impose to use specific Graphic Cards).

4.1.2 Exploiting the segmentations

These segmentation and tracking algorithms give access to a variety of data, including the cell lineage trees. Due to their complexity, these data are often under-exploited. For example, cell division timings are often considered individually, cell by cell, usually without taking into account the global sequence of cell divisions throughout the cell lineage trees [Moore et al., 2013; Amat et al., 2014]. While giving insight in the development, these studies does not take into account the cells in their This type of analysis gives a global idea of the cell division events at the scale of the embryo or the tissues at one time point. This vision, however, limits the comparison between cell progenies and does not take into account multiple consecutive cell division events. To access comparative analysis of the whole cell progenies, we adapted an algorithm that compares tree structures (Zhang 96, Ferraro 00). This comparison between cell lineage trees embeds the division patterns and the lifespan of all the cells in a given progeny. This new metric allowed to show that the mitotic histories of cells are a diagnostic of the tissue fates, thereby pinpointing new putative events of cell specification. This observation could not have been done without computing a distance between trees that integrates the whole progeny, division timings and cell lifespans.

This method could be further improved. Currently, the distance has to be applied between trees that contain at least 2 rounds of cell divisions. The sequences thus have to be long enough to capture several division events. Integrating data on several embryos may increase the statistical power of the distance analysis. Also, the cell lineage distance used in this PhD was designed for an organisms whose development does not use programmed cell death. Taking cell death into account would resuire the addition of new edition rules.

We could also extend the distance by taking into consideration additional information, besides tree topology and timing of cell divisions. For example metrics on cells such as the volume, the compactness, genetic expression could be added to refine the distance. These improved metrics could refine the clustering of the different cell fates of *Phallusia mammillata* for example. This could also allow comparing more complex cell lineage trees in other species.

4.1.3 Towards 4D template digital embryos.

The quantitative description of digital embryo development is a crucial step towards the analysis of embryonic variability within a species and of embryonic divergence between species. In the case of ascidian embryos, the development is known to be highly stereotyped during early stages. Our analysis revealed, however, temporal variability between the left and right halves of the reconstructed embryo. This intra individual variability suggests that there may be more variability between individuals than initially estimated, a situation perhaps analogous to what has been described in C. elegans [Schnabel et al., 1997]. A thorough characterization of the spatial and temporal structure of embryonic variability would allow addressing important questions in development and evolution.

The identification of the most constrained cells, tissues or cell arrangements could illuminate the fundamental logic of developmental processes. For example, we have shown that cell inductions rely on precise cell-cell contacts (chapter 2), made trustworthy by the existence of a quasi-invariant cell lineage. One could, thus, expect that the interfaces between communicating cells would be more constrained than those between non-communicating cells, thus providing a means to geometrically identify cell inductions.

It has been proposed that there is a universal temporal pattern of embryonic variability shared by all metazoans, refered to as the hourglass [Duboule, 1994; Prud'homme and Gompel, 2010; Raff, 1996; Richardson, 2012]. According to these authors, there is a period during central embryogenesis during which embryonic morphologies and molecular pathways are more constrained than either earlier or later developmental stages. Several ascidian soecies have transparent embryos that could be imaged using the same procedure, and segmented with ASTEC. Comparison of their cell lineages and geometries

would allow testing the validity of the hourglass theory in the ascidian taxon. Comparison of intra-population variability with inter-species divergence may also identify selective pressures acting on embryonic morphologies.

Such characterisation can only be done by comparing the embryos together and by making statistical analysis. To compare multiple embryos, the first step is to characterize average embryonic development, which can then be used to quantify deviations from it. Building an atlas of average development requires the temporal and spatial alignments of multiple sequences. Chapter 3 presented the results of our initial attempts at the spatio-temporal alignment of embryos. These singularities can rely on patterns of gene expression, when available. This approach was used the case of *Drosophila* embryos, Fowlkes et al. [2008] used gene expressions as spatio-temporal landmarks. As an alternative, Castro-González et al. [2014] defined landmarks in *Zebrafish* embryos based on the asymmetry of the embryo and the position of the yolk.

In our case, our imaging did not capture patterns of gene expression and the ascidian embryos cleave no singular asymmetry that could be used as landmarks. Thus we proposed a method to build landmarks based on the deformation the embryos undergo in time. We also proposed a second method that spatially registers segmented images from the same time point. This method takes advantage of the high stereotypy of ascidian development. The first method struggles to spatially register embryos and the second method is only spatial. These two methods should therefore be considered as important but early steps towards the accurate spatio-temporal registration of ascidian development sequences.

4.2 Exploring embryogenesis.

During my PhD, I mostly focused my work on the analysis of the cell lineage trees and the topology of the embryo from the 4D digital embryo. This led to the formulation of hypothesis on cell fate decision processes and their link to cell-cell surface of contact. It also suggested that the main reason for the existence of an invariant cell lineage is the need to accurately position inducing and induced cells with respect to one another.

Many more data can, however, be extracted from this dataset and analysed, including the dynamics of cell, tissue and embryo shapes. We saw in the introduction that cell shape changes can be categorized into only a few major types of behaviours: cell growth, cell death, cell division, cell neighbour exchanges and cells shape changes. Our analysis confirmed that cell growth and cell death do not play significant roles during ascidian embryogenesis up to the initial tailbud stage. This spatially constrained cell dance has to be quantitatively characterized in order to determine the major drivers of tissue shaping.

Among these major types of behaviours, I started to study the impact of cell division on morphogenesis. As preliminary we found that while both *C. elegans* and *P. mam-*

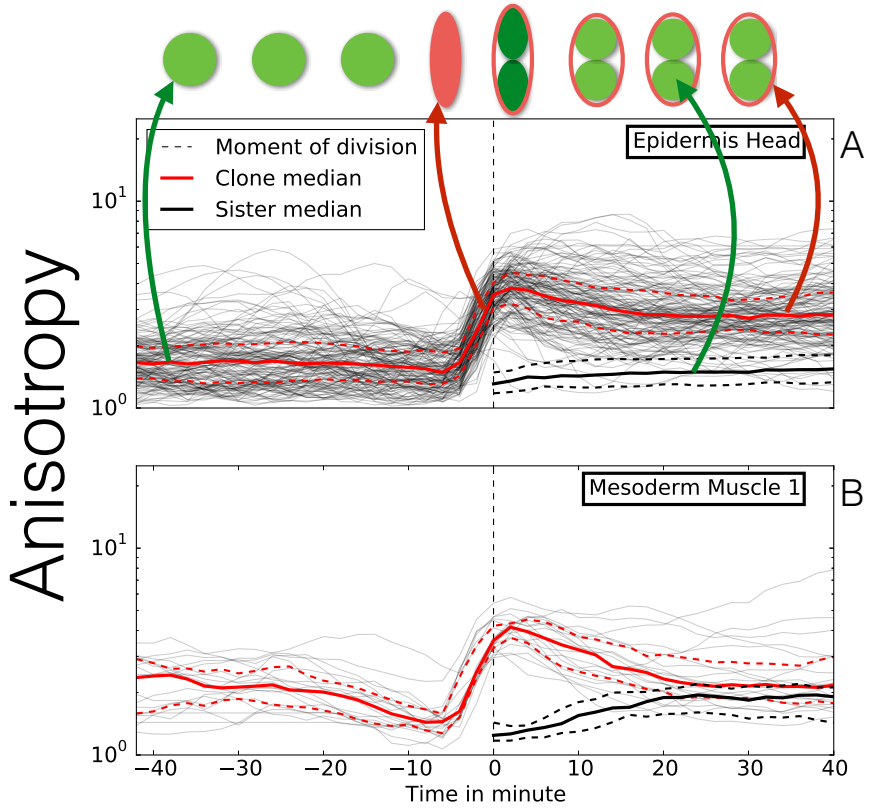


Figure 4.1 – ***Evolution of cell anisotropy around cell division events.*** *The evolution of anisotropy is shown. Before the division it is the anisotropy of the cell, after the division it is the anisotropy of the progeny of this cell (the two sister cell combined). The hashed line represent the moment of division. The black thick line represents the median of the anisotropy of the individual sister cells after the division. The anisotropy of epidermal cells (A) is stable during its life-span, increase right before the division and then remains stable. The progeny gains in anisotropy around the division process. This might let think that cell divisions drive epiboly in ascidians. By contrast, in the muscle, the anisotropy decreases after the division and the progenies revert to their original shapes.*

millata share an elongation along the anterio-posterior axis, *P. mammillata* embryos, by contrast with *C. elegans* embryos, the cell division orientations are not bias towards its A/P axis (not shown here).

Moreover, it is unfrequently noted that cell division orientation can only directly impact morphogenesis if the spaces occupied by the mother cell and by the clone of its two daughter cells differ, that is if cleavage has a direct net effect on shape. This has so far only been investigated in tissue culture conditions, revealing daughter clones are both more elongated and narrower along the division axis than their mother [Wyatt et al., 2015]. Our dataset allowed testing this idea for the first time in whole embryos

(example shown in Figure 4.1).

In all tissues, the aspect ratio (long/intermediary axis) of the daughter clones is larger than in their mother cells a few minutes before cleavage. The magnitude of this effect was maximal for the epidermis (Figure 4.1), in which the aspect ratios of each daughter cell were nearly identical to those of its mother. Epidermal cell division thus appears to be "fractal": it generates from one cell two cells of the same shape, but of half the volume. This peculiar geometry of cleavage, in a context of stable cell volumes, is a major contributor to the spreading of the epidermis

Ultimately, in order to detect crucial information to extract from such rich datasets, it is necessary to be able to browse it. In my opinion, this ability to browse the datasets is what lacks the most at the time being. For the moment, to the best of my knowledge, such data exploration cannot be done quickly. Such tools would allow the visualization of a segmented organism in time and space. Moreover, the computed metrics, such as the volumes, the anisotropy, the compactness, the fates should be possible to project on to the cells.

Some efforts have been done in this direction recently [Sommer et al., 2011; Peng et al., 2014; Barbier de Reuille et al., 2015] but the tools proposed are still quite rudimentary. They usually do not account for the multi-scale properties of the data, the 4D required visualization or the metric projection on the data. Such software would allow biologists to dig quickly and deeply into the data, which would allow to orient the analysis on to the segmented cells rather than screening the different cell properties.

References

Amat, F., Lemon, W., Mossing, D. P., McDole, K., Wan, Y., Branson, K., Myers, E. W., and Keller, P. J. (2014). Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. Nat Meth, 11(9):951–958. 121, 122

Barbier de Reuille, P., Routier-Kierzkowska, A.-L., Kierzkowski, D., Bassel, G. W., Schüpbach, T., Tauriello, G., Bajpai, N., Strauss, S., Weber, A., Kiss, A., Burian, A., Hofhuis, H., Sapala, A., Lipowczan, M., Heimlicher, M. B., Robinson, S., Bayer, E. M., Basler, K., Koumoutsakos, P., Roeder, A. H., Aegerter-Wilmsen, T., Nakayama, N., Tsiantis, M., Hay, A., Kwiatkowska, D., Xenarios, I., Kuhlemeier, C., and Smith, R. S. (2015). Morphographx: A platform for quantifying morphogenesis in 4d. eLife, 4. 126

Castro-González, C., Luengo-Oroz, M. A., Duloquin, L., Savy, T., Rizzi, B., Desnoulez, S., Doursat, R., Kergosien, Y. L., Ledesma-Carbayo, M. J., Bourgine, P., Peyriéras, N., and Santos, A. (2014). A digital framework to build, visualize and analyze a gene expression atlas with cellular resolution in zebrafish early embryogenesis. PLoS Comput Biol, 10(6):e1003670. 124

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate bauplan and the evolution of morphologies through heterochrony. Dev Suppl, pages 135–142. 123

Fernandez, R., Das, P., Mirabet, V., Moscardi, E., Traas, J., Verdeil, J.-L., Malandain, G., and Godin, C. (2010). Imaging plant growth in 4d: robust tissue reconstruction and lineaging at cell resolution. Nat Meth, 7(7):547–553. 121

Fowlkes, C. C., Hendriks, C. L. L., Keränen, S. V., Weber, G. H., Rübel, O., Huang, M.-Y., Chatoor, S., DePace, A. H., Simirenko, L., Henriquez, C., Beaton, A., Weiszmann, R., Celniker, S., Hamann, B., Knowles, D. W., Biggin, M. D., Eisen, M. B., and Malik, J. (2008). A quantitative spatiotemporal atlas of gene expression in the drosophila blastoderm. Cell, 133(2):364 – 374. 124

Keller, P. J., Schmidt, A. D., Wittbrodt, J., and Stelzer, E. H. (2008). Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. Science, 322(5904):1065–1069. 121

Liu, K., Lienkamp, S., Shindo, A., Wallingford, J., Walz, G., and Ronneberger, O. (2014). Optical flow guided cell segmentation and tracking in developing tissue. In Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on, pages 298–301. 121

Long, F., Peng, H., Liu, X., Kim, S. K., and Myers, E. (2009). A 3d digital atlas of c. elegans and its application to single-cell analyses. Nat Meth, 6(9):667–672. 121

Moore, J. L., Du, Z., and Bao, Z. (2013). Systematic quantification of developmental phenotypes at single-cell resolution during embryogenesis. Development, 140(15):3266–3274. 122

Peng, H., Bria, A., Zhou, Z., Iannello, G., and Long, F. (2014). Extensible visualization and analysis for multidimensional images using vaa3d. Nat. Protocols, 9(1):193–208. 126

Prud'homme, B. and Gompel, N. (2010). Evolutionary biology: Genomic hourglass. Nature, 468(7325):768–769. 123

Raff, R. A. (1996). The Shape of Life : Genes, Development, and the Evolution of Animal Form. University Of Chicago Press, Chicago. 123

Richardson, M. K. (2012). A phylotypic stage for all animals? Developmental cell, 22(5):903–904. 123

Schnabel, R., Hutter, H., Moerman, D., and Schnabel, H. (1997). Assessing normal embryogenesis in caenorhabditis elegans using a 4d microscope: variability of development and regional specification. Dev Biol, 184(2):234–265. 123

Sommer, C., Straehle, C., Kothe, U., and Hamprecht, F. (2011). Ilastik: Interactive learning and segmentation toolkit. In Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, pages 230–233. 126

Wyatt, T. P. J., Harris, A. R., Lam, M., Cheng, Q., Bellis, J., Dimitracopoulos, A., Kabla, A. J., Charras, G. T., and Baum, B. (2015). Emergence of homeostatic epithelial packing and stress dissipation through divisions oriented along the long cell axis. Proceedings of the National Academy of Sciences, 112(18):5726–5731. 125

Appendix A

Embryo cell membranes reconstruction by tensor voting

EMBRYO CELL MEMBRANES RECONSTRUCTION BY TENSOR VOTING

Gaël Michelin[⋆,1] *Léo Guignard*[†,2] *Ulla-Maj Fiuza*[†] *Grégoire Malandain*[⋆,1]

[⋆] INRIA, 06900 Sophia Antipolis, France
[†] CRBM, UMR 5237, CNRS, Univ. Montpellier 1 & 2, 34293 Montpellier, France

ABSTRACT

Image-based studies of developing organs or embryos produce a huge quantity of data. To handle such high-throughput experimental protocols, automated computer-assisted methods are highly desirable. This article aims at designing an efficient cell segmentation method from microscopic images. The proposed approach is twofold: first, cell membranes are enhanced or extracted by the means of structure-based filters, and then perceptual grouping (i.e. tensor voting) allows to correct for segmentation gaps. To decrease the computational cost of this last step, we propose different methodologies to reduce the number of voters. Assessment on real data allows us to deduce the most efficient approach.

***Index Terms*—** fluorescence microscopy, image segmentation, cell membrane

1. INTRODUCTION

The organism formation during animal or plant development is a major question in developmental biology. One challenge is to describe shape evolution as a geometrical output of gene activity [1, 2]. Due to recent progress in microscopy, time-lapse 3D images of living organisms can be recorded for several hours at a high frequency that permits to monitor cell deformations and divisions [3]. This results in huge data sets that require highly robust and efficient computational image analysis tools for segmenting the cells, registering the images if necessary, and extract the cell lineages [4].

The present work focuses on the cell segmentation step. The fluorescence of cell membranes or walls makes watershed [5] a method of choice [6, 7]. However, imaging noise and artifacts as well as fluorescence defects may cause segmentation errors. While deformable models may help to produce well-shaped cells, they may be computationally expensive if a large number of cells has to be segmented. Tensor voting is a means to correct for defects in segmentation [8]. The ACME method [7] demonstrates that, combined to a membrane enhancement filter, it overperformed the watershed algorithm applied on original data, but requires an important computational effort.

We propose to follow the same methodology as in [7], with the following contributions: an alternative structure-based detection filter, and several strategies to reduce the computational cost of the tensor voting step that is directly proportional to the number of retained points. Since several approaches can be built by combination of the filters and selection strategies, we perform a quantitative assessment on real data that allows us to select the most efficient ones.

Contact: {gael.michelin,gregoire.malandain}@inria.fr
[1]GM and GM are with INRIA team Morpheme.
[2]LG is also with INRIA team Virtual Plants, UMR AGAP, Montpellier, France.

2. MEMBRANE ENHANCEMENT

2.1. Membrane enhancement filters

Hessian-based filters have been proposed to detect vessels in medical imaging, because of their ability to characterize line-like structures [9]. More generally, assuming that the structures of interest are bright over a dark background, and that the eigenvalues λ_i of the Hessian matrix are ordered by their magnitude ($|\lambda_3| \geq |\lambda_2| \geq |\lambda_1|$), a line-like structure is characterized by large (and negative) λ_3 and λ_2 and a small $|\lambda_1|$ (the associated eigenvector $\mathbf{v}_1$ gives the line direction), while a plane-like structure is characterized by a large (and negative) λ_3 (the associated eigenvector $\mathbf{v}_3$ gives the orthogonal direction to the plane) and small $|\lambda_2|$ and $|\lambda_1|$. These properties have been exploited in medical imaging to detect vessels, either by building an ad-hoc response function based on the Hessian eigenvalues [10] or by integrating an edge detection response around a vessel center candidate [11]. In other words, both approaches characterize the structure orientation thanks to the Hessian eigenvectors, while the response amplitude is based respectively on the local second-order derivative amplitudes in Frangi's formulation and on the distant first-order derivative amplitudes in Krissian's one.

Mosaliganti and col. [7] have adapted Frangi's formulation to design a membrane enhancement filter (see Eq. 1):

$$R_F(P) = \begin{cases} 0 \text{ if } \lambda_3 \geq 0 \\ \left(1 - \mathrm{e}^{-\frac{\sum \lambda_i^2}{2\gamma^2}}\right) \mathrm{e}^{-\frac{\left|\frac{\lambda_2}{\lambda_3}\right|^2}{2\alpha^2}} \mathrm{e}^{-\frac{\left(\frac{\sqrt{|\lambda_1 \lambda_2|}}{|\lambda_3|}\right)^2}{2\beta^2}} \left(1 - \mathrm{e}^{-\frac{2c^2}{\lambda_1^2}}\right) \end{cases} \quad (1)$$

We propose here to adapt Krissian's approach to membranes. Although straightforward, this has not be done yet to the best of our knowledge. The response function at a point P of an image I is calculated by integrating an edge response at a distance r to a membrane center candidate:

$$R_K(P) = \begin{cases} 0 \text{ if } \lambda_3 \geq 0 \\ \frac{1}{2}\left(\nabla I(P - r\mathbf{v}_3).\mathbf{v}_3 - \nabla I(P + r\mathbf{v}_3).\mathbf{v}_3\right) \end{cases} \quad (2)$$

2.2. Extrema extraction

The two above filters are designed so that the response is maximal at the membrane center (with respect to its orthogonal direction). Thus suppressing the non-maxima will help to keep only pertinent information while suppressing the spurious one. This is done by extracting the directional (with respect to $\mathbf{v}_3$) extrema of the response, i.e.

$$E_X(P) = \begin{cases} 0 & \text{if } R_X(P) \leq R_X(P \pm \mathbf{v}_3) \\ R_X(P) & \text{else} \end{cases} \quad (3)$$

where X is respectively K for Krissian-like filter and F for Frangi-like one.

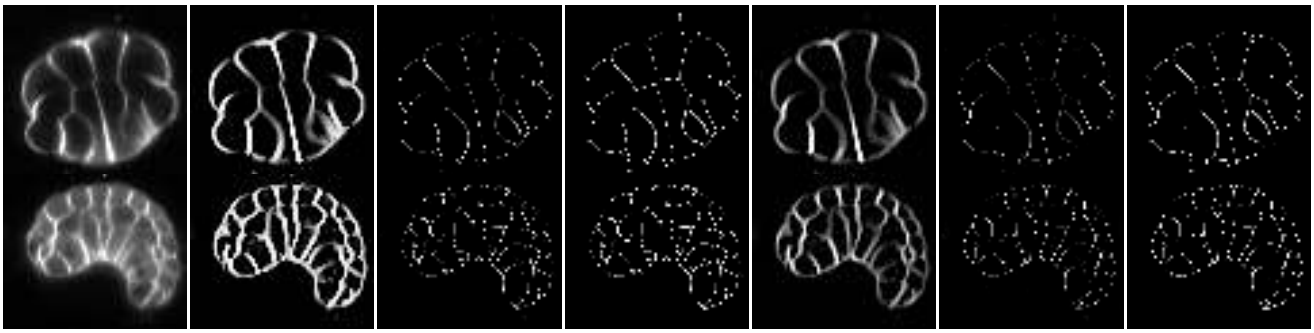


Fig. 1. First row, from left to right: a 2D cross-section of respectively the 3D image #1 (32 cells stage), the Frangi-like response, the Frangi extrema, the thresholded extrema of Frangi extrema, the Krissian-like response, the Krissian extrema, and the thresholded extrema of Krissian extrema. The second row depicted the same for image #91 (162 cells stage).

2.3. Extrema thresholding

The extrema are binarized by an hysteresis thresholding. The thresholds are chosen manually to obtain a visually good compromise between false positives and false negatives. It results a binary image B_X where $X \in \{F, K\}$ as before.

3. TENSOR VOTING

The above filters are designed to enhance plane-like structures. However, they will fail to enhance them at junctions or when the signal is too weak. Perceptual grouping, by the means of tensor voting, may address the second point, while the junctions issue can be resolved by a post-process of the tensor voting result.

3.1. Structural representation

Structures are locally represented (at each point P) by a 2^{nd} order tensor, $\mathbf{T}(P)$, i.e. a 3×3 real positive, symmetric matrix. Its decomposition in eigenvalues $\kappa_3 \geq \kappa_2 \geq \kappa_1 \geq 0$ and associated eigenvectors $\mathbf{e}_i$ allows to rewrite it as a linear combination of three generic tensors:

$$\mathbf{T} = \underbrace{\kappa_1 \sum_{i\in\{1,2,3\}} \mathbf{e}_i\mathbf{e}_i^t}_{\mathbf{T}_B} + \underbrace{(\kappa_2-\kappa_1) \sum_{i\in\{2,3\}} \mathbf{e}_i\mathbf{e}_i^t}_{\mathbf{T}_P} + (\kappa_3-\kappa_2)\underbrace{\mathbf{e}_3\mathbf{e}_3^t}_{\mathbf{T}_S} \quad (4)$$

$[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$ defines a basis where

$$\mathbf{T}_S = \begin{pmatrix} 0\,0\,0 \\ 0\,0\,0 \\ 0\,0\,1 \end{pmatrix}, \mathbf{T}_P = \begin{pmatrix} 0\,0\,0 \\ 0\,1\,0 \\ 0\,0\,1 \end{pmatrix} \text{ and } \mathbf{T}_B = \begin{pmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 0\,0\,1 \end{pmatrix}$$

These generic tensors are respectively named *stick*, *plate* and *ball* unit tensors.

- A stick tensor expresses uncertainty of data orientation in the two directions $\mathbf{e}_1$ and $\mathbf{e}_2$, it corresponds to a planar structure.
- A plate tensor expresses uncertainty in the direction $\mathbf{e}_1$ and corresponds to a line structure.
- A ball tensor does not express any orientation preference, it is the case in junction points.

3.2. Tensor voting

Tensor voting consists in building a tensor map from the votes of points P or *tokens*, that can be points without structural information ($P \in \mathcal{B}$), or points from lines ($P \in \mathcal{P}$) or planes ($P \in \mathcal{S}$), i.e. associated to some privileged directions. For each structure type, a tensor voting field is built (see [8] for details) that aims at expanding the structures along their preferential directions according to a scale parameter σ_T (figure 2). The result of tensor voting is then a tensor image $\mathbf{J}$:

$$\mathbf{J}_{\sigma_T}(M) = \sum_{\mathcal{X}\in\{\mathcal{B},\mathcal{P},\mathcal{S}\}} \sum_{P\in\mathcal{X}} \alpha_{\mathcal{X}}(P)\mathbf{V}_{\mathcal{X},\sigma_T}(\mathbf{PM}, \mathbf{e}_i(P)) \quad (5)$$

where $\mathbf{V}_{\mathcal{X},\sigma_T}(\mathbf{PM}, \mathbf{e}_i(P))$ denotes the vote of token P of type $\mathcal{X} \in \{\mathcal{B}, \mathcal{P}, \mathcal{S}\}$ at point M at voting scale σ_T, weighted by $\alpha_{\mathcal{X}}(P)$.

(a) (b) (c) (d)

Fig. 2. Voting fields: (a) a cut of the voting field for stick component; (b) a cut of the voting field for plate component (direction $\mathbf{e}_1$ is normal to the page); (c) and (d) surfaceness maps S after tensor voting step.

We are only interested here in plane-like structures (the membranes), hence for an image $I \in \{R_X, E_X, B_X\}$, we only consider stick tensor votes with

$$\mathbf{J}_{\sigma_T}(M) = \sum_{P/I(P)>0} I(P)\mathbf{V}_{\mathcal{S},\sigma_T}(\mathbf{PM}, \mathbf{v}_3(P)) \quad (6)$$

where the input stick tensor is built from the eigenvector $\mathbf{v}_3$ (i.e. $\mathbf{e}_3 = \mathbf{v}_3$) of the Hessian matrix, and its vote $\mathbf{V}_{\mathcal{S},\sigma_T}$ is determined as:

$$\mathbf{V}_{\mathcal{S},\sigma_T}(\mathbf{PM}, \mathbf{v}_3(P)) = \mathrm{e}^{-\frac{s^2+c\kappa^2}{\sigma_T^2}} \mathbf{n}\mathbf{n}^t \quad (7)$$

where s is the arc length and κ is the curvature of the arc of the osculating circle at P (i.e. normal to $\mathbf{v}_3$ in P) which goes through

M. $\mathbf{n}$ is the unit vector normal to the arc in M. The parameter c controls the degree of vote's decay with curvature and is given by $c = -16\log(0.1) \times (\sigma_T - 1) \times \pi^{-2}$. Note that no votes are cast if the angle between $\mathbf{v}_3$ and $\mathbf{n}$ is larger than 45°. Since we only consider points with non-null intensity as tokens, the computational cost of tensor voting is obviously ordered from $I = R_X$ to B_X, as it linearly depends on the number of non-null points P. Please note that votes are weighted by the filter response for $I \in \{R_X, E_X\}$, the thresholded extrema value being either 0 or 1.

4. CELL SEGMENTATION

From the tensor map $\mathbf{J}$, a surfaceness map S is computed with $S(M) = \kappa_3(\mathbf{J}(M)) - \kappa_2(\mathbf{J}(M))$ (see eq. (4)) that is subsequently used to segment the cells. For that purpose, the watershed method is used. However, this approach is known to be prone to oversegmentations and since some gaps may still exist at junctions, we design a dedicated seed extraction method, and the labeled seeds will be used as sources for the watershed instead of all minima of the S image.

First, h-minima are extracted from S [12]. Since membrane segmentation gaps form bridges between two adjacent cells, we recognize them by computing a distance map inside the extracted h-minima and then by extracting the h-maxima of the distance map. These labeled h-maxima are used as sources for a watershed segmentation with a regularized (i.e. convoled by a Gaussian) version of S, i.e. $S * G_{\sigma_W}$, in order to solve the junctions issue.

5. EXPERIMENTS

5.1. Data

We imaged a simple chordate organism, *Phallusia mammillata*, embryos. Embryo's membranes are marked by a lipophilic dye (FM4-64 which responds at $\sim 750nm$ from an excitation at $595nm$). We started the imaging session at the end of the 32 cells stage of the embryo and stoped it during its 172 cells stage. The embryo was imaged every minute from 4 different angles for 2 hours with a light-sheet microscope, the MuViSPIM [13], yielding at each timestep 4 images of around 200 slices of 1200×1200 pixels, with a pixel size of $0.26\mu m$ and a slice thickness of $1\mu m$. The 4 images were then fused to mitigate image acquisition defects due to, for example, light diffraction and/or microscope anisotropy.

One special characteristic of this setup is that the dye is slowly internalized inside the cytoplasm. It has, as impact, a degradation through time of the signal to noise ratio. This defect will allow us to have different image qualities and therefore to test two different experiment conditions. We choose for our tests the 1^{st} and the 91^{th} (after 1h30min of imaging) images from the sequence, corresponding to respectively 32 and 162 cells. While the first image may be considered as acquired under ideal imaging conditions, the second one corresponds to degrading conditions. The visual comparison of the two images (see Figure 1) depicts clearly the dye internalization.

5.2. Computational issues

The described method relies on a number of parameters. First, Frangi's and Krissian's filters require the computation of the image derivatives, which is achieved by convolving the image with the Gaussian derivatives. Although these filters can be embedded into a multi-scale approach to handle difference of sizes of the structures to segment, we choose to use only one scale denoted by σ_R since the membranes have a homogeneous thickness. We test $\sigma_R \in \{2, 3, 4, 5, 6\}$, with $r = \sigma_R$ in eq. 2. The extrema thresholding step is performed by fixing manually the thresholds for each extrema image in order to favour false negatives rather than false positives. The size of the tensor voting field $\mathbf{V}_S$ is governed by an other standard deviation σ_T that has to be chosen accordingly to the size of the gaps to be filled. We fix $\sigma_T = 10$ for all the experiments. The height h for the h-minima step is an important issue since it directly depends on the brightness of membrane structures from the surfaceness map S. We test $h \in \{10, 15, 30\}$. The h-maxima's height is less important to determine since it only has an influence on the size of the detected seeds. We fix the h-maxima parameter at 5 for the whole tests. The regularized version of S used for the watershed segmentation is processed with convolutions by the derivatives of a Gaussian filter of scale $\sigma_W = 3$ in order to remove junction gaps in S.

5.3. Evaluation methodology

The purpose of this work is to design an efficient method for cell segmentation. We want to assess the use as tokens for tensor voting of either the filter response R_X, the extracted extrema E_X, or the thresholded extracted extrema B_X for two filters, namely the Frangi-like one ($X{=}F$) and the Krissian-like one ($X{=}K$). This yields 6 token images to be compared, multiplied by the number of tested parameter sets. Note that using R_F as token image is similar to the ACME method [7], thus we have a direct comparison with this approach.

The 3D images are also processed by an Fernandez's method [6]. Briefly, this is a watershed on the (regularized) original data with an ad-hoc seed/source detection. The obtained images have been manually corrected, yielding ground truth (GT) segmentation. Since we do not use the original image for the watershed, there will be unavoidable differences at the cell borders between the ground truth segmentation and the ones we obtain. In addition, we are more interested in evaluating the topological errors (i.e. the number of over-segmentations and of missed cells) than the precision of the border of the segmented cells. For these reasons, we design three measures to quantify these errors based on the comparison of the detected seeds for watershed (see section 4) against the segmented cells of the ground truth, instead of comparing the segmentations (for instance with a Dice index).

- True detections (`TD`) characterize a one-to-one mapping between a GT cell and a seed: the cell contains only one seed, and this seed does not intersect any other cell.
- An over-detection (`OD`) occurs if a cell contains more than one seed, and there will be as many over-segmentations of this cell as there are supplementary seeds (a cell containing 3 seeds counts for 2 `OD`).
- An under-detection (`UD`) can occur by two different ways, firstly if one cell does not contain any seed, and secondly if a seed intersects more than one cell.

5.4. Results

Table 1 presents the combination of tokens images and parameters that yield the largest number of True Detections (`TD`) together with the smallest error measures. For each combination, we present the different error measures (i.e. `TD`, `OD` and `UD`) but also the computational cost of the tensor voting step defined as the computational time normalized by the computational time of the ACME method (that has then a computational cost of 1 by definition).

Table 1. Errors measures for the best combinations of token images and parameters.

Img. (♯cell)	Tokens	σ	$h_{\min}$	TD	OD	UD	TV cost
Im1 (32)	B_K	5	30	30	3	0	$2.98\,10^{-2}$
	B_F	5	30	30	3	0	$3.23\,10^{-2}$
	R_F	3	10	28	4	0	1.00
	E_K	4	10	28	4	0	$12.66\,10^{-2}$
	Fernandez [6]			31	9	0	
Im2 (162)	B_K	4	15	149	15	3	$3.74\,10^{-2}$
	E_K	5	30	144	11	8	$10.84\,10^{-2}$
	R_F	4	15	139	18	8	1.00
	B_K	3	15	136	29	2	$3.96\,10^{-2}$
	Fernandez [6]			128	40	5	

Results on the 162 cells image demonstrate that the structure-based approaches followed by tensor voting clearly outperform a direct watershed segment for poor quality images (because of the dye internalization, some interior cell points may have higher intensities than points of low contrast membranes).

It has to be pointed out that Mosaliganti's method [7] is in the top 4 best approaches for both test images. Moreover, all the best structure-based approaches yield comparable results in terms of segmentation quality for the high quality image (the 32 cells image). However, some differences can be noticed for the low quality image (the 162 cells image), where either the extrema or the binarized extrema of the Krissian-like filter seems to slightly outperform ACME.

More important, these two methods exhibit a significantly smaller computational cost (almost 1 or 2 order of magnitude) than the ACME method for the tensor voting step, making them the methods of choice for cell segmentation.

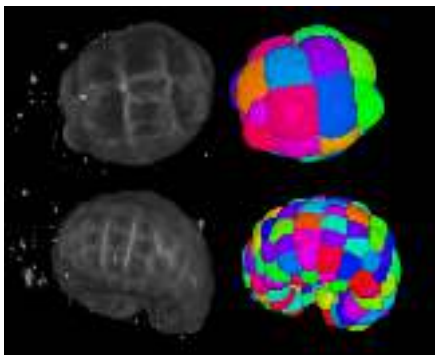


Fig. 3. From left to right: a 3D view of the 32 cells image, its segmentation (first row); the same for the 162 cells image (second row).

6. CONCLUSION AND FUTURE WORK

We investigated different segmentation methods, relying on a structure-based filter followed by a perceptual grouping step. The results demonstrate that: such methods outperform a direct watershed, the computational cost of tensor voting can be significantly reduced by extracting pertinent information from the structure-based filter, and a new structure-based filter (inspired from Krissian's work on vessels) slightly outperforms the Frangi-like filter. Apart of slight improvements (eg automated computation of the extrema thresholds), next steps will consist in evaluating the proposed methodology (i.e. tensor voting on thresholded Krissian's extrema) on whole 3D+t sequences, to extract embryos cell lineages.

Acknowledgements

Lars Hufnagel (EMBL, Heidelberg, Germany) provided a valuable help for image acquisition and technical discussions. Ulla-Maj Fiuza is partially funded by the *Fondation pour la Recherche Médicale*.

7. REFERENCES

[1] E Munro, F Robin, and P Lemaire, "Cellular morphogenesis in ascidians: how to shape a simple tadpole," *Curr Opin Genet Dev*, vol. 16, no. 4, pp. 399–405, 2006.

[2] J Traas and O Hamant, "From genes to shape: understanding the control of morphogenesis at the shoot meristem in higher plants using systems biology," *C R Biol*, vol. 332, no. 11, pp. 974–85, 2009.

[3] PJ Keller, "Imaging morphogenesis: technological advances and biological insights," *Science*, vol. 340, no. 6137, pp. 1234168, 2013.

[4] TV Truong and W Supatto, "Toward high-content/high-throughput imaging and analysis of embryonic morphogenesis," *Genesis*, vol. 49, no. 7, pp. 555–69, 2011.

[5] L Vincent and P Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans Pattern Anal Mach Intell*, vol. 13, no. 6, pp. 583–598, 1991.

[6] R Fernandez, P Das, V Mirabet, E Moscardi, J Traas, JL Verdeil, G Malandain, and C Godin, "Imaging plant growth in 4-d: robust tissue reconstruction and lineaging at cell resolution," *Nat Meth*, vol. 7, pp. 547–553, 2010.

[7] KR Mosaliganti, RR Noche, F Xiong, IA Swinburne, and SG Megason, "Acme: automated cell morphology extractor for comprehensive reconstruction of cell membranes," *PLoS Comput Biol*, vol. 8, no. 12, 2012.

[8] G Medioni, MS Lee, and CK Tang, *Computational Framework for Segmentation and Grouping*, Elsevier Science Inc., New York, NY, USA, 2000.

[9] C Lorenz, IC Carlsen, TM Buzug, C Fassnacht, and J Weese, "Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2D and 3D medical images," in *CVRMed-MRCAS'97*. 1997, number 1205 in LNCS, pp. 233–242, Springer.

[10] AF Frangi, WJ Niessen, KL Vincken, and MA Viergever, "Multiscale vessel enhancement filtering," in *MICCAI'98*. 1998, vol. 1496 of *LNCS*, pp. 130–137, Springer.

[11] K. Krissian, G. Malandain, N. Ayache, R. Vaillant, and Y. Trousset, "Model-based detection of tubular structures in 3d images," *Comput Vis Image Underst*, vol. 80, no. 2, pp. 130–171, 2000.

[12] P Soille, *Morphological image analysis: principles and applications*, Springer, 1999.

[13] U Krzic, S Gunther, TE Saunders, SJ Streichan, and L Hufnagel, "Multiview light-sheet microscope for rapid in toto imaging," *Nat Methods*, vol. 9, no. 7, pp. 730–3, 2012.

Title Quantitative analysis of animal morphogenesis: from high-throughput laser imaging to 4D virtual embryo in ascidians

Abstract

Ascidian embryos develop with stereotyped and evolutionarily conserved invariant cell lineages to produce in a few hours or days tadpole larvae with a small number of cells. They thus provide an attractive framework to describe with cellular resolution the developmental program of a whole organism. During my PhD, I developed a quantitative approach to describe the evolution of embryonic morphologies during the development of the ascidian *Phallusia mammillata*. I then used this approach to systematically characterize in detail the logic of cell fate induction events.

To quantitatively characterize cell behaviors during embryogenesis, we used multi-angle light-sheet microscopy to image with high spatio-temporal resolution entire live embryos with fluorescently labeled plasma membranes. To extract biological information from this imaging dataset, I then developed a conceptually novel automated method for 4D cell segmentation, ASTEC. Applied to a *Phallusia mammillata* embryo imaged for 6 hours between the 64-cell and the initial tailbud stages, this method allows the accurate tracking and shape analysis of 1030 cells across 640 cell divisions. The resulting 4D digital embryo can be formalized as a dynamic graph, in which cells are represented by nodes, linked within a time point by edges that represent their spatial neighborhood, and between time points by temporal edges describing cell lineages.

Based on this quantitative digital representation, we systematically identified cell fate specification events up to the late gastrula stage. Computational simulations revealed that remarkably simple rules integrating measured cell-cell contact areas with boolean spatio-temporal expression data for extracellular signalling molecules are sufficient to explain most early cell inductions. This work suggests that in embryos establishing precise stereotyped contacts between neighboring cells, the genomic constraints for precise gene expression levels are relaxed, thereby allowing rapid genome evolution.

Keywords Development; Segmentation; Cell Tracking; Atlas 4D; Ascidians

Titre Analyse quantitative de la morphogenèse animale : de l'imagerie laser haut-débit à l'embryon virtuel chez les ascidies

Résumé

Les embryons d'ascidies se développent avec un lignage cellulaire stéréotypé et évolutionairement conservé pour produire en quelques heures ou jours un têtard comportant un petit nombre de cellules. De ce fait, ils fournissent cadre intéressant pour décrire avec une résolution cellulaire le programme de développement d'un organisme complet. Pendant mon doctorat, j'ai développé une approche quantitative pour décrire l'évolution morphologique embryonnaire pendant le développement de Phallusia mammillata. J'ai ensuite utilisé cette approche pour systématiquement caractériser en détail les logiques des événements de spécifications de destin cellulaire.

Pour caractériser quantitativement les comportements cellulaires pendant l'embryogenèse, nous avons utilisé de la microscopie à feuille de lumière multi-angles pour imager des embryons entiers à haute résolution spatio-temporelle. Les membranes plasmiques étaient marquées pour permettre l'identification des cellules. Pour extraire les informations biologiques de ce jeu de donnés, j'ai développé une nouvelle méthode pour segmenter les cellules en 4D, ASTEC. Une fois appliquée aux embryons de Phallusia mammillata imagés pendant 6 heures entre le stade 64 cellules et le début des stades bourgeon caudal, cette méthode a permis de récupérer la forme et de suivre 1030 cellules pendant 640 divisions. L'embryon digital 4D résultant peut être formalisé par un graphe dynamique, dans lequel les cellules sont représentées par des sommets reliés par des arrêtes représentant au sein d'un point de temps leur voisinage spatial, et entre différents points de temps leur lignage cellulaire.

Basé sur cette représentation digitale et quantitative, nous avons systématiquement identifié les événements de spécification cellulaire jusqu'au dernier stade de la gastrulation. Des simulations informatiques ont révélé que des règles remarquablement simples intégrant les aires de contacts cellulaires et les expressions spatio-temporelles booléennes de signaux moléculaires extracellulaires sont suffisantes pour expliquer les inductions cellulaires au cours du développement précoce. Ce travail suggère que pour les embryons établissant des contacts stéréotypés et précis entre cellules voisines, les contraintes génomiques sont relâchées, ce qui permet une évolution plus rapide du génome.

Mots-clés Développement ; Segmentation ; Suivi cellulaire ; Atlas 4D ; Ascidies