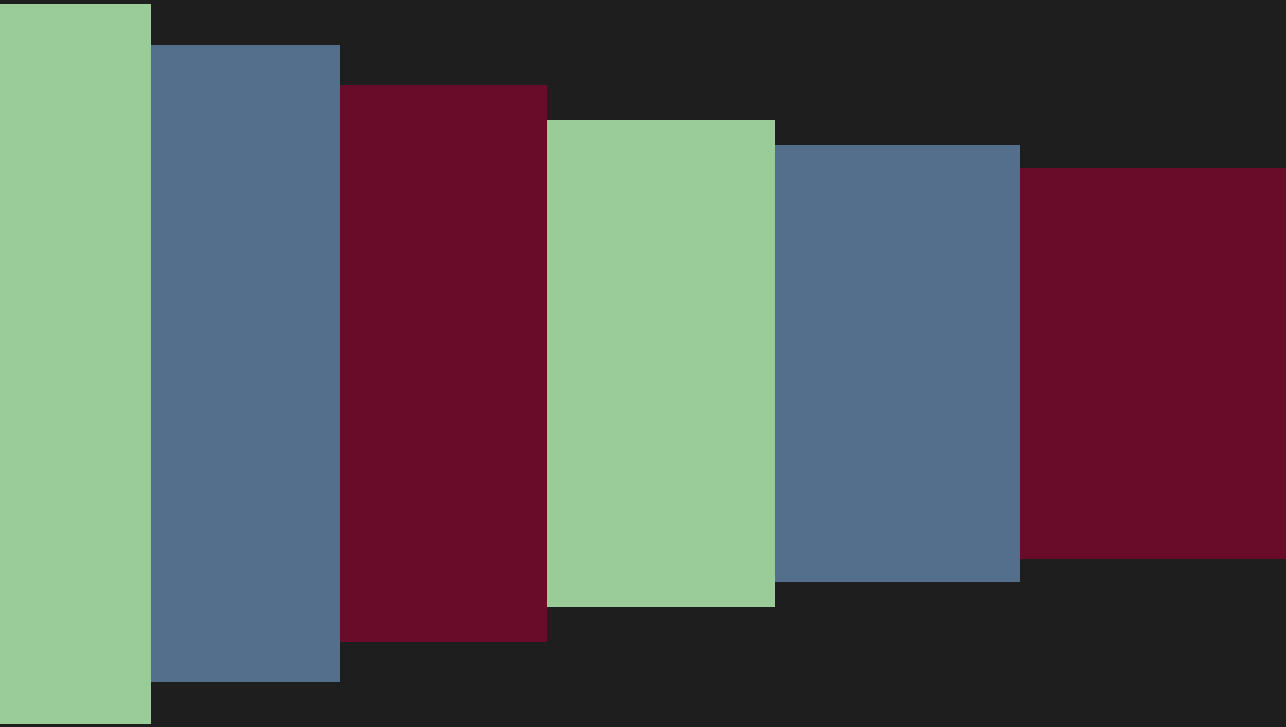


Of clocks and waves, stripes and shapes

Evo-devo models of sequential segmentation



Renske M.A. Vroomans

Of clocks and waves, stripes and shapes

Evo-devo models of sequential segmentation

Renske M.A. Vroomans

Cover Renske Vroomans
Print Gildeprint, Enschede
ISBN 978-90-393-6690-5

No part of this thesis may be reproduced in any form, by any print, microfilm, or any other means, without prior written permission of the author.

Of clocks and waves, stripes and shapes

Evo-devo models of sequential segmentation

Van klokken en golven, strepen en vormen

Evo-devo modellen van sequentiële segmentatie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 6 januari 2017 des middags te 4.15 uur

door

Renske Maria Anna Vroomans

geboren op 21 juni 1988 te Den Haag

Promotor: Prof.dr. P. Hogeweg
Copromotor: Dr. K.H.W.J. ten Tusscher

Aan mijn ouders...

Contents

| | | |
|----------|---|-----------|
| 1 | General Introduction | 1 |
| 1.1 | Preface | 2 |
| 1.2 | The how and the why of development, evolution and modelling | 4 |
| 1.3 | Modes of segmentation | 6 |
| 1.4 | Overview | 15 |
| | | |
| I | Clocks, waves and stripes | 17 |
| 2 | Analysing the influence of wavefront dynamics on clock evolution | 19 |
| 2.1 | Introduction | 21 |
| 2.2 | Methods | 23 |
| 2.3 | Results | 31 |
| 2.4 | Discussion | 47 |
| 3 | Modelling asymmetric somitogenesis: a pointer to potentially non-neutral species differences | 55 |
| 3.1 | Introduction | 57 |
| 3.2 | Results | 60 |
| 3.3 | Discussion | 82 |
| 3.4 | Methods | 85 |
| 3.5 | Movies | 93 |

| | | |
|-----------|--|------------|
| II | Clocks, stripes and shapes | 95 |
| 4 | In silico evo-devo: reconstructing stages in the evolution of animal segmentation | 97 |
| 4.1 | Introduction | 99 |
| 4.2 | Methods | 101 |
| 4.3 | Results | 108 |
| 4.4 | Discussion | 120 |
| 4.5 | Acknowledgements | 123 |
| 5 | Segment-specific adhesion as a driver of convergent extension | 131 |
| 5.1 | Introduction | 133 |
| 5.2 | Results | 135 |
| 5.3 | Discussion | 147 |
| 5.4 | Methods | 152 |
| 5.5 | Acknowledgements | 156 |
| 5.6 | Supplementary information | 164 |
| 5.7 | supplementary videos | 166 |
| 6 | Summarising Discussion | 167 |
| 6.1 | Making waves – frequency gradients | 168 |
| 6.2 | Making shapes – axis extension | 170 |
| 6.3 | Using developmental and evolutionary models | 173 |
| 6.4 | Moving models beyond standard clock-and-wavefront | 174 |
| 6.5 | Concluding remarks | 176 |
| | Bibliography | 177 |
| | Samenvatting | 189 |
| | Curriculum Vitæ | 193 |
| | List of Publications | 195 |
| | Acknowledgements | 197 |

...en mijn broertje

“The whole of life is just like watching a film. Only it’s as though you always get in ten minutes after the big picture has started, and no-one will tell you the plot, so you have to work it out all yourself from the clues.”

Terry Pratchett, *Moving Pictures* (1990)

1

General Introduction

1.1 Preface

Between the fertilisation of an egg and the birth of an animal, one of life's most dramatic processes takes place: embryonic development. The first major event is gastrulation, which transforms the embryo from a simple ball of similar-looking cells to a complex torus-like structure, with different tissue layers forming the gut, nervous tissue, skin and the material for internal organs. In bilateral animals, gastrulation coincides with the formation of the main body axis that runs from head to tail, perpendicular to the dorsal-ventral and the left-right axis (Wolpert, 2007). This process involves extensive tissue reshaping, cell ingression and divisions in a posterior undifferentiated zone. In at least three bilaterian clades – the annelids, arthropods and vertebrates –, axis formation is paired with segmentation (Balavoine and Adoutte, 2003, Blair, 2008, Davis and Patel, 1999, Peel and Akam, 2003). Segmentation divides the main body axis into repeated elements, which can then be used to specify ribs and vertebrae in vertebrate animals, or e.g. multiple pairs of legs in centipedes (Fig. 1.1A). Segments first manifest in the developing embryo as an alternating stripe pattern of so-called segmentation genes across the body axis (Fig. 1.1B). In most segmented animals, these stripes are laid down in an anterior-to-posterior progression, appearing one after another with precise periodic timing from a posterior zone with undifferentiated cells; this is known as sequential segmentation.¹

The regular pattern and clock-like nature of sequential segmentation has inspired theoreticians for decades. This started in 1976 with the conceptual “clock and wavefront model” for vertebrate somitogenesis by Cooke and Zeeman, who proposed that a cellular clock could underlay the process (Cooke and Zeeman, 1976), in the form of an oscillatory cell state. This clock is paired with a wavefront of cell differentiation, which retracts towards the posterior due to tissue growth, and transforms the temporal dynamics of the clock into a spatial pattern of somites.² Evidence for this model was first found in 1997, when Palmeirim *et al.* (Palmeirim *et al.*, 1997) demonstrated that *hairy* mRNA oscillates in the vertebrate undifferentiated zone, and that these oscillations are crucial for somitogenesis. Since then, evidence has accumulated in favour of the “clock and wavefront” model with the discovery of more oscillating genes essential for somitogenesis, and the discovery that signalling gradients of Wnt, FGF and retinoic acid (RA) can modify the point where somites are formed, resembling a wavefront (Aulehla and Pourquié, 2010). The recent discovery that gene expression oscillations are also involved in insect segmentation, supports the idea that a clock-and-wavefront mechanism could also be at work outside the vertebrate clade (Choe *et al.*, 2006, El-Sherif *et al.*, 2012, Sarrazin *et al.*, 2012).

¹Some insect species (like the fruit fly *Drosophila*) form segments differently: all stripes are laid down simultaneously and well before gastrulation and formation of the main body axis.

²More information is given in section 1.3.1

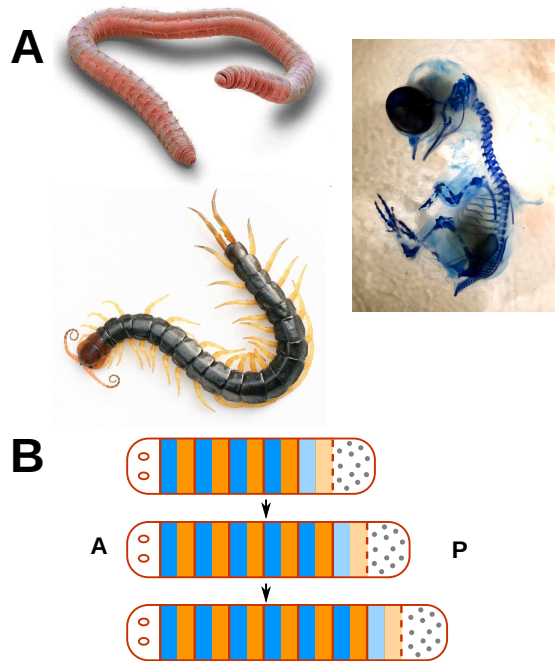


Figure 1.1. Segments in three bilaterian clades A) Manifestations of segments in the three animal clades that display overt segmentation: the external cuticle rings (annuli) of an annelid worm; the clearly visible, leg-bearing segments of the centipedes; and the spinal column in a chick embryo. (Centipede: Yasunori Koide - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=16029265>, annelid worm: myscienceacademy.com, chick: Larsson lab, McGill university) B) a schematic depiction of sequential segmentation. Left is anterior. The undifferentiated growth zone is depicted with the grey dots.

However, despite decades of experimental and theoretical research on segmentation and its evolution, there are still many open questions on how it exactly works, and why it evolved to work this way. For instance, how are the oscillations translated into somites or segments (Beaupeux and François, 2016, Cotterell *et al.*, 2015)? How did sequential segmentation evolve into a simultaneous mechanism like that of *Drosophila* (El-Sherif *et al.*, 2012, Peel, 2004)? Was the common ancestor of bilaterians segmented, and what did that mechanism look like (ten Tusscher, 2013)? New experiments and theoretical models continue to challenge and expand our notions on segmentation in different animals (Clark and Akam, 2016, Cotterell *et al.*, 2015, Shih *et al.*, 2015, Tsiaris and Aulehla, 2016). In this thesis, we use a variety of computational models to address a broad range of developmental and evolutionary questions regarding the how and why of sequential segmentation.

1.2 The how and the why of development, evolution and modelling

In order to understand how models can aid the study of segmentation, we have to discuss the general questions of development and evolution. In a sense, the type of questions is the same regardless of the specific developmental process: the same questions can – and have been – applied to e.g. butterfly and fruit fly wings (Lack *et al.*, 2016, McMillan *et al.*, 2002) and tooth development (Salazar-Ciudad and Jernvall, 2010).

During embryonic development, interactions between a genetic program and molecules, cells, and tissues give rise to a phenotype. On a subcellular level, proteins, RNAs and other molecules form an interaction network that changes the expression of genes within a cell (cell state) as development progresses. Changes or refinement of gene expression may alter the chemical and biomechanical properties of a cell by regulating cytoskeletal or adhesion proteins, or lead to changes in the division behaviour. At the tissue level, extracellular cues such as cell-cell communication and diffusible proteins (morphogens) provide feedback to the intracellular network, coordinating cell states over larger distances: this can cause distinct spatial patterns of different cell states to form across the tissue. The physical interactions between cells may then lead to morphogenesis: the acquisition of tissue shape through cell movement and divisions. By changing the environment in which cells find themselves, morphogenesis may feed back on (the network responsible for) cell state, changing or refining it until a stable tissue state is reached.

These multi-scale, multi-feedback interactions make it hard to disentangle *how* a developmental mechanism generates a particular trait (Jaeger and Sharpe, 2014). Computational models can be a powerful tool for understanding developmental mechanisms, by allowing us to first investigate the independent action of the different components and then to combine them incrementally to study what higher-level properties emerge. For instance, a model of the self-organisation of the slime mould *Dictyostelium discoideum* showed how a few basic processes (signalling, cell differentiation and extracellular matrix production) cooperate to produce the complex movements both during slug migration and the formation of the fruiting body (Marée and Hogeweg, 2002, Marée and Hogeweg, 2001). A multiscale model of somitogenesis was able to identify inconsistencies and gaps in the available knowledge by integrating several models of subprocesses (Hester *et al.*, 2011). If the underlying molecular details of the developmental process are not yet known, a more coarse-grained model may be useful for generating new hypotheses – like the Cooke and Zeeman model mentioned earlier (Cooke and Zeeman, 1976).

As Dobzhansky put it, “Nothing in biology makes sense except in the light of evolution” (Dobzhansky, 1973), and embryonic development is no exception – evolution shapes the developmental process. To study this interaction, we take the evolution of the embryonic structure itself as a given, and ask how its developmental mechanism evolved: how did incremental changes lead to the complex processes we see today? And – perhaps even more interesting – *why* did this particular developmental mechanism evolve to form this particular structure? From this perspective, we investigate what properties of the developmental mechanism gave a selective advantage over the other possible mechanisms that generate *the same* structure. For instance, one developmental mechanism may have a greater robustness against gene expression noise or environmental disturbances than other mechanisms. However, because development is part of a complex genotype-to-phenotype map, a developmental process also influences the future evolutionary potential of a multicellular organism.³ For example, some mechanisms may be more evolvable: they can easily be changed such that potentially favourable new phenotypes emerge. Thus, as has been noted before (Gissis *et al.*, 2011), Dobzhansky’s statement is also true in a reversed form: nothing in evolution makes sense except in the light of development, for multicellular organisms at least.

Our current knowledge of the evolution of developmental processes is based on studying extant animals and a very sparse fossil record. We therefore do not know what ancestral developmental processes did not stand the test of time, or existed in poorly fossilisable animals. While evolutionary models cannot prove that a simulated mechanism is physically feasible or that a particular evolutionary trajectory was taken, they are useful for assessing the likelihood of evolutionary outcomes under various conditions. There are two computational tactics for evo-devo questions: the ensemble approach and evolutionary simulation. In an ensemble approach, either all possible topologies of small gene networks, or a large collection of randomly generated networks, is assessed for the ability to generate a certain phenotype – a striped pattern for instance (Cotterell and Sharpe, 2010, Jiménez *et al.*, 2015, Solé *et al.*, 2002). The advantage of this approach is that a wide range of genotypes can be assessed, exposing many, if not all, possible patterning mechanisms. In the context of segmentation, these approaches have shown that there is a limited number of stripe-generating mechanisms, that differ in their robustness and evolvability; within their mutational neighbourhood, they have different alternative phenotypes (Jiménez *et al.*, 2015). Furthermore, these mechanisms are not connected in genotype space: they cannot mutate from one to the other without losing the selected-for phenotype along the way (Cotterell and Sharpe, 2010). These outcomes suggest a role for historical contingency in the evolution of developmental mechanisms. Still, it remains to be seen how well these outcomes apply to the behaviour of actual biological networks, which tend to contain many genes and interactions, and may therefore more easily allow

³But remember that this complex map itself is evolved! Evolution shapes the process that shapes evolution, not only in development (Hogeweg, 2012)

for connections between distinct regions of the genotype-phenotype map.

In this thesis, we will only use evolutionary simulations to answer our “why” questions. In contrast to ensemble approaches, each evolutionary simulation starts with a population of individuals, each carrying a single randomly generated genotype that is subject to mutation and selection: *in-silico* evolution. We assume that having a particular embryonic structure confers fitness, which is formalised in the fitness criterion: the better the mechanism (encoded by the genotype) can generate the structure, the higher the fitness of the individual. We then study what types of mechanism can evolve, given the embryonic structure that we select for. Because these models simulate the process of evolution, they have the advantage that a perfect “fossil” record is preserved of earlier mechanisms and the routes between them. Multiple simulations can be run to assess the likelihood of evolving a certain mechanism. By giving the evolutionary model sufficient degrees of freedom, mechanisms may evolve that match real-world examples, as well as alternative mechanisms that only evolve *in silico*; these can then be compared with regards to evolvability and robustness (Salazar-Ciudad *et al.*, 2001a, ten Tusscher and Hogeweg, 2011). The evolved gene expression networks can also be dissected in detail, assessing how certain network modules shape the developmental mechanism and the course of evolution (Fujimoto *et al.*, 2008, Kohsokabe and Kaneko, 2016). Additionally, the selection pressures and starting conditions (what genes are expressed, what other developmental processes are already present) can be varied to assess how these shape the likelihood of evolving particular mechanisms. For instance, a study on circadian clocks demonstrated that a particular combination of seasonality in the day-night rhythm and environmental noise is required for the evolution of realistic, complex circadian clock networks (Troein *et al.*, 2009).

In summary, computational models are invaluable tools for answering the mechanistic “how” questions of developmental processes and the proximate and ultimate “why” questions of their evolution. Of course, theoretical models still require experimental data to provide input for basic assumptions and details on the process, and to test predictions that follow from the model. Below, we will therefore discuss what both experiments and models have already elucidated about sequential segmentation.

1.3 Modes of segmentation

As briefly mentioned earlier, there are roughly two ways in which segments are generated during development: simultaneously and sequentially. In most segmented animals, the segments are generated sequentially from a posterior zone, in a rhythmical anterior-to-posterior progression. The exact process however differs between clades; the timing

between axial elongation and segmentation can differ, as well as the mechanism which specifies segment boundaries and segment polarity.

In vertebrates, somites are formed from the paraxial, presomitic mesoderm (PSM) that is laid down on either side of the notochord during gastrulation (Bénazéraf and Pourquié, 2013). Before somites split off the PSM, they are first specified with a gene expression pattern in the anterior PSM. This pattern is generated by a segmentation clock with the Notch pathway as the main oscillating module in all vertebrates, and Delta-Notch signalling synchronises these oscillations between neighbouring cells (Hubaud and Pourquié, 2014). Gradients of Wnt and FGF emanating from the posterior keep the PSM in an undifferentiated, oscillating state; FGF is counteracted in the anterior by a gradient of retinoic acid (RA), which is produced by the somites (Aulehla and Pourquié, 2010). The clock is fastest in the posterior PSM and slows down towards the anterior (a frequency gradient), causing the appearance of a kinematic wave of gene expression travelling across the PSM. In the anterior, this wave defines the pre-pattern of somite boundaries and rostro-caudal polarity within the somites by upregulating other genes⁴. Oscillations eventually cease before the somite buds off from the PSM (Hubaud and Pourquié, 2014).

In short-germ insects like *Tribolium*, the segmentation process bears considerable similarity to somitogenesis. The pair-rule genes that define segments display oscillations in the posterior unsegmented zone (Brena and Akam, 2013, Sarrazin *et al.*, 2012); with regular periodicity a stripe emerges from this zone, travelling a short distance before halting (El-Sherif *et al.*, 2012). Unlike in vertebrates, this stripe splits in two to generate the definitive segment pattern. Furthermore, while in vertebrates the elongation of the main body axis precedes somitogenesis, in *Tribolium* this occurs in the region of stripe formation anterior to the posterior undifferentiated zone (Nakamoto *et al.*, 2015).

In annelids, while segments are created in an anterior-to-posterior order, no evidence has so far been found for a segmentation clock. Segments are generated via divisions of posterior cells called teloblasts (Balavoine, 2014, Shimizu and Nakamoto, 2001). In some annelids, daughters of these teloblasts proceed with stereotyped divisions to contribute specific parts of a segment; in others the segmentation mechanism is unknown (Balavoine, 2014, Shimizu and Nakamoto, 2001). Segment patterning thus seems to happen through asymmetric division rather than oscillatory gene expression (Shimizu and Nakamoto, 2001).

The simultaneous mode of generating segments has so far only been found in long-germ insects, and has been studied in detail in the model species *Drosophila* (Peel, 2004). Segmentation occurs very early in development, before gastrulation and axis extension.

⁴The point at which this determination takes place is commonly called the wavefront or determination front

A cascade of gene regulation, starting from gradients of maternally deposited mRNA, subdivides the egg into progressively narrower regions until the repeated pattern of segmentation genes is specified (Ingham, 1988, Peel, 2004). Each stripe is therefore created by a different combination of activating and inhibiting factors, in contrast to clock patterning which repeats the same process for each segment.

Most experimental and modelling efforts thus far have focused on *Drosophila* and vertebrate segmentation. Models of *Drosophila* patterning have for instance elucidated the interactions between the gap genes, which translate the maternal gradients into a number of distinct regions along the egg (Crombach *et al.*, 2012, Jaeger *et al.*, 2004). In this thesis, we will mainly focus on sequential segmentation. Below, we discuss some of the major open questions regarding segmentation, axis extension and the evolution of segmentation, and the modelling approaches that have thus far been used to these questions.

1.3.1 The clock-to-stripe transition

Models have been used to study vertebrate somitogenesis on all levels of organisation: how genetic interactions lead to gene expression oscillations (Goldbeter and Pourquié, 2008, Lewis, 2003, Schröter *et al.*, 2012, Tiedemann *et al.*, 2007), how oscillations are synchronised between cells (Herrgen *et al.*, 2010, Lewis, 2003, Morelli *et al.*, 2009, Terry *et al.*, 2011, Tiedemann *et al.*, 2014), and how somites take shape (Dias *et al.*, 2014, Grima and Schnell, 2007, Hester *et al.*, 2011). Despite the tremendous progress that has been made, at the larger scale of somite determination it is still heavily debated how somite boundary definition occurs, how this depends on gene expression oscillations and whether and how the oscillation phase is translated to intrasomite polarity. Several functions have also been proposed for the frequency gradient that causes the travelling wave of gene expression across the PSM: whether and how it determines somite polarity (Jaeger and Goodwin, 2001, Murray *et al.*, 2011), is required for somite boundary formation (Beaupeux and François, 2016, Harima and Kageyama, 2013), or for somite size scaling (Lauschke *et al.*, 2013). Finally, FGF and Wnt have been thought to determine the position of the determination front, since adding or depleting these signals leads to smaller or larger somites, respectively (Bajard *et al.*, 2014, Dubrulle *et al.*, 2001). However, the necessity of the signalling gradients for determining the wavefront position has recently been questioned (Cotterell *et al.*, 2015, Mallo, 2016, Murray *et al.*, 2011).

At the time of the original “clock and wavefront” model, the existence of internal somite polarity or travelling waves of gene expression were not yet known; Cooke and Zeeman therefore only needed to explain the formation of discrete blocks of somites (Cooke and Zeeman, 1976). They assumed that as long as a block of cells differentiates at the same time, they can become separated as a somite from the rest of the PSM via

differential adhesion. They proposed that a timing gradient (the wavefront, probably informed by an actual morphogen gradient) retracts across the tissue towards the posterior, and determines for each cell at what moment it starts to differentiate. An intracellular oscillator interacts with this wavefront to transform the smoothly retracting signal into a stepwise process: one phase of the clock speeds up this wavefront and the other slows it down, so that an entire block of cells becomes a somite at the same time (Cooke and Zeeman, 1976).

Subsequent models attempted to explain somite determination in conjunction with gene expression waves (i.e. a frequency gradient) and/or intrasomite polarity patterning. Hans Meinhardt was the first to suggest that the oscillation phase becomes memorised upon somite determination, informing both somite polarity and inter-somite boundaries (Meinhardt, 1986). In the mechanism that he proposed, the posteriorly retracting determination front, the travelling wave and the memorisation of the oscillation phase arise as emergent properties due to interactions between two cell states. These states are auto-activating and inhibit each other locally, but activate each other on a larger length scale, causing non-cell-autonomous oscillations throughout the tissue (Meinhardt, 1986). In the anterior, the boundary of a previously formed half-somite (with either anterior or posterior identity) is responsible for the formation of a new half with opposite identity. A gradient of morphogen is thus only required for generating the first somite boundary, after which somite formation progresses independently.

Later models adopted Meinhardt's notion that the anterior or posterior identity of a cell within a somite results from "freezing" the phase of the oscillator, but incorporated cell-autonomous oscillators rather than reaction-diffusion. Some of these models assumed that the frequency gradient responsible for forming the travelling waves, is also a prerequisite for the freezing of oscillations: it brings the oscillation frequency to zero in the anterior PSM, stopping the oscillations entirely and thereby forming a stable somite pre-pattern (Jaeger and Goodwin, 2001, Murray *et al.*, 2011). In other models, a threshold concentration of FGF or Wnt determines the point at which the clock can switch on somite polarity genes, before oscillations halt completely (Hester *et al.*, 2011, Mazzitello *et al.*, 2008). In both cases, the clock phase is directly memorised as somite polarity, and somite determination occurs in a smooth, cell-by-cell manner rather than block-wise. The frequency gradient itself has been proposed to be the result of cell ageing (Jaeger and Goodwin, 2001), cell-cell synchronisation with non-oscillatory cells in the anterior (Murray *et al.*, 2011), or interactions between the morphogen gradient Wnt or FGF and the clock (Gibb *et al.*, 2009, Hester *et al.*, 2011, Mazzitello *et al.*, 2008).

Recent experiments demonstrated that oscillation frequency decreases by only 50% before the somite is formed and oscillations stop (Niwa *et al.*, 2011, Shih *et al.*, 2015), while somite boundaries are established earlier in the PSM (Akiyama *et al.*, 2014).

Moreover, there is evidence that the oscillator phase is not directly memorised into somite polarity, but that somite polarity formation is a “secondary” process. First the somite becomes specified as a nearly uniform block by upregulation of differentiation genes (e.g. *Mesp2* and *Ripply*), then this block obtains polarity through interactions between these genes (Harima and Kageyama, 2013, Niwa *et al.*, 2011, Oginuma *et al.*, 2008, 2010) (modelled in (Oginuma *et al.*, 2010, Tiedemann *et al.*, 2012)). This discards the possibility that somite boundaries and polarity are formed from the opposite phases of a fully slowed down and frozen oscillator, or even from the memorisation of the oscillation phase upon passing a morphogen concentration threshold. As a consequence, the functional relevance of the frequency gradient and clock phase for somitogenesis is again an open issue.

Recently, two models were proposed that fundamentally differ from the above-described clock-and-wavefront models (Beaupeux and François, 2016, Cotterell *et al.*, 2015, Lauschke *et al.*, 2013). The first model proposes that two oscillators function in the PSM, one which maintains a constant frequency and one which slows down towards the anterior, so that the two oscillators shift out of phase in the anterior (Lauschke *et al.*, 2013). They proposed that this shift is detected by the rest of the gene regulatory network and that when a certain phase difference is reached in the anterior PSM, a somite is formed (Beaupeux and François, 2016). The frequency gradient is therefore crucial for somite formation in this model. The other model, called PORD (Progressive Oscillatory Reaction Diffusion model), revives the old idea of Meinhardt, but with different interactions between the two genes (cell states in the Meinhardt model): one is an activator and the other an inhibitor. This interaction resembles a Turing system, but additional input from an FGF gradient prevents stripes from spontaneously emerging everywhere in the tissue at the same time. Instead, the amount of activator and inhibitor oscillates and travelling waves are formed, while oscillation arrest in the anterior is due to diffusion of the repressor from the last-formed stripe (Cotterell *et al.*, 2015). Thus, as in the Meinhardt model, the wavefront of somite patterning arises as an emergent property, rather than as a direct result of the FGF gradient. However, this model is capable of generating cell-autonomous oscillations (as has been observed *in vivo* (Webb *et al.*, 2016)), whereas in the Meinhardt model, diffusion between cells was required.

While the new two-oscillator model suggests that the frequency gradient has an important function in somitogenesis (Beaupeux and François, 2016), it is important to consider that not all animals generate segments under a frequency gradient. In the cephalochordate *Amphioxus* (a close relative to vertebrates) for instance, somites are formed without a travelling wave, and also evolutionary simulations so far yielded mechanisms without a frequency gradient (François *et al.*, 2007, ten Tusscher and Hogeweg, 2011). This shows that a frequency gradient is not always necessary, and in **chapter 2** we investigate under what conditions it does and does not evolve.

Above, we discussed several alternative models for somite determination, and it remains to be seen which one best describes vertebrate somitogenesis. But is there just one correct model, or could some models apply to one species, and others to another? After all, while somitogenesis looks very similar in the different vertebrates, there are also some notable differences. For example, the set of oscillating genes differs greatly between the three main model organisms: zebrafish, chick and mouse (Krol *et al.*, 2011). More differences become apparent in experiments where RA, one of the morphogens pervading the PSM, is removed: somitogenesis becomes asymmetric, but in a different manner in all three species. We will explore these species differences in asymmetric somitogenesis in **chapter 3**, to see if this allows us to discover potential differences in somite determination mechanism.

1.3.2 Axis extension and segmentation

In many animals, the main body axis is formed through convergent extension: the simultaneous narrowing and elongation of tissue. There are several processes that may contribute to axis extension, such as oriented cell divisions, cell shape changes and cell intercalation (Nikolaidou and Barrett, 2005, Tada and Heisenberg, 2012). In the latter process, cells in one row move between the cells of another, reducing the number of cell rows in one direction and increasing them in another (Fig. 1.2A).

The mechanism of axis extension, and its timing with segmentation, differs between clades (section 1.3). In vertebrates, axis extension mostly precedes somitogenesis (Bénazéraf and Pourquié, 2013, Tada and Heisenberg, 2012). The shape of the main axis and some of the PSM is formed through collective cell migration and convergent extension, and later the tailbud produces the remainder of the PSM and neural tissue by adding cells through additional ingression and divisions (depending on the species) (Bénazéraf and Pourquié, 2013, Steventon *et al.*, 2016, Tada and Heisenberg, 2012, Wilson *et al.*, 2009); once the somites become patterned, they no longer elongate significantly. In short-germ insects (*Tribolium*) instead, elongation occurs by convergent extension that is concomitant with segment formation (Nakamoto *et al.*, 2015), which suggests that these two processes could interact. Divisions occur everywhere along the anterior-posterior axis with random orientation, making it unlikely that they contribute significantly to elongation (Nakamoto *et al.*, 2015). In *Drosophila*, axis extension follows well after segmentation.

Convergent extension requires tissue polarity information to instruct the direction of convergence and extension. In vertebrates, the Wnt signalling pathway and planar cell polarity (PCP) pathway are involved in giving each individual cell an internal polarity that is coordinated with the overall tissue polarity (Fig. 1.2B). In insects, the segmentation genes are required for instructing tissue and cell polarity for convergent extension

(Benton *et al.*, 2016, Irvine and Wieschaus, 1994, Mao and Lecuit, 2016, Paré *et al.*, 2014, Zallen and Wieschaus, 2004). How such tissue polarity is set up is a field of study in itself, but we want to mention one general example, used for both plant and animal cell polarity (Abley *et al.*, 2013). This computational model showed how a cell-autonomous polarity generating mechanism, combined with cell-cell coupling and a tissue polarity organiser, can generate coordinated tissue polarity (Abley *et al.*, 2013).⁵ Models and experiments on convergent extension have focused on tissues in which polarity is predefined, addressing questions such as: what cellular mechanisms could explain the observed cell movements? How robust is the mechanism to imperfect tissue polarity information, and what is the effect of feedback of convergent extension on tissue polarity?

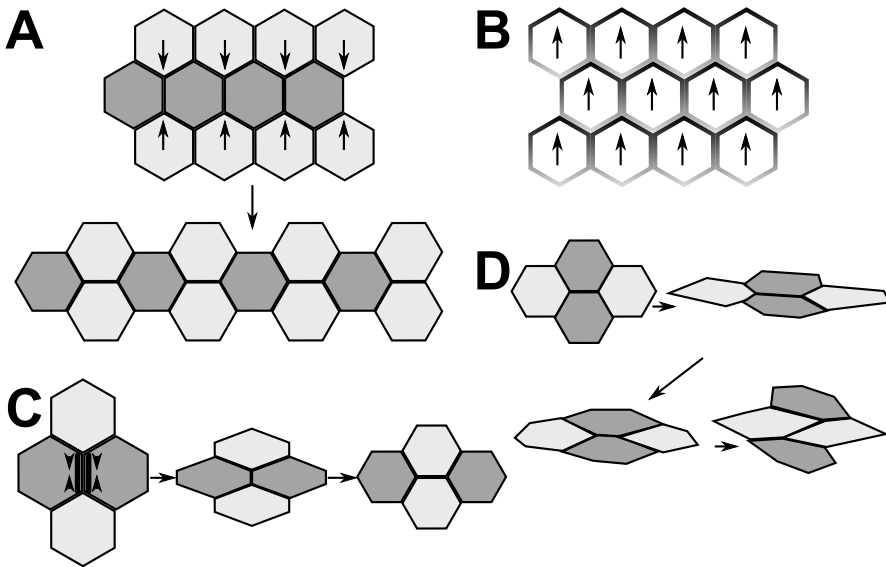


Figure 1.2. Mechanisms of convergent extension via cell intercalation A) General principle of intercalation. One row of cells inserts itself between another, leading to extension of the tissue in one direction and narrowing in another. B) Schematic representation of tissue polarity. Each cell has its own internal polarity, which often manifests itself in different proteins at the cortex (picture as in (Abley *et al.*, 2013)). C) Convergent extension via junctional tension: the boundary indicated with arrows has increased actomyosin flow (as in (Honda *et al.*, 2008, Tada and Heisenberg, 2012)). D) Convergent extension through cell elongation (without lamellipodia pulling; as in (Honda *et al.*, 2008)).

⁵This is reminiscent of some of the somitogenesis models, with cell-autonomous oscillators coupled for synchronisation and directed by organising gradients.

A number of mechanisms has been proposed that could drive convergent extension and intercalation in different animals. For instance, in *Drosophila*, cell contacts with a certain orientation have an increased actomyosin flow at the cortex (Bertet *et al.*, 2004, Rauzi *et al.*, 2010); computational modelling showed that the increased tension generated by this flow is sufficient for cell intercalation (Honda *et al.*, 2008, Rauzi *et al.*, 2008) (Fig. 1.2C). In the *Xenopus* notochord instead, cells become elongated in the direction perpendicular to the direction of elongation (Tada and Heisenberg, 2012). It was proposed that these elongated cells adhere more strongly at their elongated sides than at the tips, which causes them to align their long axes and intercalate (Zajac *et al.*, 2003). Other models showed that protrusive activity and elongation by itself is sufficient for convergent extension, as long as this elongation is restricted on two sides of the tissue (Weliky *et al.*, 1991) or if the elongation is alternated with cell relaxation (Honda *et al.*, 2008) (Fig. 1.2D). In other models instead, cell protrusions (lamellipodia or filopodia) actively pull on other cells, dragging them closer and causing neighbour exchanges (Belmonte *et al.*, 2016, Brodland, 2006). One of these models also demonstrated that imperfect tissue polarity (where cells have slightly deviating polarity with respect to the overall tissue polarity) may still yield correct convergent extension; individual deviations from the correct polarity are buffered by the feedback from surrounding cells (Belmonte *et al.*, 2016).

In short, convergent extension requires cell and tissue polarity information, and there exist several alternative mechanisms which can translate this polarity into cell movement and tissue reshaping. Existing models consider the tissue to be homogeneous: each cell has its individual polarity, but all cells are identical and the tissue is not otherwise patterned. However, we know that at least in insects, a segmented pattern is already present during convergent extension. In *Drosophila*, it has been shown that precautions are in place that prevent tissue reshaping from disrupting the segmentation pattern (Monier *et al.*, 2011). This suggests that in absence of such constraints, convergent extension may have a profoundly disturbing effect on existing tissue patterns. On the other hand, segmental patterning can be a source for the polarity information that instructs convergent extension (Irvine and Wieschaus, 1994, Zallen and Wieschaus, 2004). We will investigate the potential interplay between segmentation and convergent extension in **chapter 5**.

1.3.3 Evolution of segmentation

The evolution of segmentation in bilateral animals has been a topic of ongoing debate for the last two decades, and it is currently unresolved whether the bilaterian ancestor was segmented, or segmentation evolved independently two or three times (Balavoine, 2014, Balavoine and Adoutte, 2003, Budd, 2001, Seaver, 2003). Arguments in favour of a segmented ancestor cite the similarities in segmentation dynamics in the three clades:

the anterior-to-posterior progression and fact that all clades share genes involved in segmentation (Balavoine, 2014, Budd, 2001). There are other clades which display partial segmentation in just one or a few organs, so segmentation could be more widespread than usually thought (Balavoine, 2014, Budd, 2001) (Fig. 1.3). Moreover, the fossil record has brought up very early examples of bilaterian-looking animals with segment-like structures, supporting the idea of a segmented ancestor (Gold *et al.*, 2015). However, there are also considerable differences in the gene set used for segmentation, and the overlap in the genes involved in segmentation could be due to a limited toolbox and convergent evolution (Chipman, 2010). While the mechanism used in vertebrates and arthropods looks similar, segmentation occurs in different tissue layers and with a different temporal order of axis extension versus segmentation, as discussed above (section 1.3.2). The mechanism in annelids differs even more (see section 1.3). Another long-standing issue is the evolution of long-germ, simultaneous segmentation (*Drosophila*-like). This segmentation mode likely evolved a number of times from short-germ, sequential segmentation (Peel, 2004), but the evolutionary trajectory from one mode to the other is so far unresolved.

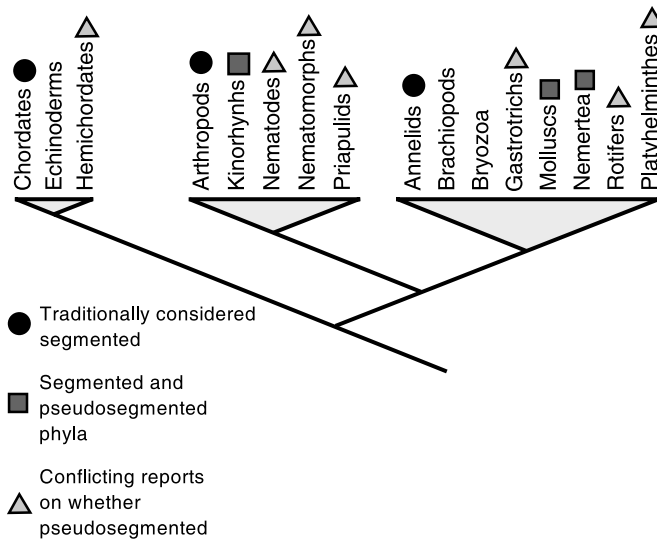


Figure 1.3. Tree of bilaterians Adapted from (Hannibal and Patel, 2013). When also pseudosegmented phyla (having some form of repetition along the main body axis) are taken into consideration, segmentation is not a rare trait.

Computational models addressing the evolution of segmentation collectively show that only a few distinct classes of mechanism evolve for generating stripes, and the type of mechanism that emerges depends strongly on the morphogen dynamics and the fitness criterion (reviewed in (ten Tusscher, 2013)). Simulations with more specific fitness criteria (precise number and location of stripes) and/or a static morphogen gradient yield hierarchical or emergent mechanisms that generate all stripes at roughly the same time (François *et al.*, 2007, Fujimoto *et al.*, 2008, Kohsokabe and Kaneko, 2016, Salazar-Ciudad *et al.*, 2001a), with hierarchic networks being more robust to mutations (Salazar-Ciudad *et al.*, 2001a, ten Tusscher, 2013). Although these do not truly resemble *Drosophila*-like simultaneous segmentation, their evolution for specific fitness criteria is in line with the idea that a hierarchic network evolved secondarily, having to maintain the position of existing segments. Instead, simulations with a sweeping wavefront of morphogen often evolve sequential segmentation using gene expression oscillations, as the wavefront provides a timing mechanism that can transform the temporal oscillations into a spatial pattern (François *et al.*, 2007, ten Tusscher and Hogeweg, 2011). Compared to alternative mechanisms that evolved, the sequential mode typically evolves faster and is more evolvable and robust to gene expression noise than alternative mechanisms (François *et al.*, 2007, Fujimoto *et al.*, 2008, ten Tusscher, 2013, ten Tusscher and Hogeweg, 2011).

In short, these latter simulations show that oscillatory sequential segmentation is a likely and robust outcome of selection for segments. Although not conclusive, this supports the idea that sequential segmentation could have evolved independently in the three clades. However, these results were obtained assuming that an anterior-to-posterior timing mechanism is already present in the shape of a retracting wavefront, which mimics posterior growth with a morphogen gradient. Thus, the question remains whether sequential segmentation still evolves so readily if such posterior growth and a wavefront also have to evolve. In **chapter 4**, we therefore extend an evo-devo model to evolve segmentation and a growing axis. With this model, we investigate under what prior conditions and selection pressures sequential segmentation with posterior growth becomes a likely outcome of evolution.

1.4 Overview

This thesis consists of two parts of each two chapters. In the first section, we study the clock to stripe transition, with emphasis on the more thoroughly studied process of vertebrate somitogenesis. In the second part we investigate the developmental and evolutionary interplay between segment patterning and axis formation. The chapters divide each part into an evolutionary and a developmental question.

Part I: clocks, waves and stripes Arguably, the most striking feature of vertebrate somitogenesis is the travelling wave of gene expression, traversing the PSM from posterior to anterior for each cycle of somite formation. As we discussed in section 1.3.1, a number of functions has been proposed for the frequency gradient that causes this travelling wave. In **chapter 2**, we put some of these proposals to the test by running evolutionary simulations with realistic vertebrate tissue properties such as posterior growth and a morphogen gradient (using the framework of (ten Tusscher and Hogeweg, 2011)). We then investigate under what circumstances a frequency gradient evolves.

From a distance, somitogenesis looks very similar in all vertebrates. When instead somite symmetry is disrupted due to the removal of RA, the resulting asymmetry looks different in each of the three most frequently studied model species – zebrafish, chick and mouse. In **chapter 3**, we take a simple phenomenological model (as used in (Morelli *et al.*, 2009)) to study these different asymmetric phenotypes. We then adapt this model based on experimental data on different species, to investigate whether differences in the somite formation mechanism can explain the observed differences in the asymmetric phenotype.

Part II: clocks, stripes and shapes In **chapter 4**, we return to the observation that most animals use a superficially similar looking segmentation process involving posterior addition of segments. We investigate what factors could have caused this prevalence of sequential segmentation by extending the evo-devo model with regulation of divisions, to evolve segmentation and axis formation at the same time. We vary the presence of a persistent tissue polarity signal (in the form of a posteriorly emanating morphogen) and additional selection pressures, and assess under what conditions sequential segmentation is most likely to evolve.

In section 1.3.2 we discussed how convergent extension relies on tissue patterning and polarity. Previous models assumed that the tissue undergoing convergent extension is homogeneous, each cell having the same identity as its neighbours. Since convergent extension takes place in segmented tissue in some species, the cell rearrangements may interfere with this pattern. We investigate whether different convergent extension mechanisms can disrupt a segmented tissue pattern – and how this could be prevented – in **chapter 5**. Incidentally, we find a new mechanism to generate convergent extension.

These chapters cover a wide variety of questions concerning segmentation. In **chapter 6** we will summarise and integrate our findings, discussing their implications for our understanding of body axis segmentation and for promising directions of future work.

Part I

Clocks, waves and stripes

2

Analysing the influence of wavefront dynamics on clock evolution

RENSKE M.A. VROOMANS, PAULIEN HOGEWEG AND KIRSTEN H.W.J. TEN
TUSSCHER

Manuscript in prep.

Abstract

The predominant mode of segmentation in all three segmented animal clades is sequential segmentation, where segments are formed in an anterior to posterior manner from a posterior undifferentiated zone. In short-germ insects, segments form close to the posterior zone using a seemingly simple genetic oscillator. In contrast, vertebrate segmentation is characterised by a complex oscillator that generates travelling waves across an extended presomitic mesoderm (PSM) beyond which segments are laid down. These particular properties of vertebrate segmentation have been suggested to contribute to robustness, but this hypothesis has thus far not been thoroughly tested.

Here we extend previously used computational models for the evolution of segmentation to investigate factors that influence the evolution of segmentation clock properties. We include an explicit posterior division zone coupled with a realistic, decay-driven morphogen gradient. Furthermore, we vary the presence of cell-cell signalling, gene expression noise and the developmental time available for segment patterning. The mechanisms that evolve in our simulations are compared in terms of oscillator complexity, oscillator dynamics and evolutionary trajectories.

We find that oscillatory segmentation evolves significantly faster in the presence of a steep morphogen gradient, and results in qualitatively different evolutionary outcomes compared to segmentation evolved with a shallow gradient. In absence of noise, under a steep morphogen gradient often damped oscillators evolve. In contrast, for shallow gradients persistent oscillators dominate, and sloped frequency profiles resulting in propagating waves regularly evolve. Still, upon changing morphogen decay, we observe ready evolutionary adaptation to the altered gradient, including associated oscillator characteristics. Incorporating gene expression noise leads to a substantial increase in the evolution of persistent oscillators for steep morphogen gradients and of sloped frequency profiles for shallow gradients, supporting their roles in increasing developmental robustness. Surprisingly, we find that oscillator complexity is not correlated with oscillator slowing, suggesting that these properties may not have evolved simultaneously. Instead, oscillator complexity may have evolved first under a steep morphogen gradient, in response to the requirement for a robust, persistent oscillator; later a transition to shallow morphogen gradients may have led to the evolution of a frequency gradient.

2.1 Introduction

Evolutionary developmental biology aims to understand how the complex developmental programs arose that shape multicellular organisms, and why particular mechanisms evolved to create the observed gene patterns and tissue dynamics. Segmentation, the division of the body axis into repeated units, is one of the most intensely studied patterning processes, both on the level of the generating mechanism and from the broader evolutionary perspective; it is considered a major evolutionary innovation. There are only three lineages of animals with a clearly segmented body plan: annelid worms, arthropods and chordates (Davis and Patel, 1999, Peel and Akam, 2003). It is still a major unsettled issue whether the common ancestor of these three lineages had a segmented body plan which was subsequently lost in the majority of lineages, or the three segmented lineages evolved segmentation in parallel (Balavoine and Adoutte, 2003, Budd, 2001, Chipman, 2010, Couso, 2009, Davis and Patel, 1999, Minelli and Fusco, 2004, Seaver, 2003, Tautz, 2004). While similarities in used genes have been taken as support for a single origin, an alternative case can be made for the likelihood of parallel recruitment due to the limited size of the developmental toolkit (Chipman, 2010). Parallel evolution is further supported by specific differences in the segmentation mechanism used in the three different lineages.

In most segmented animals, segments are generated from a posterior zone and laid down in a regular anterior-posterior sequence. Sequential segmentation has been studied in most detail in vertebrates, where somites emanate sequentially from a posterior undifferentiated zone, the presomitic mesoderm (PSM). A wavefront retreating across the PSM transforms oscillatory gene expression into a spatial pattern of segments (for review, see e.g. (Hubaud and Pourquié, 2014)). Most arthropods appear to deploy a similar sequential segmentation mode although the molecular details underlying oscillations and the transformation to segments are still poorly understood. While vertebrates and sequentially segmenting arthropods specify segments before the intra-segmental fates of cells is laid down, in annelids this order is reversed. Thus, while the segmentation process in annelids is also sequential, cell lineages with a different future fate appear to undergo distinct parallel sequential segmentation processes before fusing into segments (Shankland and Seaver, 2000).

To unravel the evolutionary processes that have shaped these sequential segmentation programs, several complementary approaches exist: bioinformatic comparisons, experimental developmental comparisons and *in silico* evolutionary simulation studies. While the first two approaches are essential for establishing how organisms pattern their body and how this process has diversified across different lineages, they are less suited for answering why these particular mechanisms evolved. With evolutionary simulation we can rerun the evolutionary process again and again to determine the likelihood of specific

evolutionary outcomes, and which factors may bias evolution towards a certain solution. This approach is also useful for assessing the probability of parallel evolution.

Earlier evo-devo simulation studies demonstrated that the oscillation-driven sequential segmentation observed in vertebrates, arthropods and annelids, readily evolves from scratch (François, 2014, François *et al.*, 2007, ten Tusscher, 2013, ten Tusscher and Hogeweg, 2011, Vroomans *et al.*, 2016), provided that a posterior signalling centre has previously evolved (Vroomans *et al.*, 2016). Furthermore, these studies have shown that this type of segmentation mechanism should be expected to evolve due to its increased robustness and future evolutionary potential relative to alternative strategies. However, thus far the evolutionary forces that have shaped particular more detailed aspects of sequential segmentation have remained unresolved. In vertebrate segmentation, the oscillations are generated by a complex regulatory network of three coupled oscillator motifs (Aulehla and Pourquié, 2008, Dequéant *et al.*, 2006, Goldbeter and Pourquié, 2008). One can speculate that coupling multiple oscillators contributes to the robustness of oscillation period by generating redundancy. Similarly, so-called kinematic waves of gene expression have been observed that traverse the unsegmented region from posterior to anterior in both vertebrate and arthropod segmentation. This dynamics reflect the gradual slowing down of oscillations before they arrest into segments (Dequéant and Pourquié, 2008, Jaeger and Goodwin, 2001, Kaern *et al.*, 2000, Palmeirim *et al.*, 1997). Again, one can speculate that such a sloped oscillation frequency profile enhances the robustness of the segmentation process (El-Sherif *et al.*, 2014). Thus far, these hypotheses have not been rigorously tested. As a consequence, it remains unclear why these particular properties of sequential segmentation evolved.

Here, we take the first steps required to answer these questions. First, we develop a more detailed modelling framework for studying the evolution of sequential segmentation, to enable the evolution of more complex oscillators and different types of frequency profiles. Rather than running a moving wavefront across a preformed tissue, we explicitly model a posterior growth zone characterised by cell division and high expression of the morphogen gene. Due to slow decay of the morphogen protein in cells outside of the growth zone a decay-driven morphogen gradient is generated. We have previously demonstrated that the presence of such a posterior signalling centre and gradient forms a major determinant for the evolution of sequential segmentation (Vroomans *et al.*, 2016). In this study, we vary the rate of decay to test the impact of gradient length scale and slope on the type of segmentation that evolves. Additionally, we investigate the influence of gene expression noise and the developmental time available for segment patterning. To efficiently study for large numbers of simulations whether or not complex oscillators or a sloped frequency profile evolves, we build an automated alternative to manually analysing the details of network topologies and dynamics, as was done in previous studies. Our analysis pipeline determines the number and size of loops present in the evolved

networks, as well as performing a detailed Fourier analysis of the oscillations in gene expression as a function of morphogen concentration.

We demonstrate that oscillation-driven sequential segmentation with complex multi-loop networks, and / or a sloped frequency profile that generates kinematic waves, automatically evolves in a subset of simulations. We find that different morphogen decay rates, resulting in differently sloped morphogen gradients, lead to somewhat different outcomes, which are amplified by the addition of noise. Simulations incorporating slow decay resulting in shallow morphogen gradients often lead to the evolution of persistent oscillations, sloped frequency profiles and complex networks with many and large regulatory loops. In contrast, simulations with fast morphogen decay resulting in steep gradients evolve sequential segmentation faster, requiring a slightly smaller genome and simpler network with fewer loops. However, evolutionary simulations with a steep gradient resulted more often in damped oscillations. A disadvantage of these damped oscillators is their sensitivity to perturbations, and their reduced evolutionary potential for developing a larger body with more segments.

To perform a preliminary assessment of the functional significance of network complexity and sloped frequency profiles we performed evolutionary transition experiments, in which individuals evolved under a steep morphogen gradient were transferred to a situation with a shallow gradient and vice versa, after which evolution continued. The increases or decreases in genome size, loop numbers and occurrence of persistent oscillators and sloped frequency profiles match our earlier results, suggesting that these properties are not evolutionary neutral but tailored to the specific morphogen profile. Finally, we find that sloped frequency profiles may have an important role in mitigating the adverse effects of gene expression noise. In the future we aim to use the methodology developed here to further investigate the evolutionary forces that have driven the formation of complex oscillatory networks and sloped frequency profiles.

2.2 Methods

2.2.1 The model

General setup

We use a individual based model of a population of organisms evolving on a lattice, as has been applied before to evolution of segmentation and domains (ten Tusscher and Hogeweg, 2011) (Fig. 1A). Each organism has a so-called “pearls-on-a-string” genome consisting of genes (transcription factors) and upstream regulatory regions with transcription factor binding sites (TFBS) (Crombach and Hogeweg, 2008). Furthermore, they consist of a one-dimensional row of cells which starts out small and grows during the course

of the individual's development, instead of starting at full length as in previous models (for review, see (ten Tusscher, 2013)). The individuals reproduce in a fitness-dependent fashion, with fitness dependent on the number of segments in their final gene expression pattern.

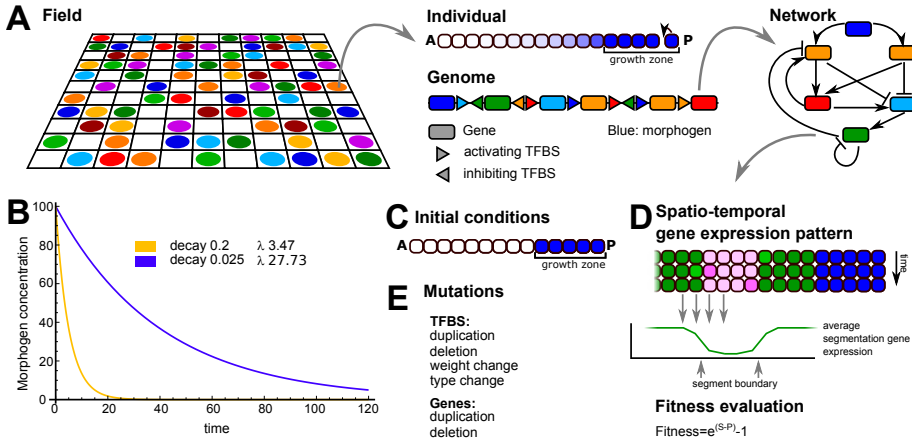


Figure 2.1. Overview of the model. A) The developing individuals live on a 2D lattice. Each individual consists of a row of cells, of which the posterior-most cell divides at regular intervals. Within the growth zone, the morphogen (in blue) is maintained at a high concentration; it decays in cells outside of this zone. The genome of the individual codes for a network of regulatory interactions, which determines the spatiotemporal dynamics of the proteins within each cell (see D). B) The gradients resulting from the different morphogen decay rates (d) used in our simulations. The lambda indicates the position (or time) at which the morphogen concentration is half-maximal, i.e. 50 : $\lambda = \ln(2)/d$. C) The initial conditions for each new individual at the start of its development. There is a growth zone with high morphogen, and a 'head' region without morphogen. D) At the end of development, the expression of the segmentation gene is averaged over a number of timesteps, and from this the segment boundaries are determined. E) The mutational operators acting on the genome.

Individuals

genome, network and genes The genome codes for a gene regulatory network. The genes in the genome form the nodes of the network; the set of TFBS upstream of each individual gene in the genome dictate the incoming regulatory edges of the GRN (Fig. 2.1A). Outgoing edges follow from genes matching the type of the TFBS in front of another gene. The regulatory interactions between genes can be repressive (strength -1) or activating (strength 1). The network governs gene expression dynamics and subsequent protein levels. Gene expression is modelled with ordinary differential equations

as shown in eq. 2.1:

$$\frac{dG_i}{dt} = \text{Max}\left(\frac{A_j^n}{A_j^n + H^n}\right) * \prod_{k=1}^m \left(\frac{H^n}{I_k^n + H^n}\right) * E - \delta * G_i \quad (2.1)$$

Transcription of gene i is determined by the activating genes A_j ($j = 1 \dots l$) with the highest expression (a max function), resulting in a so-called activating OR-gate, while repressive inputs I_k ($k = 1 \dots m$) are multiplied, resulting in a repressive AND-gate (l and m are the total number of activating and repressing inputs for gene i). Note that these choices are somewhat arbitrary, as for both activating and repressive TFs AND as well as OR or even different types of integration may occur; however, the main point here is to incorporate at least partially the highly complex, non-linear integration of TF inputs into gene expression levels. E is the maximum expression level; δ is the degradation rate; H is a Hill constant, the transcription factor concentration level at which half-maximal activation or repression occurs; and n is the Hill coefficient governing the steepness of the transition from low to high gene expression depending on transcription factor concentrations.

There are 16 types of genes, indicated with a number from 0 to 15.

Gene 0 is the morphogen: it is not regulated by any of the other genes, but instead is set to high expression in the cells of the growth zone, while decaying with a predefined rate in the rest of the embryo (Fig.1B). We run simulations with either a large or a small morphogen decay rate, yielding a steep or a shallow morphogen gradient, respectively.

Gene 1 and gene 2 are signalling genes, in a subset of simulations. Both function via direct membrane bound ligand induced signalling (e.g. Delta-Notch signalling). As a consequence they can only affect gene expression dynamics of directly neighbouring cells. Note that expression of these signalling genes does not affect the cell expressing the gene itself.

Gene 5 is the segmentation gene, whose final pattern after development determines the fitness of the organism.

gene expression noise In a subset of simulations, we implemented gene expression noise, as follows. First, we computed the expected gene expression rates based on Eq. 2.1. Next, we computed the actual gene expression rate by sampling from a Gaussian distribution around the expected gene expression rate. Specifically, we assume a Gaussian distribution with a mean equal to the computed expected gene expression rate R_{expr} ($\mu = R_{expr}$, $\sigma = 0.07 * R_{expr}$), and a standard deviation of 0.07 times this mean. Note that by scaling the standard deviation with the mean, the noise which is defined as the standard deviation divided by the mean, is kept constant independent of the mean gene expression rate. We avoid negative gene expression rates by capping any negative gene expression rates due to noise to zero: $R_{actual} = \text{Max}(0, R_{expr} + noise)$.

developmental dynamics Individuals start their development with a short row of 14 cells, where 5 cells form the primordial “growth zone” in which the morphogen concentration is high; in the remaining 9 cells (the “head”), the morphogen is absent (Fig. 2.1C). The other genes have an expression level of 0 in all cells. This means that no gene expression will occur in the anterior-most 9 cells. We ignore the developmental processes generating the head part of the body and their evolution, and focus solely on the developmental processes governing formation of more posterior body parts and their evolutionary history. The posterior-most cell of the growth zone divides at regular intervals, pushing the other cells forward so that they eventually move out of this zone. Once a cell leaves the growth zone, the morphogen protein starts decaying. As a result, a gradient of the morphogen is formed due to the age difference of the cells (Fig. 2.1A,B). (The 4 cells in the growth zone that do not divide are there for cosmetic reasons; it makes it easier to see the dynamics in the growth zone on a time-space plot.) Throughout development, the concentrations of the other proteins (i.e. all except the morphogen protein) are updated according to the genetically specified network interactions (Eq. 2.1). The posterior cells stops dividing after 120 divisions, after which developmental dynamics continue for a while longer such that also the youngest cells reach a low morphogen concentration and can converge on a stable gene expression pattern.

fitness evaluation By the end of development, the expression pattern of the segmentation gene is evaluated to determine the number of segments formed outside the growth zone (Fig. 2.1D). Segments should be at least 7 cells wide, and boundaries between segments should consist of a clear transition of the expression of the segmentation gene from a high to a low level, or vice versa, within 5 cells (similar to earlier definitions (François *et al.*, 2007, ten Tusscher and Hogeweg, 2011)). Given that the tissue grows out to be 134 cells, of which 9 form the head segment and 5 form the growth zone, the maximum number of segments that can be formed is 18. The number of well-formed segments (i.e. fulfilling the above requirements) determines an individual’s fitness. In addition, some penalties are applied. First, we require that at least one gene of each type is present in the genome; if this requirement is not met, the individual is not allowed to reproduce. Second, too-narrow segments are penalised. Third, small fitness penalties are used for gene and TFBS numbers in order to prevent excessive genome growth. Finally, when determining the number of segments, rather than considering the expression of the segmentation gene at the last timestep of development, we average expression of the segmentation gene over the last 100 developmental steps. This averaging helps ensure temporally stable segmental patterning, as it will not reward oscillatory segmentation that fails to converge on a constant spatial pattern. To further ensure stability of the final developmental pattern we apply an additional fitness penalty for the amount variance of the pattern from the average (pattern instability) within these final 100 developmental steps.

The fitness then becomes $e^{\max(0,F)} - 1$, where F is:

$$\begin{aligned}
 F = & \text{nr good segments} \\
 & - \text{nr narrow segments} \\
 & - G * \text{gene nr} \\
 & - T * \text{TFBS nr} \\
 & - U * \text{nr unstable cells}
 \end{aligned}
 \tag{2.2}$$

See table 2.2 for parameter values.

Evolution

initial conditions, mutations and simulations The population is initialised with 50 identical individuals. The population resides on a lattice of size 30x30, imposing an upper boundary of 900 individuals to the population size. The genome of the initial individuals contains a single copy of each gene, in randomised order and with an average of 2 TFBS or random type upstream. Individuals compete in a 7x7 neighbourhood for the opportunity to reproduce into an empty spot (this done for computational efficiency, and the realism of local competition). An individual's chance to reproduce is proportional to its fitness divided by the sum over the fitness values of itself and the other individuals neighbouring the empty position: $P_i = \frac{f_i}{\sum_{j=1}^{nb} f_j}$. Death occurs with a constant probability d , and individuals move on the lattice via Margolus diffusion.

Upon reproduction, the genome is mutated via duplications and deletions of both genes and TFBS, with a per-element probability (Fig.1E). In addition, TFBS may also mutate their type (which gene binds) and weight (activating or repressing), and new TFBS may appear *de novo* as an innovation. Gene duplication also copies the associated TFBS, and results in multiple genes of the same type. The expression of all genes of the same type therefore contributes to the concentration of a single protein. Note that since we do not include mutations that change gene type, gene duplication cannot be followed by subsequent divergence. In order to simplify our model and decrease the number of different mutation rates in our simulations we do not evolve maximum gene expression rates, protein decay rates or TF activation and deactivation thresholds (parameters E , D and H in equation 1) similar to the approach taken in (ten Tusscher and Hogeweg, 2011). We run 8 series of 60 simulations, varying between a high and low morphogen decay rate, the presence and absence of CCS and/or gene expression noise, and the duration of the relaxation period.

2.2.2 Analysis pipeline

It is impossible to derive the patterning strategy deployed by an evolved network from mere network architecture. Even for small networks evolved to the simple task of patterning a single stripe along the body axis, identical network architectures may lead to different patterning dynamics for different regulatory parameter settings (Schaerli *et al.*, 2014). For these small networks it may still be feasible to determine the patterning strategy by examining the expression dynamics of individual genes; this strategy however will not provide a solution for larger networks evolved towards more complex patterning tasks, such as the one considered here. As a consequence, mostly individual case studies were previously used to unravel the evolved developmental mechanism, analysing only a few network architectures and their gene expression dynamics in detail (Beaupeux and François, 2016, François *et al.*, 2007, Salazar-Ciudad *et al.*, 2001a,b, ten Tusscher and Hogeweg, 2011, Vroomans *et al.*, 2016). However, if we aim to study the evolutionary forces driving evolution of complex oscillator networks and / or of sloped oscillator frequency gradients, large numbers of simulation outcomes need to be assessed to see which evolutionary forces may contribute to the evolution of these properties. Detailed manual analysis of each individual simulation outcome would be prohibitively slow. Furthermore, a different type of approach is needed to determine the nature of the evolved segmentation oscillator, i.e. whether it generates damped or persistent oscillations, and whether oscillation amplitude or period changes gradually or abruptly as a function of morphogen concentration. Therefore, we developed an automated analysis pipeline that can determine measures of network complexity and oscillatory frequency profiles for large numbers of simulations.

Complexity Analysis

Our pipeline starts by extracting from each simulation the genome of a single fit individual present in the population at the end of evolution (Fig. 2.1A). Because an evolved genome consists partly of redundant interactions, we first prune the genomes via a repeated process of trying to remove genes and binding sites in the genome, while keeping the final spatial expression pattern of the segmentation gene the same (ten Tusscher and Hogeweg, 2011). We will refer to these pruned genomes and networks as core genomes and networks, as they embody the essential core necessary to generate the segmentation pattern. To obtain measures for the complexity of the evolved networks we determine genome size (number of genes and TFBS), the number of regulatory loops present in the network encoded by the genome, the size (nr of genes) of these loops, and the number of positive and negative feedback loops (Fig. 2.1B). All measures are obtained for the core genomes and networks.

Fourier frequency profile analysis

Since the model incorporates posterior growth, we expect a significant part of the evolutionary runs to evolve a vertebrate-like solution, transforming temporal gene expression oscillations into a spatial segment pattern (Vroomans *et al.*, 2016). To determine the precise nature of the oscillations, we apply a Fast Fourier Transform (FFT, C library `fftw3.h`) to the gene expression dynamics and quantify how the amplitude and frequency of oscillations changes as a function of morphogen concentration. Since each cell leaving the posterior growth zone experiences the same morphogen decay, such an analysis will reveal both the temporal oscillation dynamics of an individual cell as well as the spatial oscillation profile across the tissue at a single time point. This method will therefore allow us to determine whether, in case of persistent oscillations, a sloped frequency profile is present and kinematic oscillation waves are to be expected.

In principle, one could apply Fourier analysis directly to the gene expression dynamics of a cell as it leaves the growth zone and experiences morphogen decay. However, cells leaving the growth zone undergo only few oscillations in a short amount of time, and there is only a limited number of timepoints per individual morphogen concentration level. This makes it hard to extract the precise oscillatory dynamics as a function of morphogen concentration, especially when the morphogen decays rapidly. Furthermore, such an analysis would not be able to distinguish whether, at any given morphogen concentration, oscillations are stable or damped. Therefore we decided to obtain longer time series of gene expression by running the evolved networks multiple times, each time with a different but constant morphogen concentration, using a linear set of concentration levels occurring in the morphogen gradient (Fig. 2.2A). This ensures that the same amount of data and detail is available for oscillators evolved under fast and slow morphogen decay.

After developing this series of gene expression dynamics for different morphogen concentrations, we apply a Fourier analysis for each individual gene for each of these different time series (Fig. 2.2A). Subsequently we select the gene oscillating with the largest amplitude. For this gene, we then plot the frequency distributions (amplitude per frequency) for each morphogen concentration next to each other in a 2D heatmap, creating the so-called “frequency profile” (Fig. 2.2A). We give examples of the resulting plots in figure 2.2C, first column. Note how the frequency of the oscillations may or may not change with the morphogen concentration. A side effect of using this Fourier analysis is that, in addition to detecting the frequency of the genetic oscillator as the dominant mode, it also detects one or more so called eigenmodes of this frequency, as can be clearly seen in Fig. 2.2C, second row. These eigenmodes have no particular biological meaning.

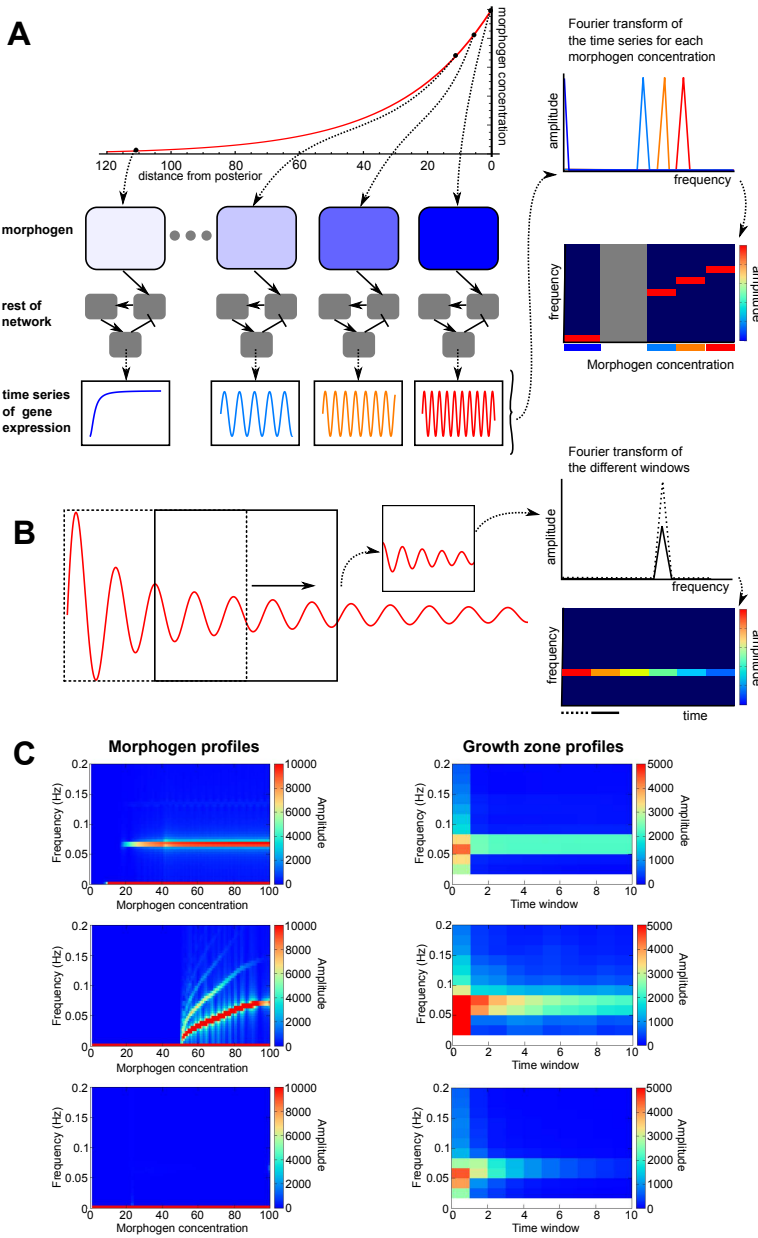


Figure 2.2. Explanation of the Fourier analysis procedure. A) We run the evolved network for 1800 steps with several, fixed concentrations of the morphogen. For every gene, we take the Fourier transform of the temporal gene expression dynamics to find the gene’s oscillation frequency for that particular morphogen concentration. We plot the Fourier transform data of all concentrations together in one heatmap, where the colour intensity represents the amplitude at every frequency for every concentration. See also C for a “real-life” example. (...)

Figure 2.2. (...) B) For the network run at the highest morphogen concentration (representing the growth zone), we also perform a sliding-window analysis: here, we take subsets of the time series generated as in A, and apply the Fourier transform to every window to visualise the change in frequency and amplitude over time in the growth zone. The rest of the procedure is the same as in A. C) Examples of frequency profiles from real simulations. The plots in the left column are generated as explained in A, and those on the right as in B.

To investigate whether the frequency or the amplitude of oscillations changes in the growth zone, we also apply this Fourier analysis to different subsections of the time series for the high morphogen concentration occurring in the growth zone (Fig. 2.2B). The procedure for making the frequency profile heatmap remains the same, but now the x-axis represents developmental time rather than morphogen concentration. Examples can be found in figure 2.2C, second column.

Oscillator classification

To compare the evolutionary outcomes arising for the different decay rates and hence morphogen gradient shapes, we would like to classify the obtained frequency profiles into the three different categories illustrated in Fig. 2.2C. First, we distinguish between damped and persistent oscillators depending on the fourier profile obtained from the growth zone. This is done by simple visual inspection of the profile, determining whether or not oscillations of non-zero amplitude persist throughout the time window. Next, within the category of persistent oscillators we determine whether a frequency profile is constant across the morphogen gradient or rather has a sloped appearance. This classification was formalised as follows: we measure the maximum oscillatory frequency occurring for the high morphogen concentrations in the posterior as well as the minimum frequency of the oscillations just prior to the ceasing of oscillations. Next we determine the difference between these oscillation frequencies, indicating the extent of oscillator slowing across the morphogen gradient. We choose a particular threshold value for this frequency difference (0.02). For frequency differences larger than this threshold we classify the oscillator as one with a sloped frequency profile, for smaller frequency differences we denote it as an oscillator with an approximately constant frequency profile.

2.3 Results

2.3.1 Evolutionary outcomes

To test how the length scale and slope of the morphogen gradient influences the evolution of segmentation, we ran two sets of 60 simulations: one with a small and one with a large morphogen decay rate, leading to shallow and steep gradients respectively. For both sets,

nearly all simulations manage to evolve a tissue pattern with 10 or more segments (59 of 60 simulations with a shallow gradient, 60 out of 60 simulations with a steep gradient, Fig. 2.3A). Of these successful simulations, 10 shallow-gradient and 11 steep-gradient simulations manage to evolve the maximum number of 18 segments (the median nr of segments is 16 for both cases). Typical space-time plots for both kinds of gradient are shown in figure 2.3B.

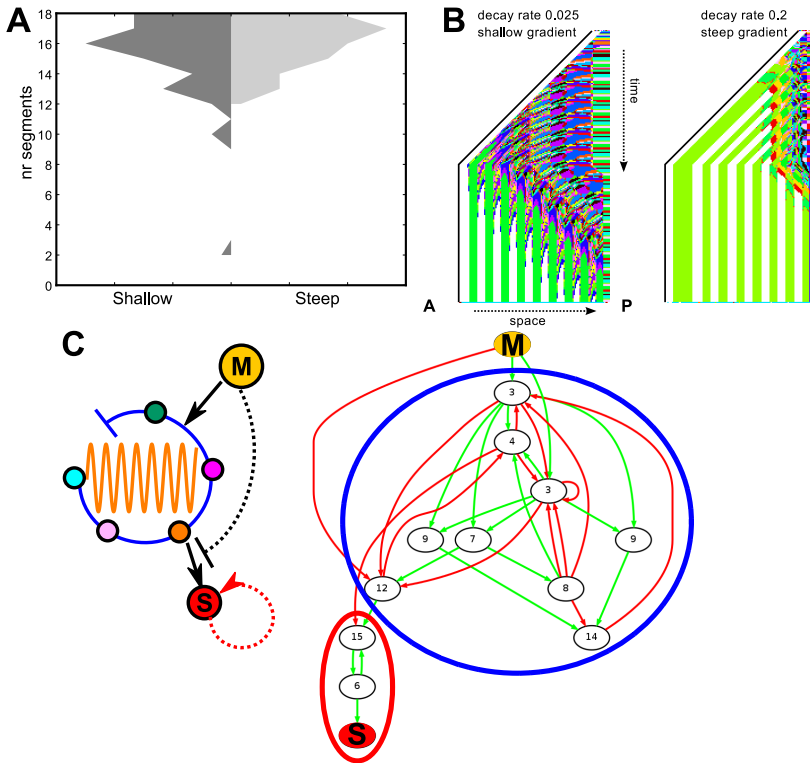


Figure 2.3. Summary of simulation results. A) Histograms of the number of segments formed in simulations with shallow and steep gradients. B) Examples of the resulting space-time plots from an individual at the end of a simulation. The posterior growth zone on the right is anchored, the other cells shift position when the tissue grows. The colour reflects the cell type, defined by the expression levels of all genes within a cell (see box at end of chapter). Note the regular alternation of gene expression in the posterior growth zone. C) Left: a simplified representation of the gene regulatory networks that evolve in our simulations; right: an example of an evolved network (pruned, see Methods). The clock that generates gene expression oscillations is indicated in blue, the bistable switch in red. D) A simplified representation of the phase space of the evolved networks. At high morphogen concentrations, gene levels oscillate (blue). These oscillations make the network state alternate between the basins of attraction of the two stable states that emerge for a low morphogen (red).

The evolved networks generally contain a genetic oscillator that consists of multiple negative feedback loops combined with a bistability motif. One or more of these feedback loops are regulated by the morphogen, so that oscillations occur for high morphogen concentrations but cease below a certain threshold level of morphogen (illustrated in figure 2.3C). Below this threshold, the expression of the segmentation gene becomes stabilised by the bistability generating positive feedback loop.

Due to the non-linearity of gene expression regulation in our model, the positive feedback on the segmentation gene generates a bistability so that either high or low expression can be maintained. While the morphogen concentration is high, both the high and the low state are unstable and the network can keep oscillating between the two regions that form the future basins of attraction of these states (Fig.2.3D). When morphogen concentrations drop, oscillations terminate, the two states become stable and the network converges to one of them. Whether it converges to high or low segmentation gene expression depends on the phase of the cycle at which oscillations stopped. Thus, the bistability allows for a translation of oscillations into a stable segmented gene expression pattern. This structure is similar to those in (ten Tusscher and Hogeweg, 2011), although the pruned networks tend to remain somewhat larger. Variations on this general theme do occur, for example the segmentation gene and the genes in the positive feedback loop may be part of a negative feedback loop of the oscillator. In other variants the inhibition by the morphogen may be indirect. Still, the overall mechanism generating morphogen dependent oscillations and translating them into a stable segmentation pattern is the same.

2.3.2 Fourier analysis as a signature of gene expression dynamics reveals spontaneous evolution of travelling segmentation waves

We need to validate whether the computed frequency profiles of the Fourier analysis correctly reflect the gene expression dynamics that actually occur during development. To perform this validation, we show the tissue dynamics in a number of different ways for three cases with qualitatively different frequency profiles in figure 2.4A. First, we show the frequency profile obtained using our Fourier analysis method. (Fig. 2.4A). Next, we show developmental space-time plots of cell types (Fig. 2.4B) and of segmentation gene expression (Fig. 2.4C). We also show snapshots of the spatial expression pattern of the segmentation gene at different time points (Fig. 2.4D Supp. Video 1-3), and line graphs showing the spatial expression pattern of the segmentation gene for different time points (Fig. 2.4E).

To discuss gene expression and oscillatory dynamics, and whether or not actual dynamics and the dynamics as reflected by our Fourier analysis correspond, we introduce the concept of a transient. This is the region or period of time in which the cells are

no longer in the growth zone but have not formed a stable segment yet. Clearly, individuals with a slow morphogen decay and shallow gradient have a longer transient, with segment formation occurring later and further away from the posterior end of the body (Fig. 2.4A,A' vs. A''). Because a longer transient makes developmental dynamics easier to interpret by eye, we will focus on results from evolutionary simulations with a shallow gradient for the validation of our Fourier frequency profile.

In the first column, the computed frequency profile clearly shows a slope, implying the occurrence of slower oscillations for lower morphogen concentrations (from here on called a sloped frequency profile).

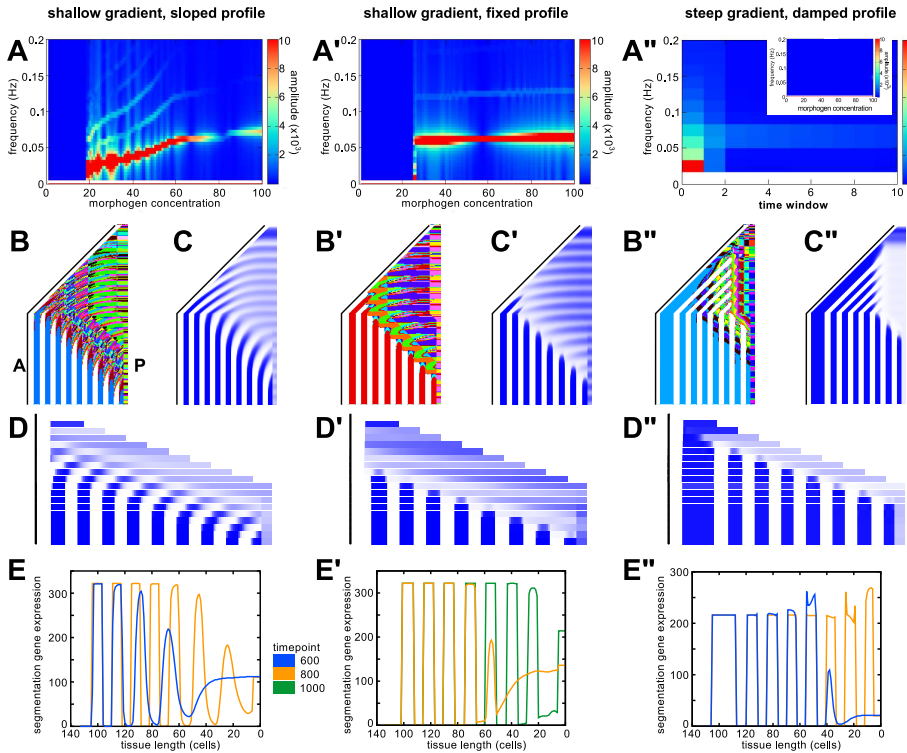


Figure 2.4. Frequency profiles correctly predict tissue-scale dynamics. A-A'') Space-time plot of celltypes: each colour represents a unique combination of gene expression values within each cell. Note how the tissue grows in the first half of development. B-B'') Space-time plot of the segmentation gene expression. C-C'') Line graphs depicting the expression level of the segmentation gene at two different timepoints (indicated by the colour). Cell 0 is the dividing cell in the growth zone. D,D'') Frequency profile of gene 5. D'') Frequency profile of gene 10 (the strongest oscillating gene) E, E'') Snapshots of the tissue dynamics for gene 5, or E'') gene 10. The anterior ends (indicated by the black bars) are aligned for greater clarity. The pictures are taken 12 steps apart.

In the space-time plots we can see this slowing down from the bent shape of the waves of segmentation gene expression. In the snapshots of segmentation gene expression in figure 2.4D we see that every segment starts as a travelling wave from the posterior, and becomes narrower and more strongly expressed as it arrives at the anterior. Thus, a sloped frequency profile translates to travelling waves across the tissue, much like those observed in vertebrate development.

In contrast, the individual used as an example in the middle column of figure 2.4 (A'-D') has a constant frequency profile, implying that oscillations have a constant frequency for a range of morphogen concentrations and then suddenly cease to exist for lower morphogen concentrations (a constant frequency profile). Both developmental space-time plots and snapshots of segmentation gene expression dynamics show that indeed, most of the tissue oscillates synchronously and that only the very anterior end shows a minor deviation of this dynamic immediately prior to segment stabilisation. Based on our frequency plot we can deduce that in this small region, the cells are already in a non-oscillatory regime, converging towards one of the two stable states that allow for a segmented pattern. Note that this is different from the individual in the left column with travelling waves, where the anterior tissue that is out of sync with the posterior end is in a regime of sustained but slower oscillations.

Finally, in the right column of figure 2.4 (A''-D'') we display an individual whose frequency profile only shows oscillatory dynamics for the high morphogen concentrations that occur in the posterior growth zone. The sliding-window analysis of these growth zone oscillations reveals that these oscillations are damped, reducing their amplitude over time (a "damped frequency profile"). This is confirmed by both the developmental space-time plot and the snapshots of segmentation gene expression dynamics, which show a clear decrease in oscillation amplitude in the growth zone.

In all 3 cases illustrated above, there is a clear correspondence between the developmental dynamics as suggested by the computed Fourier frequency profile and the actual observed developmental dynamics. We therefore conclude that the Fourier frequency analysis is a valid method to compute and categorise in an automated manner the type of oscillatory dynamics produced by evolved networks. Note that while the above examples are easily distinguishable, clear-cut cases, unfortunately not all computed frequency profiles are that easy to interpret or fall into these three clear categories. Some profiles have a very modest slope, in other cases oscillations extend beyond the growth zone but for only a limited part of the entire morphogen concentration range, in yet other cases oscillations may be damped for the high morphogen concentrations in the growth zone yet persistent for a range of lower concentrations (Supp. Fig. S2.1). The difference between the tissue dynamics of the different profiles then becomes harder to see, as they span a smaller

region. Still, we maintain that also for these more complicated cases the frequency profile reliably reflects the actual oscillatory developmental dynamics.

2.3.3 A shallow gradient more often yields sustained oscillations and sloped frequency profiles

Comparing figure 2.4A and A' to A'' illustrates that the transient (where cells have left the growth zone but have not yet stably formed a segment) is longer for shallow morphogen gradients than for steep gradients. After all, the transition from oscillations to segments occurs at low morphogen concentrations, which are reached later and further away for shallow gradients. Thus, an interesting question is whether this spatially and temporally extended transient has evolutionary consequences in terms of network complexity and types of oscillatory dynamics that evolve. Indeed, when we classify all simulations into the three broad categories as displayed in figure 2.4 (see also Methods), the simulation set with a shallower posterior gradient has a lower fraction of profiles with only damped oscillations, and a higher fraction of sloped frequency profiles (Table 2.1), while the two sets contain a similar number of simulations with a constant frequency profile. To test the robustness of these results, we measured the frequency difference within a profile (Fig. 2.5), rather than categorising the profiles using somewhat arbitrary cutoffs to distinguish sloped from fixed profiles. The distribution of these frequency differences makes clear that not only do runs with shallower gradients more often have a sloped profile, but they also tend to have a larger frequency difference across their profile (Fig. 2.5).

The difference in the number of simulations with damped oscillation profiles is intuitively understandable. In the case of persistent oscillations, cells oscillate with a constant amplitude, which causes them to alternate between two regions, which form the future basins of attraction of two fixed point states. These states only appear in the system when the morphogen concentration drops below a certain threshold and the stable limit cycle supporting persistent oscillations disappears (Supp. Fig. S2.2A). The stable oscillations thus serve as a memory for the oscillation phase at which a cell left the growth zone and will determine the cells final segmental state. In case of damped oscillations, cells oscillate at a decreasing amplitude and are on a trajectory towards the stable equilibrium residing inside an unstable limit cycle (Supp. Fig. S2.2B). If this is combined with a shallow morphogen gradient and hence a long transient, oscillation amplitude may have decreased so far that cells have already become stuck in the basin of attraction of only one of the two stable states, thus preventing segmentation. Put differently, due to the reduced amplitude the cells lose the information of the oscillation phase with which they left the growth zone – and thus what type of segment they should become. With a steep gradient, the bistable regime is reached fast enough and therefore the nature of the oscillation (damped or persistent) is less relevant.

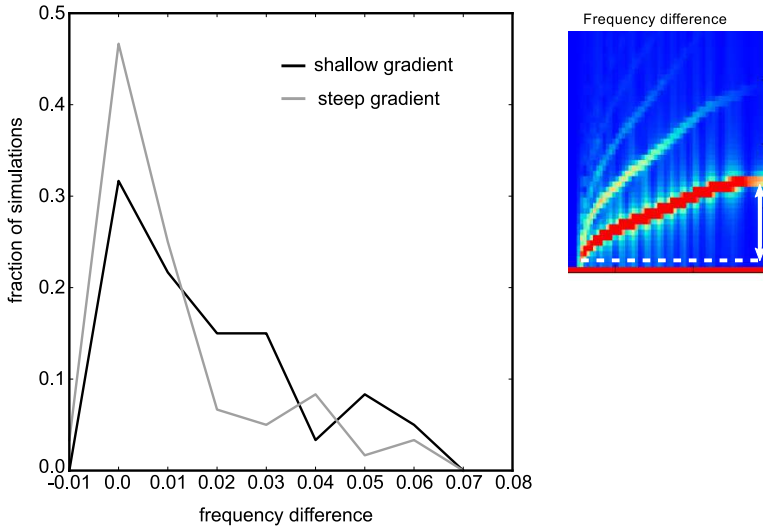


Figure 2.5. A shallow gradient more often yields a sloped frequency profile Histogram of the frequency difference between oscillations in the growth zone and at the end of the profile (see illustration). Note that the damped oscillators are grouped in the bin with 0.0 frequency difference. Bin size: 0.01

It is less clear why individuals with a shallow gradient more often evolve a sloped frequency profile and waves of gene expression. If we assume that there is no inherent difference in functionality between having a constant or a sloped frequency profile, the increase in the number of sloped profiles could simply be due to the more general need for sustained oscillations when the gradient is shallow. In that case, a sloped profile represents just one of two ways of achieving persistent oscillations. On the other hand, if a sloped frequency profile were to have any additional functionality such as its suggested larger robustness, it would have more space and time to exert this functionality under a shallow, more spread out, morphogen gradient. This could possibly contribute to its more frequent evolution under shallow gradients.

2.3.4 A shallow gradient more often yields complex solutions

Above we established that the steepness of the morphogen gradient influences the type of oscillatory frequency profile. Next, we investigated whether differences in morphogen gradient slope also cause differences in the complexity of the evolved oscillatory networks. We find that individuals with shallow morphogen gradients may evolve to slightly larger core networks, but this difference is not significant due to the large amount of variation between simulations with the same type of morphogen gradient slope (Fig. 2.6A). Still, the networks evolved under a shallow morphogen gradient do tend to contain more

feedback loops, especially more of the negative FBLs that are needed to construct an oscillator (Fig. 2.6B). In addition, loops are also larger on average (Fig. 2.6C). However, this increase in average loop number and size is mainly caused by a subset of 14 simulations (out of a total of 59) which have more than 20 negative feedback loops, rather than all simulations evolving to larger networks. These simulations also have the largest genomes (Supp. Fig. S2.3). The increased complexity in this subset may reflect the stronger requirement for stable oscillations under a shallow gradient.

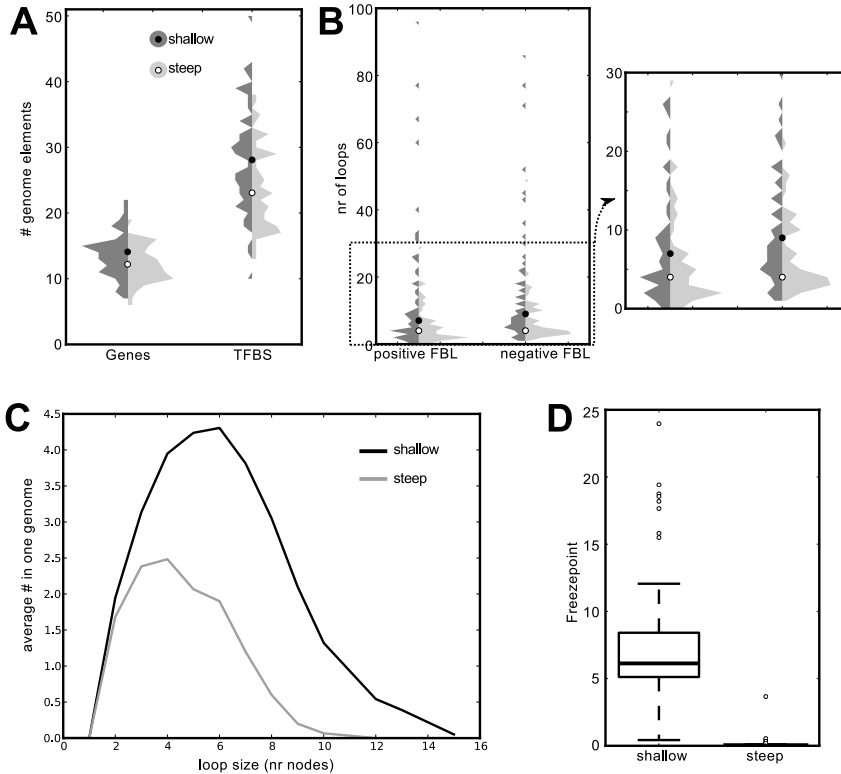


Figure 2.6. Comparison of genome and network properties A) Violin plots (vertical histogram) of the number of genes and transcription factor binding sites (TFBS) in the pruned genomes of shallow-gradient (dark) and steep-gradient (light) simulations. B) Violin plots of the number of positive and negative feedback loops in the pruned networks of the shallow- and steep-gradient simulations. C) Histogram of the number of loops (FFL and FBL) of a certain size. All histograms of individual simulations have been summed for this average histogram. D) Boxplot of the morphogen level at which individuals reach a stable expression (after the transition from the oscillatory to the non-oscillatory regime).

Finally, individuals evolved with a shallow gradient clearly freeze their temporal oscillations into spatial stripes at a higher concentration of the morphogen (Fig. 2.6D). This may reflect a partial compensation of the long transient by reducing the time until segment formation. On the other hand, it may simply be a result of the lack of a sharp transition from high to low morphogen concentration in a shallow gradient. In that case, it is somewhat arbitrary at which posterior morphogen concentration the transition to stable segments takes place, while for a steep gradient, a morphogen concentration close to zero is the logical point for this transition.

2.3.5 Network complexity and frequency profile slope are uncorrelated

As discussed above, regulatory networks are on average more complex and oscillator profiles are on average more sloped under a shallow morphogen gradient. However, this increase in average complexity and slopedness resulted from substantial increases in network complexity or profile slopedness in a subset of the simulations rather than a modest increase in most simulations. This raises the question to what extent the evolution of complex networks and sloped frequency profiles is correlated. Note that vertebrate segmentation is characterised by both a complex three-part oscillatory network and a sloped oscillatory frequency profile, which may be taken to suggest that these two properties evolve in concert.

Supplementary figure S2.4 shows a scatter plot of network complexity (as measured by the number of negative feedback loops in the pruned network) versus frequency difference across the profile. Interestingly it illustrates the absence of a clear correlation between network complexity and slopedness. This suggests that different evolutionary trajectories may be involved in generating complex networks and sloped frequency profiles and demonstrates that complex networks are not necessary for sloped frequency profiles to occur. This fits with our hypothesis that the shallow gradient simulations require a complex segmentation networks to generate stable rather than damped oscillations, independent of whether or not the frequency of these stable oscillations depends strongly on the morphogen concentration. These results suggest that network complexity and frequency gradient may have evolved independently in vertebrate segmentation.

2.3.6 Gradient steepness influences evolutionary innovation speed

We thus established that both the type of oscillations, and the underlying network generating the oscillations may differ between individuals with a shallow or steep morphogen gradient. Subsequently we investigated whether this difference in final evolutionary outcome is reflected by differences in the evolutionary trajectories leading up to these outcomes. In figure 2.7 we illustrate that the evolutionary trajectories of individuals evolved under the two different gradient slopes differ significantly. Simulations with a steep gradient yield individuals with more than 10 segments very early in evolution (Fig. 2.7A). In contrast, while many of the simulations with a shallow gradient are also able to find a segmented solution quickly, others need a much longer evolutionary timespan to yield individuals with 10 or more segments. Much of this time, these simulations are stuck in a primitive, two-segment stage, where the entire tissue that is generated by the growth zone expresses the segmentation gene while the head does not (Fig. 2.7B). This indicates that it is harder for evolution to find a segmentation pattern for a shallow gradient.

To further investigate this difference, we removed the “head” from the initial tissue (see fig. 2.1C). As discussed in the methods section, the head region is the part of the tissue in which the morphogen gradient is absent and no gene expression whatsoever occurs. As a segment boundary is defined as the transition from low to high expression of the segmentation gene or vice versa, simply expressing the segmentation gene in the non-head part of the tissue suffices to generate the first segment. Thus, removing the head region will make it harder for evolution to discover the first segment, and may therefore in some cases make it impossible to find a solution. The rate of success of evolutionary simulations indeed decreases significantly in absence of a head region, and considerably more so for shallow than steep morphogen gradients. Only 28 out of 60 simulations find a solution for a shallow gradient, while 51 out of 60 simulations evolve a segmented pattern with at least 10 segments for a steep gradient. This further supports our observation that a segmented body pattern evolves more easily for steep morphogen gradients.

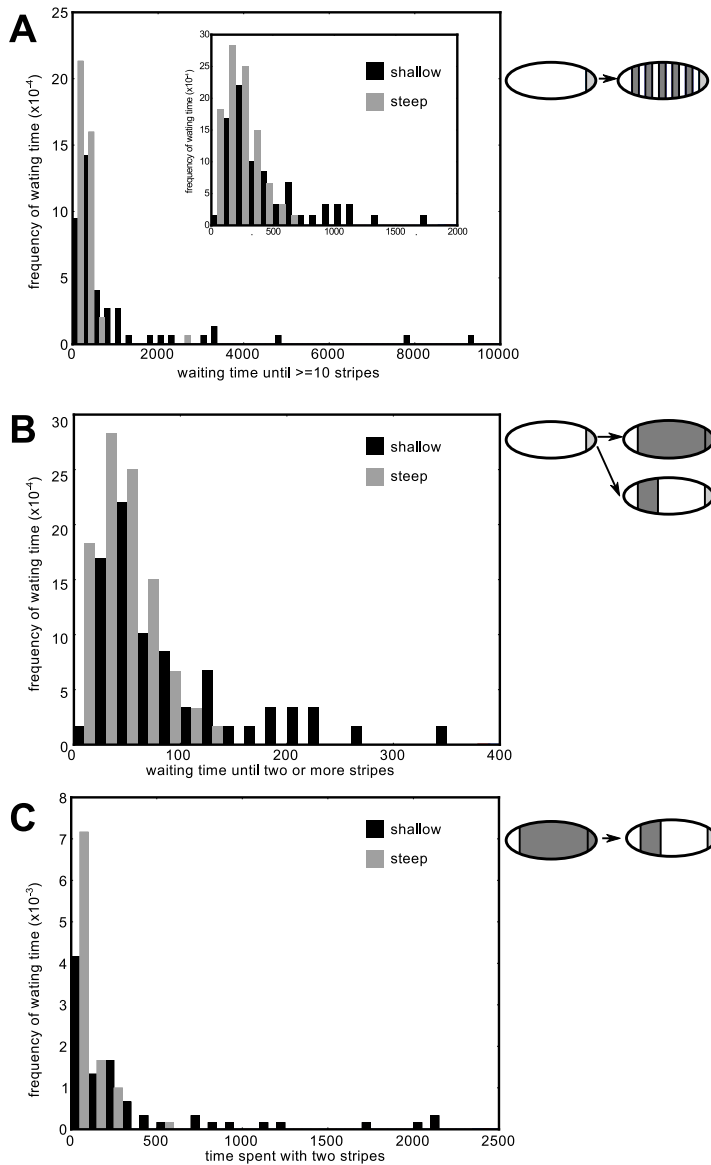


Figure 2.7. A shallow gradient takes longer to find a solution A) Histogram of the number of generations it took for simulations to make 10 or more stripes. Bin size=100 B) The waiting time until individuals with two or more stripes appear in the simulation. Bin size=25 C) The number of generations each simulation spent with only two stripes (see space-time plot). Note that the first bin includes those individuals which immediately find more than two stripes. Bin size=50.

2.3.7 Evolved segmented bodyplans adapt easily to a different morphogen gradient

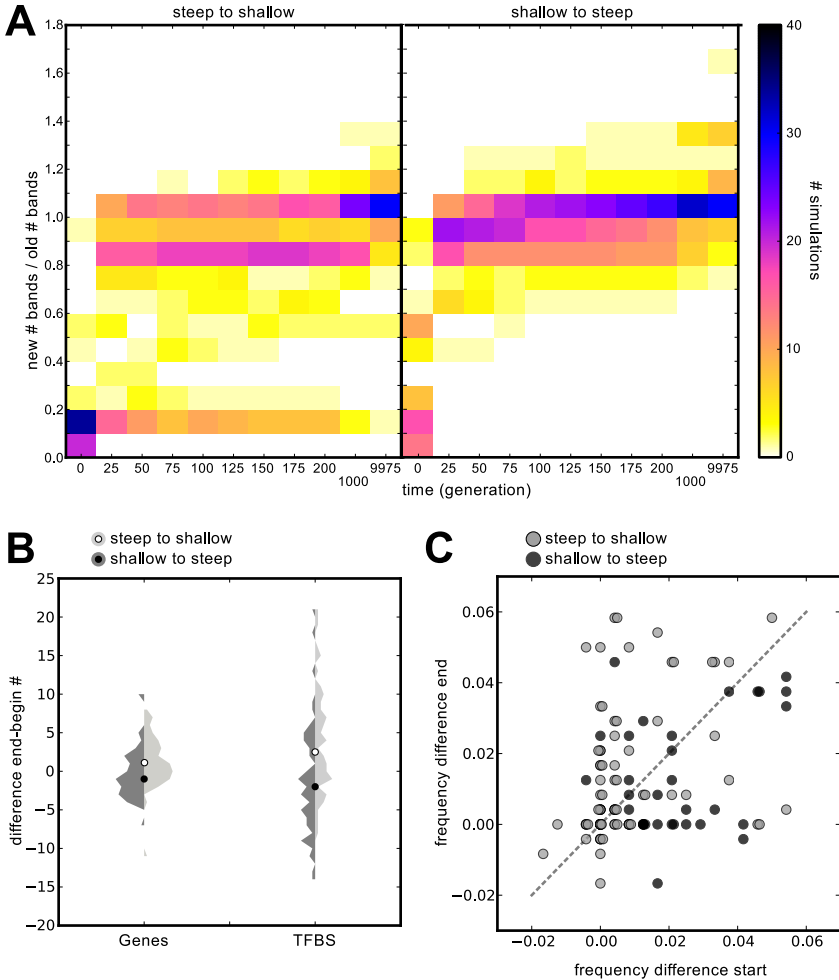


Figure 2.8. Switching to another decay reveals functionality of differences between gradients A) Heatmap of the number of segments (ratio original number of the transplanted individual / current nr of segments maximum fit individual) after switching the decay rate of all individuals. B) Violin plot of the difference in the number of genes and TFBS in the pruned genome between the start and end of the simulation. C) Scatterplot with the frequency difference (see fig. 2.5C) of the Fourier profile at the start and the end of the decay-switch run. Blue is from steep to shallow, red from shallow to steep.

The differences in the number of regulatory loops, the evolutionary history and oscillatory frequency profiles suggest that the steepness of the morphogen gradient influences the functional requirements for the evolved networks. To further assess the functionality of these differences while at the same time investigating whether these differences are evolutionary mutually exclusive, we extract individuals evolved in a steep or shallow gradient, and let them continue evolution while now applying a gradient of the opposite steepness.

When the evolutionary transition goes from a shallow to a steep morphogen gradient, 22 out of 59 simulations are immediately able to generate more than three segments (Fig. 2.8A). In contrast, for the evolutionary transition from a steep to a shallow gradient, only 6 out of 60 simulations can still generate more than 3 segments directly after the transition (Fig. 2.8A). Still, in both cases evolution generally needs fewer than 100 evolutionary time steps (approximately 30 generations) to come to a new solution with a similar number of segments as before the transition. However, for the steep to shallow transition, 3 simulations needed more than 1000 time steps to restore their prior segmentation pattern.

We conclude that the segmentation strategies evolved under either shallow or steep morphogen concentrations are not mutually exclusive but can in fact easily be converted into one another. However, these results also imply that the evolutionary transition from shallow to steep is easier than that from steep to shallow, thus confirming our earlier findings on the difference in speed with which segmentation patterns evolve under shallow and steep gradients and that shallow gradients require more complex oscillatory networks.

Next, we looked at the difference in (pruned) genome size between the original individuals and an individuals at the end of the evolutionary transition simulation. We observe that for a transition from a shallow to a steep gradient, genome size is more likely to decrease, while for a transition from a steep to a shallow gradient genome size is more likely to increase (Fig. 2.8B). Although the observed differences are admittedly small, they are in line with the differences in genome size we saw in figure 2.6. This suggests that this difference is functionally significant.

Finally, in figure 2.8C we illustrate that the frequency profile also changes in accordance with our earlier results. For the evolutionary transition from a shallow to a steep gradient, the slope of the frequency profile is more likely to decrease (27 decrease, 20 increase) and the number of damped oscillators increases (from 4 to 18). For the opposite evolutionary transition the slope of the frequency profile is more likely to increase (32 increase, 19 decrease), and the number of damped oscillators decreases (16 to 7). This result suggests that also the differences in frequency profile are functionally significant.

2.3.8 Other influences on evolutionary outcome

cell-cell signalling Although our model is more realistic in comparison to previous models because it explicitly incorporates growth and the formation of a morphogen gradient through decay, it still contains many simplifications. To name but a few, our model ignores the role of cell-cell signalling, gene expression noise and the two or three dimensional aspects of the segmented tissue. To test the robustness of our results, we repeated part of our experiments in the presence of direct, contact-mediated cell-cell signalling (see Methods).

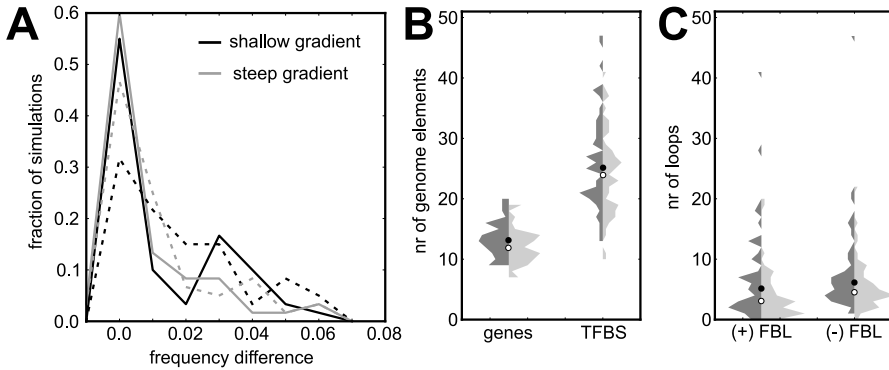


Figure 2.9. Simulations with cell-cell signalling A) Histograms of the frequency difference between oscillations in the growth zone and at the end of the profile. The dotted lines are the histograms of the simulations without CCS (Fig. 2.5) B) Violin plots (vertical histogram) of the number of genes and transcription factor binding sites (TFBS) in the pruned genomes of shallow-gradient (red) and steep-gradient (blue) simulations. C. Violin plots of the number of positive and negative feedback loops in the pruned networks of the shallow- and steep-gradient simulations.

Table 2.1. Prevalence of frequency profiles

| simulations | decay rate | sloped profiles | fixed freq. profiles | damped profiles | not classified |
|-------------------------------|------------|-----------------|----------------------|-----------------|----------------|
| Without CCS | 0.025 | 0.25 | 0.61 | 0.05 | 0.09 |
| | 0.2 | 0.13 | 0.58 | 0.23 | 0.06 |
| With CCS | 0.025 | 0.31 | 0.45 | 0.20 | 0.04 |
| | 0.2 | 0.08 | 0.58 | 0.27 | 0.07 |
| shorter stabilisation period | 0.025 | 0.25 | 0.58 | 0.08 | 0.09 |
| | 0.2 | 0.1 | 0.44 | 0.32 | 0.13 |
| gene expression noise and CCS | 0.025 | 0.37 | 0.37 | 0.02 | 0.23 |
| | 0.2 | 0.13 | 0.74 | 0.08 | 0.05 |

Interestingly, damped oscillators now occur at an approximately similar rate for steep and shallow morphogen gradients (Table 2.1), suggesting that cell-cell signalling can partly mitigate the effect of damped oscillators converging too early to a single equilibrium. However, sloped frequency profiles still evolve more often under a shallow than under a steep morphogen profile (Fig. 2.9A). This provides further support for the notion that a sloped profile is functional in a shallow gradient but is not directly related to the occurrence of sustained oscillations, as the percentage of damped oscillators is now the same for the two types of morphogen gradients.

The difference in genome size between individuals evolved under steep and shallow morphogen gradients is even smaller in the presence of cell-cell signalling, and also the difference in number of negative feedback loops is smaller (Fig. 2.9B-C). This further supports the hypothesis that a complex regulatory network is not required for a sloped oscillatory frequency profile, but rather that it is necessary to deal with the need for sustained oscillations, something that appears a less severe constraint in the presence of cell-cell signalling.

A shorter stabilisation period Simulations with a shallow gradient reliably yield mechanisms where the transition from oscillations to segments occurs at a higher morphogen concentration than in mechanisms evolved with a steep gradient (Fig. 2.6D). We hypothesised that this transition at higher morphogen concentrations is necessary to obtain stable segment patterning within the available developmental time window. To test this we reduced the amount of time the developmental process has after tissue growth (during which further morphogen decay occurs) to achieve final segment patterning. Halving the duration of this period reduces the number of successful simulations with a shallow gradient but not a steep gradient (Fig. 2.10A). Furthermore, the shallow gradient simulations that do manage to evolve segments, form these at a considerably higher morphogen concentration than before (Fig. 2.10B).

However, the percentage of simulations that evolve a sloped frequency profile does not increase, although their average slope does seem to be a bit increased (Fig. 2.10C, table 2.1). This indicates that a sloped profile is not required to terminate oscillations at a higher morphogen concentration. It will be interesting to investigate whether the network complexity is increased in the successful simulations with a shallow gradient.

Gene expression noise A sloped profile has been suggested to increase the resilience of the segmentation process against gene expression noise (El-Sherif *et al.*, 2014), which we investigated by adding Gaussian noise on gene expression in our simulations. Adding this noise slightly reduces the number of successful simulations with a shallow gradient (51 successful simulations out of 60 Fig. 2.10D), but not of those with a steep gradient. Strikingly, the fraction of simulations with a sloped profile increases in simulations with a shallow gradient, which suggests that a sloped profile indeed plays a role in increasing

robustness against noise (Fig. 2.10E). We also find that adding noise greatly increases the fraction of simulations with a steep gradient that yield persistent oscillations (table 2.1). This can be understood from a dynamical systems perspective; noise may cause a damped oscillator to reach its stable equilibrium faster, thus causing it to stop oscillating and lose phase information.

We want to pursue these findings further, by investigating whether we can identify differences between simulations evolved with and without noise. For example, these additional data may allow us to identify correlations between the type of frequency profile and genome sizes, which was not possible before. Furthermore, we would like to assess the difference in evolvability between individuals evolved with or without noise. These experiments may provide more information on the possible functions of a sloped frequency profile.

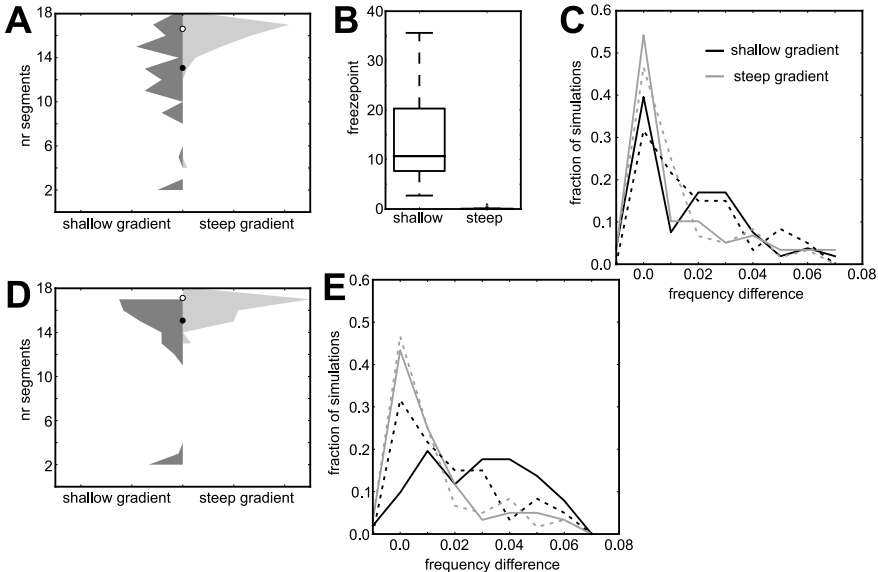


Figure 2.10. Preliminary results A-C) Simulations with a shorter stabilization period. A) Vertical histograms of the number of segments formed in simulations with shallow and steep gradients. B) Boxplot of the morphogen level at which individuals reach a stable expression (after the transition from the oscillatory to the non-oscillatory regime). Note how the level of the freezepoint for the shallow gradient is higher than before (Fig. 2.6D). C) Histograms of the frequency difference between oscillations in the growth zone and at the end of the profile, with the dotted lines representing the histograms of the original simulations (see Fig. 2.5). D-E) Simulations with gene expression noise. D) Vertical histograms of the number of segments formed in simulations with shallow and steep gradients. E) Histograms of the frequency difference between oscillations in the growth zone and at the end of the profile. Note the difference in y-axis compared to C.

2.4 Discussion

Body plan segmentation is a major evolutionary innovation exhibited by the vertebrate, arthropod and annelid clades (Davis and Patel 1999, Peel and Akam 2003). While all three clades predominantly show periodic, anterior posterior generation of segments, fundamental differences in patterning details exist. As an example, studies on *Tribolium* segmentation suggest that segmental prepatterning occurs at a considerably smaller distance from the posterior growth zone than is the case in vertebrates. This may have implications both for the shape of the posterior morphogen gradient as well as the potential to generate travelling waves of gene expression.

In the current paper we extended previous evo-devo models for the evolution of body axis segmentation by incorporating realistic growth from a posterior growth zone, with a posteriorly expressed morphogen that forms a sloped gradient through decay. We have previously shown how this biases for the evolution of oscillatory sequential segmentation (Vroomans *et al.*, 2016). In addition we developed an automatic analysis pipeline that allows us to compute parameters describing network complexity and oscillator dynamics. With this, we investigated the effect of different morphogen decay rates (resulting in differently sloped morphogen gradients), gene expression noise and the developmental time available for segment patterning. We showed that in this new model, different types of oscillators can evolve, with damped oscillators or oscillators with a constant period frequently evolving. In some simulations we also observed the spontaneous evolution of oscillators with a frequency profile resulting in a slowing down of oscillations towards the anterior. Such a profile yields travelling waves of gene expression (Dequéant and Pourquié, 2008, Morelli *et al.*, 2009), similar to those seen during somitogenesis in vertebrates. Furthermore, for these frequency profile cases we find that oscillation frequencies typically decrease by 50-60% percent rather than decreasing all the way to zero, in agreement with recent experimental measurements (Shih *et al.*, 2015).

We found that a steep morphogen gradient more often leads to the evolution of a damped oscillator. Under a shallow morphogen gradient, such damped oscillators are rare because these simulations have to deal with a long spatiotemporal transient before morphogen levels have decayed enough for segmentation to occur. Under such settings, sustained oscillations serve as a robust dynamic memory of the oscillator phase with which the cell left the growth zone. Preliminary results show that in presence of gene expression noise an increase in the number of evolved persistent oscillators arises also for steep morphogen gradients, supporting the notion that persistent oscillators contribute to robust patterning.

In addition to differences in the occurrence of damped oscillators, shallow gradients also more often yield a sloped profile, leading to travelling waves. Preliminary

simulations show that this is even more so in the presence of gene expression noise. We speculate that a sloped profile could aid in the transition to segments when the morphogen concentration changes but slowly, making the transition from the oscillatory to non-oscillatory regime harder to detect. The slowing down of oscillations keeps the state of the network longer in the basin of attraction of each of the stable states of the bistable switch. This causes the network to experience a larger difference in morphogen concentration in any part of the oscillation cycle, thus facilitating the detection of the regime change to the bistable switch. Finally, we found that the genomes evolved under a shallow gradient tend to be larger and the networks have more and larger feedback loops.

We observe interesting parallels between the simulations with a shallow morphogen gradient and vertebrate somitogenesis. In these simulations as well as in somitogenesis there is a long spatiotemporal transient before segments are laid down, that is accompanied by kinematic waves of gene expression (Lauschke *et al.*, 2013, Palmeirim *et al.*, 1997, Shih *et al.*, 2015, Soroldoni *et al.*, 2014). In addition, in our simulations with a shallow gradient segmentation networks tend to be more complex, which resembles vertebrate somitogenesis where the segmentation network consists of an entanglement of at least three separate oscillators (Goldbeter and Pourquié, 2008).

In contrast, our results for a steep morphogen gradient show parallels to *Tribolium* segmentation. In the beetle, segment formation occurs considerably closer to the posterior growth region than in vertebrates, implying a substantially shorter transient. In addition, the thus far elucidated oscillatory segmentation network is much simpler than the vertebrate segmentation network (Choe *et al.*, 2006). Although travelling waves have been observed, both the distance travelled and the wave contraction are substantially smaller than in vertebrates, suggesting only a modestly sloped frequency profile (El-Sherif *et al.*, 2012). These properties resemble our simulations with steep morphogen gradients which show shorter transients, damped or constant period oscillators and simpler networks. Based on this, it would be interesting to compare the caudal gradient that was recently suggested as the instructive morphogen in *Tribolium* segmentation (El-Sherif *et al.*, 2014), to the FGF gradient important for vertebrate segmentation. Based on the current work we predict a significantly shorter and steeper caudal than FGF gradient.

Intriguingly, in our simulations we did not observe a clear correlation between the evolution of network complexity and a frequency gradient. This suggests that these two properties could have evolved separately rather than simultaneously in vertebrates and may play a different role. Indeed, we demonstrated here that upon incorporating cell-cell signalling, shallow and steep morphogen gradients evolved damped oscillators equally often and differences in network complexity were even smaller, while differences in the number of evolved sloped frequency profiles remained. This demonstrates that cell-cell

signalling mitigates the need for sustained oscillators, and that rather than complex oscillators being necessary for generating a sloped frequency profile they are necessary for generating sustained oscillations. Finally, our preliminary results with a shorter developmental time window and gene expression noise also show independent changes in the number of persistent oscillations and sloped frequency profiles, demonstrating that sloped frequency profiles do not merely arise as a side effect of persistent oscillators.

Of course, there may have been additional requirements not accounted for in our simplified model, such as the need to scale segment size with the size of the remaining unsegmented tissue, that may cause the coordinated evolution of network complexity and sloped frequency profiles. Recent studies have indeed suggested that network complexity may reflect the need for two distinct oscillators, one with a constant frequency and one slowing down according to a sloped frequency profile, with the resulting phase difference patterning somite boundaries and polarity (Beaupeux and François, 2016, Lauschke *et al.*, 2013). Under these assumptions, network complexity and the presence of a frequency profile together function to pattern somites and hence would likely be correlated. It would be interesting to modify future evo-devo models to enable evolution of such alternative oscillatory patterning mechanisms, and investigate whether this mechanism as a whole increases robustness.

In the future, we aim to use the methodology developed here to further investigate the forces influencing the evolution of oscillatory segmentation. Specifically, we aim to further test the suggested increased robustness of sloped frequency profiles to expression and developmental noise and extend our analyses to a more realistic 2D tissue context. Interestingly, our preliminary results indicate that gene expression noise, but not noise in the morphogen gradient substantially increase the frequency with which sloped frequency profiles arise. Furthermore, we would like to investigate when steep versus shallow morphogen gradients are expected to evolve. Interestingly, our evolutionary transition experiments show that segmentation modes can easily convert from a steep to a shallow morphogen gradient and vice versa. We speculate that early in evolution there may be selection for steep morphogen gradients for which oscillatory segmentation readily evolves, while later in evolution there may be selection for shallow morphogen gradients that generate more robust and evolvable persistent oscillators. Preliminary results confirm this early selection for a steep gradient. Interestingly, amphioxus, belonging to the cephalochordate sister group of the vertebrates, indeed patterns its somites close to the posterior (Beaster-Jones *et al.*, 2008). This is consistent with a short, steep morphogen gradient. Furthermore, amphioxus lacks precision mechanisms involved in ensuring left-right symmetry (Bertrand *et al.*, 2015). This could point to a role for a long transient and a sloped frequency profile in maintaining left-right symmetry.

Table 2.2. parameter values

| parameter | values | remarks |
|----------------------------------|--------------|--|
| general | | |
| grid size | 30x30 | |
| evolutionary time steps | 10000 | |
| death rate | 0.5 | |
| initial # agents | 50 | |
| Development | | |
| developmental time steps | 900 or 1200 | the number of integration steps |
| duration of division period | 600 | divisions occur every 5 time steps |
| duration of stabilisation period | 300 or 600 | period without divisions |
| integration step size | 0.2 | forward Euler integration |
| Morphogen decay rate | 0.025 or 0.2 | |
| initial tissue size | 14 cells | of which 9 form the head |
| Gene and protein dynamics | | |
| gene product decay rate | 0.3 | |
| Hill constant of the TFBS | 60. | |
| gene transcription | 100. | |
| Mutational dynamics | | |
| Nr of gene types | 16 | |
| gene duplication | 0.006 | Note that with the gene, also its TFBS are duplicated. |
| gene deletion | 0.009 | |
| TFBS weight change | 0.001 | |
| TFBS type change | 0.001 | |
| TFBS duplication | 0.0015 | |
| TFBS deletion | 0.004 | |
| TFBS innovation | 0.001 | spontaneous emergence of new TFBS |
| Fitness | | |
| G: penalty per gene | 0.0005 | |
| T: penalty per TFBS | 0.00005 | |
| control period | 100 steps | Period over which gene expression stability is measured. |
| U: expression variance penalty | 0.1 | Penalty per cell that has a variance in segmentation gene level > 5.0 during the control period. |

celltype colour algorithm:

$$celltype = \sum_i^n s_i * i^3 \begin{cases} s_i = 0 & \text{if } G_i < ThOff \\ s_i = 1 & \text{if } ThOff < G_i < ThOn \\ s_i = 2 & \text{if } G_i \geq ThOn \end{cases} \quad (2.3)$$

If the cell expresses any genes, its celltype is translated into a colour as follows: $colour = celltype * 40 \% 251$. The *colour* number gives a point on the following colour gradient:

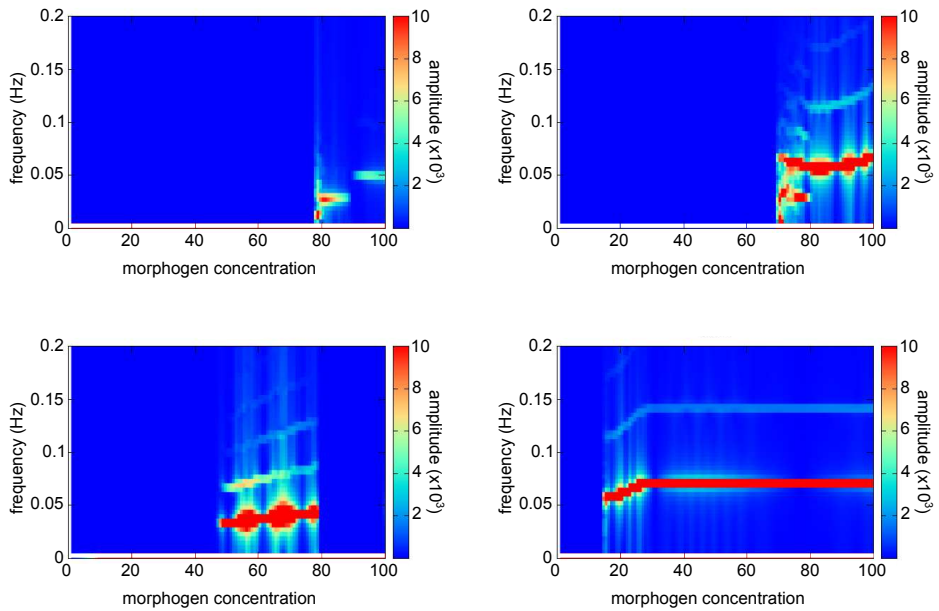


Figure S2.1. Examples of profiles that are harder to classify

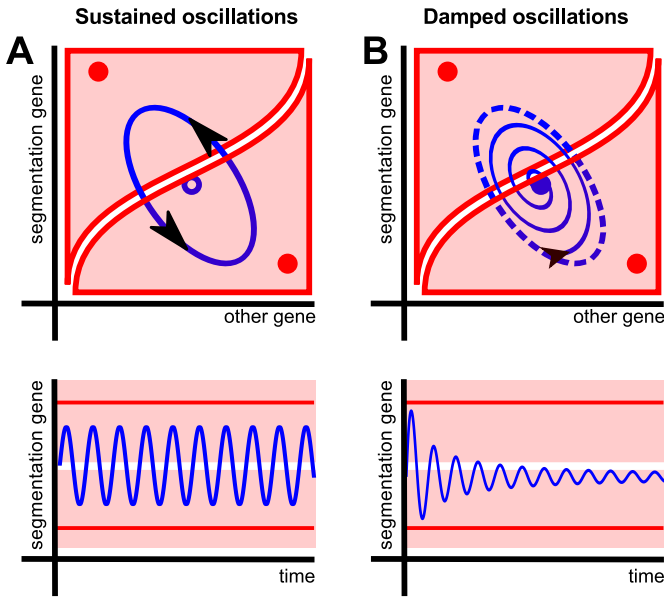


Figure S2.2. Networks with persistent and damped oscillations have different origins. A) Persistent oscillations are the result of a stable limit cycle around an unstable equilibrium (open blue dot). As long as conditions (e.g. morphogen concentration) stay constant, these oscillations continue indefinitely. When the morphogen concentration decreases, the system will reach either of the two stable states (red dots), depending on the basin of attraction (red zones) in which it finds itself. B) Damped oscillations are caused by a stable spiral. Even if all else stays constant, the oscillations lose amplitude over time, and the system will end up with fixed gene expression. Such a system “loses” the memory of the oscillations and thus of the phase with which it started.

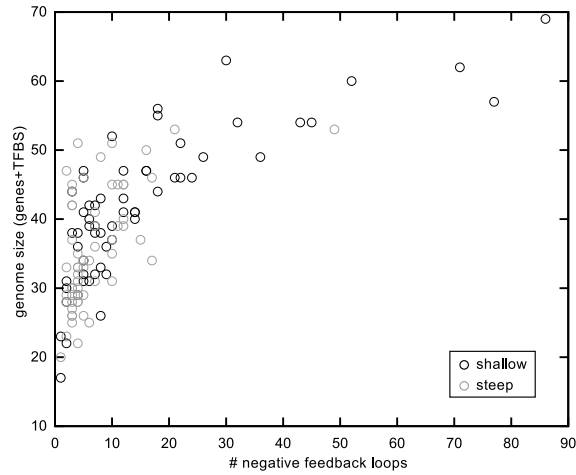


Figure S2.3. Larger genomes generate networks with more loops. Scatterplot of the number of loops in the network versus genome size. The two are clearly correlated, but note that particularly simulations with a shallow gradient (red dots) lead to larger genomes and networks with more loops.

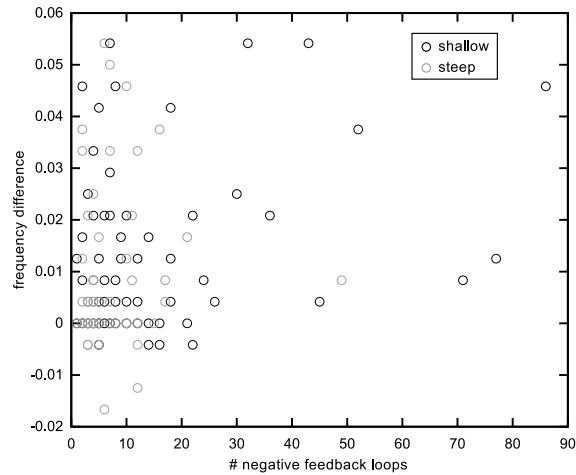


Figure S2.4. The type of frequency profile is not correlated with the number of loops Scatterplot of the number of loops in the network versus the posterior to anterior frequency difference in the profile.

3

Modelling asymmetric somitogenesis: a pointer to potentially non-neutral species differences

RENSKE M.A. VROOMANS AND KIRSTEN H.W.J. TEN TUSSCHER

Submitted

Abstract

Somitogenesis is one of the major hallmarks of bilateral symmetry in vertebrates. This symmetry is lost when retinoic acid (RA) signalling is inhibited, allowing the left-right determination pathway to influence somitogenesis. Intriguingly, the resulting left-right asymmetric phenotypes differ between the model species zebrafish, chicken and mouse. In all three cases, the clock speed becomes asymmetric, with slower gene expression oscillations on the right side. In zebrafish and mouse, the FGF8 gradient thought to constitute the wavefront also becomes asymmetric, extending further anteriorly on the right. While in zebrafish this leads to a transient delay in right sided patterning that subsequently becomes fully restored, in mouse patterning abnormalities may arise in the region on the right that experienced the delay.

While somitogenesis is generally considered as functionally equivalent among different vertebrates, substantial differences exist in the subset of oscillating genes. Variation also appears to exist in the way oscillations cease and somite boundaries become patterned. Here we apply computational models of somitogenesis to investigate the different asymmetry phenotypes. Specifically, we investigate to what extent differences can be explained from observed differences in FGF asymmetry and whether differences in determination may be involved.

We demonstrate that while a simple phenomenological clock and wavefront model can easily reproduce the chicken asymmetry phenotype, it does not readily reproduce the zebrafish and mouse phenotypes. Instead, to reproduce these phenotypes we developed models that explicitly take into account the experimentally observed dynamics of clock termination and boundary determination. Our results suggest that functional differences exist in somitogenesis between different vertebrates. Furthermore, they demonstrate that studying left-right asymmetry can unravel species differences that are not apparent from symmetric somitogenesis.

3.1 Introduction

The vertebrate body plan displays bilateral symmetry, for instance in the placement of limbs and cranial features; somitogenesis is one of the major hallmarks of this symmetry. The regular blocks of tissue patterned during somitogenesis later on give rise to the vertebrae, ribs and skeletal axial muscles. Somite pairs are generated periodically in an anterior to posterior direction from the presomitic mesoderm (PSM). This periodic patterning arises from a so-called clock and wavefront developmental patterning mechanism (Cooke and Zeeman, 1976, Hubaud and Pourquié, 2014). In the posterior part of the PSM, cells experience regular gene expression oscillations, called the somitogenesis clock (Palmeirim *et al.*, 1997, Resende *et al.*, 2014). As cells progress towards the anterior PSM, these oscillations become transformed into a stable spatial somite boundary pattern. This transition from temporal oscillations to spatial stripes is thought to arise from the transition of high to low morphogen concentrations as cells move anteriorly, and referred to as the wavefront (Aulehla and Pourquié, 2010). This process results in the periodic generation of pairs of somites flanking the notochord, with left and right somites being generated with identical timing and spacing. This symmetry becomes essential during later developmental stages when parts of the left and right somites fuse to form the vertebrae, and disturbances of somite symmetry can have severely disabling consequences such as scoliosis (Pourquié, 2011).

The somitogenesis clock, like all biological processes, is inherently noisy (Herrgen *et al.*, 2010, Jiang *et al.*, 2000). Therefore, additional levels of control are necessary to coordinate the behaviour of individual cells to ensure sharply delineated, coherent boundary formation and generate precise left right symmetry. While Delta-Notch mediated direct cell-cell signalling has been shown to synchronise directly neighbouring cells (Özbudak and Lewis, 2008, Soza-Ried *et al.*, 2014), the precise mechanism underlying left-right coordination have only been partly elucidated. Studies in zebrafish, chick and mouse have shown that somite symmetry can be compromised by disrupting the left-right signalling pathway (Brend and Holley, 2009, Kawakami *et al.*, 2005, Vermot and Pourquié, 2005, Vermot *et al.*, 2005). In a brief time window during somitogenesis, the signal of the left-right pathway passes through Hensen's node and the posterior PSM in order to confer a left- or right-handed identity to the more distal lateral plate mesoderm (LPM), from which most internal organs are generated (Brent, 2005). When components of this pathway (such as H^+/K^+ -ATPase (Kawakami *et al.*, 2005)) are knocked down, the travelling waves that lay down the somite prepatterning become asymmetric (Kawakami *et al.*, 2005, Vermot and Pourquié, 2005, Vermot *et al.*, 2005). Somite formation then becomes delayed randomly on one of the sides (without a clear bias to the left or the right) (Kawakami *et al.*, 2005), and somite alignment may be compromised ((Vermot and Pourquié, 2005) reviewed in (Brend and Holley, 2009)). Some components of the pathways involved in left-right determination are also involved in somitogenesis (FGF,

Delta-Notch and Wnt), and may become asymmetrically distributed during LR signalling (Boettger *et al.*, 1999, Huang *et al.*, 2011, Jacobs-McDaniels and Albertson, 2011, Kato, 2011, Kawakami *et al.*, 2005, Krebs *et al.*, 2003, Nakaya *et al.*, 2005, Raya *et al.*, 2003, Tanaka *et al.*, 2005, Ángel Raya *et al.*, 2004).

During normal somitogenesis, a compensatory mechanism must be active to counteract the asymmetric distribution of somitogenesis pathway components. This mechanism may then be responsible for the perturbed somite symmetry when the left-right pathway is inhibited. Experiments indicate that retinoic acid (RA) normally buffers the effects of the left-right pathway on somitogenesis, as somite symmetry is perturbed when RA is inhibited while left-right signalling remains unaltered (Kawakami *et al.*, 2005). In absence of RA, there is a clear bias in which side becomes delayed: the left side in chick and the right in zebrafish and mouse (Brent, 2005, Kawakami *et al.*, 2005, Sirbu and Duester, 2006, Vermot and Pourquié, 2005, Vermot *et al.*, 2005). In the latter two, the gradient of FGF8 also extends more anteriorly on the right. Since FGF8 is an important component of the determination front, this may explain the observed delay in zebrafish and mouse. The field of somitogenesis has a rich history of computational models to supplement experiments, but thus far somite left-right asymmetry has not been addressed. In this study, we take the first step by creating simple phenomenological models to reproduce the different asymmetry phenotypes observed in zebrafish, chick and mouse. We start with a phenomenological clock-and-wavefront model, in which a frequency gradient dictates the slowing of a simple phase oscillator. Somite pre-patterning occurs when a frequency of zero is reached (Ares *et al.*, 2012, Morelli *et al.*, 2009). With this, we investigated the necessary assumptions, parameter settings and their robustness needed to mimic the species-specific asymmetry phenotypes.

While the simple clock and wavefront model suffices to explain the chick asymmetry phenotype, it has difficulty reproducing zebrafish and mouse asymmetric somitogenesis. Since the three species differ substantially in the number and type of genes that oscillate (Krol *et al.*, 2011) and in the dynamics governing the onset of somite determination (Akiyama *et al.*, 2014, Niwa *et al.*, 2011), we develop more detailed somitogenesis models for mouse and zebrafish hoping to explain these differences. In the zebrafish model we incorporate experimental observations suggesting that a boundary of FGF signalling together with oscillation waves dictate somite determination (Akiyama *et al.*, 2014). Likewise in the mouse model we incorporate experimental observations that interactions between a non-slowng pErk oscillator and a slowng Notch oscillator together dictate oscillation termination (Niwa *et al.*, 2011).

Our analysis indicates that the more detailed models explain the zebrafish and mouse asymmetry phenotypes in a more straightforward and robust manner. These models also result in a more block-like specification of future somites well before oscillations cease,

which is in closer agreement to experimental findings (Niwa *et al.*, 2011, Shih *et al.*, 2015). In addition, while the more detailed models do not explicitly account for rostro-caudal somite polarity, they do suggest that the frequency gradient may provide temporal information for generating anterior-posterior polarity downstream of somite determination. Finally, the models make distinct, experimentally testable predictions. For instance, while the simple model predicts that the observed right-handed delay in zebrafish and mouse is caused by the right PSM skipping a few cycles of somite formation, the more extended models generate this delay without such cycle skipping. Thus, our analysis indicates that the mechanism by which oscillations cease and pre-patterning sets in warrants further study.

Glossary

tailbud: the posterior-most region where cells ingress and divide to contribute to extension of the main axis.

somite prepatterning: a determined somite whose cells have ceased to oscillate. Term used to distinguish the genetic process from somite morphogenesis, which is not included in the models.

PSM: presomitic mesoderm, the region of the paraxial mesoderm between the tailbud and the last-formed somite prepatterning.

frequency profile: A function describing how the frequency of oscillations in the PSM changes towards the anterior.

determination front: the point where oscillations stop and a somite prepatterning is formed.

travelling wave oscillator: the phase oscillator present in all simulations, which follows a decreasing frequency profile. Here used to represent oscillations of genes in the Notch

global oscillator: a phase oscillator whose frequency and phase remain uniform across the PSM.

morphogen: In the alternative models, we incorporate a morphogen which represents the FGF and/or Wnt ligand; it has a maximum in the tailbud and decays in cells in the PSM.

asymmetric somitogenesis: When in the model, we explicitly endow the right PSM with different parameters, such as oscillation frequency and either determination front position (clock-and-wavefront model) or morphogen decay. Note that the resulting somites may still end up symmetrically positioned.

catch-up somites: the somites that are formed on the right side upon return to symmetry. These are formed at a faster pace than the oscillations in the tailbud, in order to bridge the anterior-posterior difference in determination front position between left and right.

return to symmetry: when we explicitly shift the parameters back to symmetric values. Note that somitogenesis may remain asymmetric for a while after this shift.

3.2 Results

3.2.1 A clock-and-wavefront model of somitogenesis

experimentally observed asymmetric phenotypes First, we assess the asymmetric phenotypes in different vertebrate model species when RA synthesis is blocked, either chemically or genetically. In chick embryos in which RA synthesis is inhibited chemically, the first 7 somites form normally but subsequent left-sided somites are formed more anteriorly than those on the right (Fig. 3.1). Oscillations slow down on the right (Vermot and Pourquié, 2005) (see table 3.1), while the FGF8 gradient remains symmetric but extends more to the anterior, which is expected given the antagonistic interaction between FGF8 and RA (Diez del Corral *et al.*, 2003). In RA knock-out mice, after 9 symmetric somites have formed somitogenesis becomes delayed in the right PSM by up to 3 somites. Furthermore, the travelling waves desynchronise between left and right and the FGF8 gradient is shifted more anteriorly on the right (Fig. 3.1). The delayed right-hand somites are often asymmetrically positioned, and occasionally the entire region remains unpatterned (Vermot *et al.*, 2005). After left-right signalling has terminated, subsequent somites again form symmetrically. In zebrafish with inhibited RA synthesis, the asymmetry looks similar to that in mouse, with a delay occurring between somites 6 and 13 in the right PSM. In contrast to mouse, the delayed right side somites form at the same A-P position as the left, resulting in a fully symmetric spinal column (Kawakami *et al.*, 2005) (Fig. 3.1, see also the Glossary).

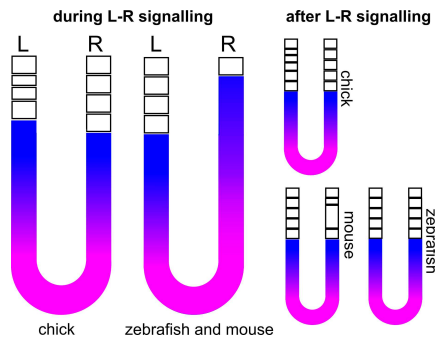


Figure 3.1. Asymmetric somitogenesis phenotypes after RA knock-down. RA-inhibited chick embryos have symmetric FGF8 distributions and skewed somite formation, where the somites on the left are placed more anteriorly. Mice and zebrafish without RA instead have asymmetric FGF8 distributions and delayed somite formation on the right. While it is not entirely clear whether chick embryos return to symmetric somite formation, mouse and zebrafish embryos do. In zebrafish, the “catch-up” somites (the somites that are formed more quickly to make up for the initial delay) are formed at symmetric positions compared to the left, while in mouse this is not always the case, and the catch-up region may even remain unpatterned.

the baseline model We first investigate whether we can capture these three different phenotypes with a simple model of somitogenesis (Jaeger and Goodwin, 2001, Morelli *et al.*, 2009). The model consists of cells forming a 2D tissue (the PSM), divided into a left and a right half (Fig. 3.2A). At the posterior end (the “tailbud”), new cells are added at regular intervals, reflecting ingression and division in the tailbud. In each cell, the somitogenesis clock is modelled as a phase oscillator. The frequency of oscillations is largest at the posterior end of the tissue and declines towards the anterior, reflecting the combined effect of the FGF and Wnt pathways on the clock. In this first baseline model, we incorporate this slowing down directly as a frequency profile: a function of the distance of each cell from the posterior end. The point where the oscillation frequency becomes zero represents the determination front. Here the temporal oscillations are transformed into a spatial pattern with cells memorising their latest oscillation phase (Morelli *et al.*, 2009)(Fig.3.2B, see also Methods). This frequency profile retracts at the pace with which cells are added to the posterior end, resulting in a constant size of the PSM. As has been shown before, the shape and length of the frequency profile determine the number of waves traversing the PSM: with a shorter and more nonlinear frequency profile, there are fewer travelling waves (Fig.3.2D-E) (Morelli *et al.*, 2009). The maximum frequency of the profile, which occurs in the tailbud, determines the pace of somitogenesis. We do not take noise into account and do not include cell-cell signalling; the effect of cell-cell synchronisation on the oscillation frequency has been extensively investigated before (Morelli *et al.*, 2009).

We start the simulations with symmetric left and right clock frequencies and wave-front positions, then at some point impose the asymmetry in frequency between left and right (Fig.3.2C). A reduction in right oscillator frequency is modelled by lowering the maximum oscillation frequency, while an anterior shift in FGF/Wnt gradient is implemented by allowing the anterior zero-point of the frequency profile to stay in place for a fixed duration, so that the profile transiently extends with the growing tissue rather than shifting posteriorly.

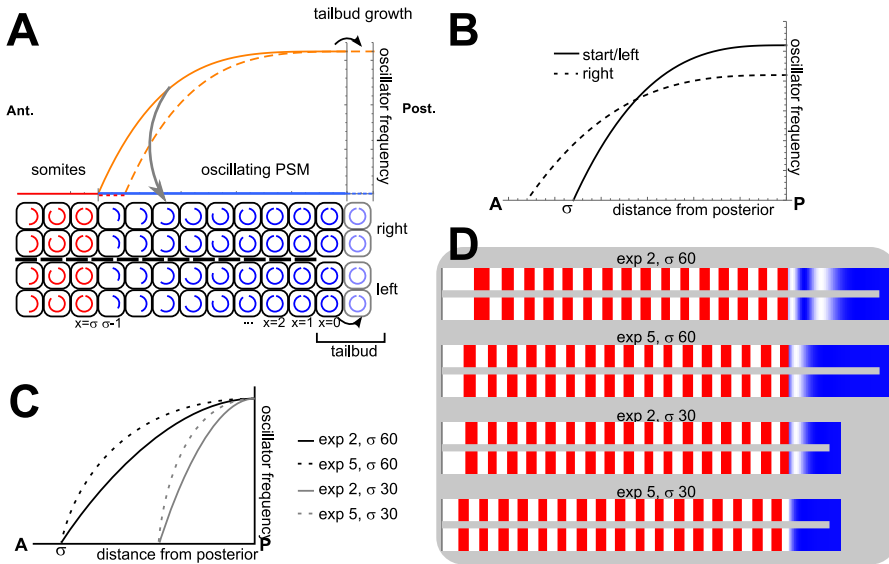
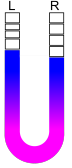
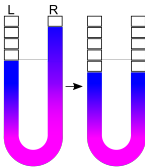
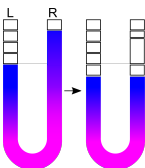


Figure 3.2. A simple model of somitogenesis. A) We model a 2D tissue of cells with a phase oscillator (Morelli *et al.*, 2009). The tissue is divided into a left and a right half. Cells are added to the PSM at the posterior end, and inherit the phase of the posterior-most cells. Above, we show the frequency profile (a negative quadratic function, see Methods). The point at which the frequency becomes zero retracts at the same speed as the extension of the PSM in the posterior. In the region where somites are defined (indicated in red), cells no longer oscillate. B) An example of how the left and the right may differ in frequency profile. C) Different frequency profiles of the simulations shown in D, differing in steepness (exp) and extent (σ). D) The number of waves traversing the PSM for the different shapes of the frequency profiles. The grey line in the figures is superimposed for clarity, and when waves are transformed to somites, we define the colours with a sharp cut-off for greater visibility.

Table 3.1. phenotypes of model organisms during somitogenesis

| organism | genetic properties | | left-right asymmetry in absence of RA | | | | | |
|-----------|---------------------|----------------------|---|------------|--------------------------|--------------------------------|------------------|--------------------|
| | pErk dynamics | oscillating pathways | left-right phenotype | Slower osc | FGF8 | delay (somite nr) | somite size diff | return to symmetry |
| chick | smoothly retracting | FGF, Wnt, Notch |  | right side | symmetric, more anterior | no; left somites smaller | yes | unclear |
| zebrafish | retracts in jumps | Notch |  | right side | right side more anterior | right side 2-3 somites delayed | no | yes |
| mouse | oscillates | FGF, Wnt, Notch |  | right side | right side more anterior | right side 2-3 somites delayed | sometimes | yes |

asymmetric oscillations and the tailbud The most general property of left-right asymmetry in all studied vertebrate species is the difference in oscillation frequency between cells in the left and right PSM. We first studied the effect of such a difference without modelling any change in the FGF gradient. Theoretically, two possibilities exist: either the left-right signalling pathway confers a frequency difference in the entire PSM, including the tailbud (Fig. 3.4B), or the difference in frequency only occurs anterior to the node, where left and right paraxial mesoderm are separated by the notochord (Fig. 3.3D). If the former is assumed to be the case, we observe that oscillation phase diverges between left and right in the tailbud, and the slower pace of oscillations on the right necessarily results in larger and fewer somites compared to the left (Fig. 3.3A-B).

However, in experimental images we do not observe clear phase differences in the tailbud, suggesting that cells in the tailbud oscillate synchronously (Kawakami *et al.*, 2005, Saúde *et al.*, 2005, Vermot and Pourquié, 2005, Vermot *et al.*, 2005). The node, which is considered the left-right signalling centre (Komatsu and Mishina, 2013), lies anterior to the tailbud and exerts its effect laterally. This would imply that the tailbud is not much affected by the left-right system. We therefore assume in all subsequent simulations that cells have the same oscillation frequency and phase in the tailbud before entering the PSM on the left or the right, after which the frequency decreases on the right (Fig. 3.3D). This yields “phase forcing” of the oscillators in the right PSM, with the faster oscillations in the tailbud taking over control and driving the oscillation pace of the slower PSM. As a consequence, right-hand somites return to a size dictated by the tailbud frequency shortly after the left-right asymmetry is imposed (Fig. 3.3C). Because of the transition period in which the somites on the right were larger, somites on the right have become shifted posteriorly, resulting in fewer somites formed on the right (Fig. 3.3C).

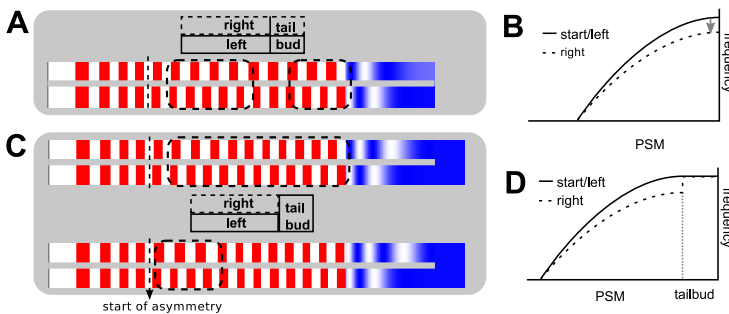


Figure 3.3. Tailbud dynamics influence the course of asymmetric somitogenesis. A -B) Asymmetric frequency profile throughout the entire tissue (also tailbud). The normal period is 90 minutes. At one third of the simulation, the period on the right side increases by 15 minutes. C-D) The tailbud is symmetric, the PSM is not. Top: period on the right is 15 minutes longer. Bottom: right period is 30 minutes longer. Regions with clear asymmetry are indicated with dashed boxes.

simulating chick asymmetric somitogenesis In chick, inhibition of RA synthesis did not induce obvious asymmetries in the FGF gradient (Vermot and Pourquié, 2005). Instead, it was symmetrically shifted anteriorly compared to wild-type embryos, as is expected from the known antagonism between RA and FGF (Diez del Corral *et al.*, 2003). Given the role of FGF in keeping cells in an undifferentiated oscillatory state (Dubrulle *et al.*, 2001), the anterior shift implies an extension of the PSM region supporting oscillations. Thus, to specifically model chick somitogenesis we applied a symmetric anterior shift of the frequency profile. The anterior shift of the frequency profile imposes a transient delay in somite formation, causing the travelling waves to compact further and yielding a number of smaller somites on both the left and the right (Sirbu and Duester, 2006, Vermot and Pourquié, 2005, Vermot *et al.*, 2005). The decrease in somite size is smaller on the right due to the simultaneously imposed lower oscillation frequency on the right. The resulting left right differences in somite size cause an offset in left and right somite positioning similar to those observed for the simulations without the determination front shift (Fig. 3.3C). Since somites on the left are smaller shortly after inhibiting RA synthesis, it appears as if somitogenesis is delayed on the left (Brent, 2005) even though the clock actually goes faster on the left. Later on, as tailbud forcing sets in and left and right somites have regained similar sizes it depends on the amount of oscillator slowing and frequency profile asymmetry whether or not symmetry becomes restored (compare top and bottom panels in Fig. 3.4A). When after a while we restore oscillation frequencies and extent of frequency profiles to their values before LR signalling (symmetric conditions), a second opposite transient occurs correcting the earlier arisen asymmetries between left and right. In the end, the left and right count the same number of somites (Fig. 3.4C, Video 1). A symmetric anterior shift of the determination front, paired with a lower oscillation frequency on the right, thus accurately captures the asymmetric phenotype of chick somitogenesis in absence of RA.

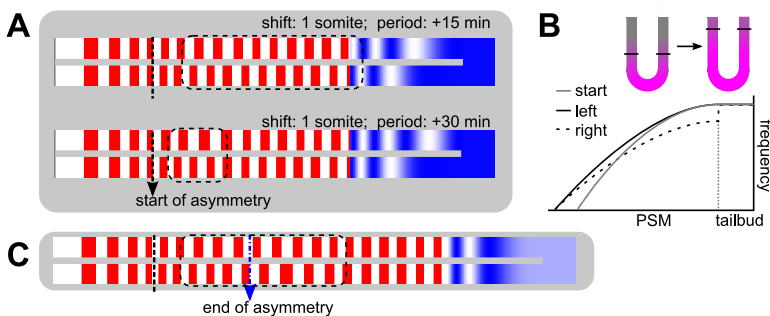


Figure 3.4. Symmetric determination front shifts during chick somitogenesis.

A-B) Asymmetry combined with an anterior shift of the frequency profile (increase in σ , measured relative to normal somite size) on both sides. Left sided period is 90 minutes. C) Simulation in which the frequency and determination front position are shifted back to normal values. Left period: 90 minutes, right period: 105 minutes, shift: 2 somites.

simulating zebrafish and mouse asymmetric somitogenesis In order to simulate zebrafish or mouse somitogenesis, we need to take into account that the FGF8 gradient shifts ahead further on the right, causing a larger anterior shift of the oscillating region and increased delay in somitogenesis. To investigate the effect of an asymmetric shift in frequency profile in isolation and in the simplest possible manner, we initially only shift the frequency profile on the right while maintaining the left frequency profile and left and right oscillator frequencies constant (using mouse oscillator frequencies (period of 120 min) and somite size as an example; Fig 3.5A, left). We observe a delay in the formation of somites on the right, with somites already pre-patterned on the left that have not yet been determined on the right (Fig 3.5A, right, top panel). Once the frequency profile shift is complete, one or more smaller somites are formed on the right (Fig. 3.5A left, bottom panel), shifting the position of all following somites to become asymmetrically positioned.

Since in zebrafish (and occasionally in mouse) the “catch-up” somites are formed symmetrically, we investigate whether the asymmetric positioning of somites in our simulations can be corrected by taking into consideration the simultaneously occurring lower oscillation frequency on the right (Fig. 3.5B, left). While the lower frequency indeed causes a temporary increase in somite size on the right, it cannot compensate for the small size of the first somite that formed directly after the anterior shift of the determination front (Fig 3.5B right). When we tune the left-right frequency difference and anterior shift relative to each other, we do eventually end up with bilaterally symmetric somites again but only after this transient with asymmetric somites (Fig. 3.5B, left, lower panel). This may suffice to explain mouse somitogenesis, but in zebrafish the “catch-up somites” are formed at a symmetric position with the left side, which is now clearly not the case.

So far the decrease in oscillation frequency and anterior shift of the frequency profile were applied at the same time. When we instead assume that the oscillations on the right slow down before the profile starts to shift, we observe that the shift may compensate almost exactly for the increase in somite size that would otherwise ensue (Fig. 3.5C, right, top panel), although the first few somites on the right do remain a little larger. When the anterior shift is larger, also a larger decrease in oscillation frequency is required, otherwise the compensation does not suffice ((Fig. 3.5C, right, bottom panel).

Next, to completely mimic the experimental phenotype upon the blocking of RA, we add the lesser anterior shift of the left determination front and a return to symmetric parameters (frequency and determination front position) after a number of tailbud oscillations (Sirbu and Duester, 2006). We find that under these conditions it becomes considerably harder to tune the size and timing of frequency difference, frequency profile shifts and return to symmetrical conditions (Fig. 3.5D, Video 2 and Supp. Fig. S3.1) to obtain

symmetrically formed somites. We repeated the simulations for zebrafish oscillator frequencies (period 30 min), and found that the tuning is equally difficult (Supp. Fig. S3.2).

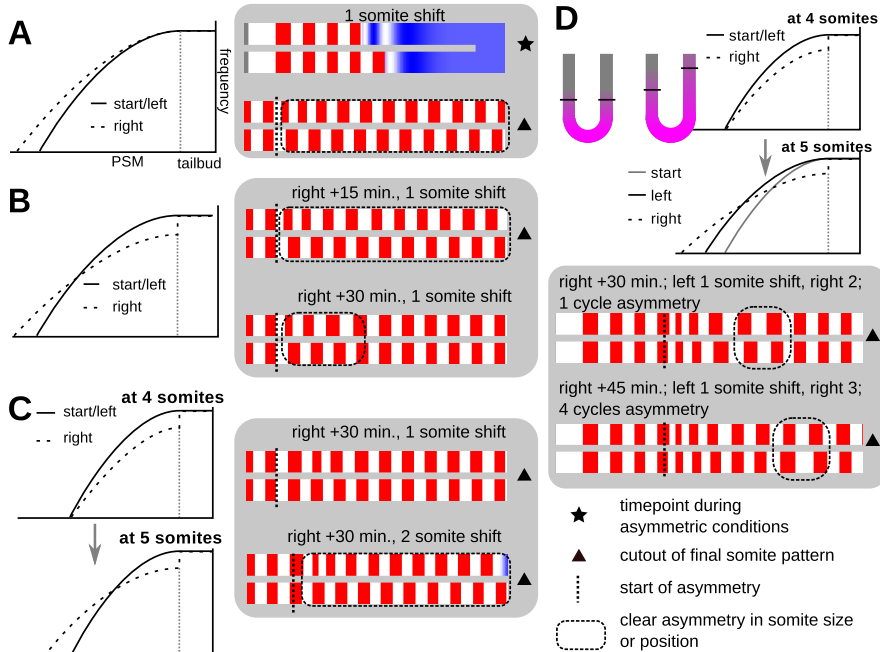


Figure 3.5. The effect of an anterior determination front shift in the right PSM during mouse and zebrafish somitogenesis. A) Simulations in which only the extent of the frequency profile (σ) is increased on the right, at one third of the simulation. B) Simulations combining a drop in oscillation frequency on the right with the anterior frequency profile shift. C) Simulations in which the frequency drop comes one cycle before the anterior shift. D) Simulations in which the frequency profile also extends by one somite length on the left. Furthermore, after a number of cycles the conditions are restored to symmetrical values. Starting values: oscillation period = 120 minutes, $\sigma = 60$ cells. Values for the right PSM upon L-R signalling indicated in the figure.

conclusion Thus, our modelling predicts that if the implicit assumptions of this simple model are met, in the right paraxial mesoderm of zebrafish oscillation speed decreases prior to the anterior shift of the determination front in order to ensure symmetric somite formation. However, our modelling also predicts that mouse and zebrafish asymmetry phenotypes are highly non-robust to the precise magnitude and timing of changes in wavefront and oscillation frequency, indicating that the assumptions of this model may in fact not be met during mouse and zebrafish somitogenesis.

3.2.2 Extended models of somitogenesis

introduction Although the clock-and-wavefront model is widely accepted, the exact mechanism by which the oscillations are translated to a somite pre-pattern is still debated (reviewed in (Hubaud and Pourquié, 2014)). While there is evidence for a conserved role of boundary determination genes such as pErk, Mesp, Ripply and Tbx6 in zebrafish, chick and mouse (for review, see (Yabe and Takada, 2016)), there are also large differences in the number and type of oscillating genes between these species (Krol *et al.*, 2011) (Table 3.1). Based on these differences, we hypothesise that the transition from oscillations to a determined somite and the feedback of this process on the progression of the determination front could be regulated differently in these model organisms. Since this could potentially explain the different left-right phenotypes in zebrafish and mouse, we aimed to create more accurate models of somitogenesis for these organisms based on experimental data.

In the models we develop next, we keep the basic phase oscillator from the clock-and-wavefront model, that generates travelling waves due to its slowing across the PSM. Hereafter we will refer to it as the travelling wave oscillator, using it as token for the slowing oscillatory dynamics of Notch and Her/Hes proteins. We add an explicit posterior morphogen gradient, representing FGF and/or Wnt signalling. This morphogen is highly expressed in the tailbud and slowly decays in the rest of the PSM, in agreement with experimental data (Fig. 3.6A). We assume that its level dictates the frequency profile of the travelling wave oscillator in a nonlinear manner, such that the resulting frequency profile yields realistic-looking travelling waves (Fig. 3.6A and Methods). Next, we decouple the frequency profile from final somite determination. That is, rather than assuming that somite determination occurs when oscillations stop (frequency becomes zero), we instead implement species-specific rules for somite determination. As we will discuss below, this often leads to somite determination well before oscillations fully cease. As a consequence, somite polarity can no longer arise from strict memorisation of oscillator phase upon somite determination, which most current clock-and-wavefront models tacitly assume.

In the sections below, we describe what is known about the determination mechanism in zebrafish and mouse, and distil a simple set of rules for somite determination for the models. We then investigate the consequences of these rules for reproducing the left-right asymmetry phenotype.

Zebrafish somitogenesis

experimental data Of the three vertebrate model organisms discussed in this study, zebrafish may be expected to have the simplest mechanism for somitogenesis. It has the fewest oscillating genes, which mostly belong to the Notch pathway; none of the FGF and Wnt pathway genes appear to oscillate (Krol *et al.*, 2011). Yet, already in zebrafish there are substantial feedbacks between at least the FGF and Notch pathways that contribute to somite pre-pattern formation (Akiyama *et al.*, 2014), creating a more complex mechanism than represented by the simple model.

In zebrafish, it was shown that the boundary of pErk expression (an effector gene downstream of FGF) determines the somite boundary of future somite S-IV, well before the first signs of somite polarity determination such as *Mespa/b* expression arise (Akiyama *et al.*, 2014) (Fig. 3.6A). While the FGF8 gradient itself retracts smoothly with the extending body axis, pErk becomes displaced by one somite length after each clock cycle, and this behaviour depends on Notch oscillations (Akiyama *et al.*, 2014). *Tbx6* is another protein important for boundary positioning, whose expression boundary lies one to two somites anterior to the pErk boundary (Wanglar *et al.*, 2014). *Tbx6* becomes displaced from posterior to anterior in somite S-III through inhibition by *Ripply1/2* (Wanglar *et al.*, 2014), which may represent an indirect dependence on Notch signalling. Thus, changes in pErk and *Tbx6* mark the early stages of somite boundary patterning and determination. We surmise that once FGF8 levels have fallen below a certain threshold level, the anterior PSM can start differentiating towards somites, with subsequent waves of Notch signalling initiating the sequence of differentiation events by repressing first pErk, then *TBX6*.

While somite boundary determination occurs posteriorly in the zebrafish PSM (at somite S-IV), Notch pathway oscillations continue up to position S0, where somite morphogenesis takes place (Shih *et al.*, 2015). The function of these continued oscillations this far anterior to somite boundary formation is unclear. A potential function could be that later waves of Notch expression switch off (or on) other genes (analogous to the switching off of pErk and *TBX6* expression), thereby inducing further steps in somite determination and boundary refinement (Shih *et al.*, 2015). The period of Notch oscillations increases toward the anterior, but only by about 50%. Thus, oscillations do not slow down infinitely before somite formation, as often assumed in models ((Morelli *et al.*, 2009, Murray *et al.*, 2011, Tiedemann *et al.*, 2012) and the models above). It is still debated what controls this oscillator period: while some research reported a link between FGF8 and frequency via regulation of *Her13.2* (Kawamura *et al.*, 2005), other studies reported no relation between the concentration of FGF8 and frequency (Webb *et al.*, 2016), maintaining that FGF8 represents only a permissive signal for oscillations but does not impact period.

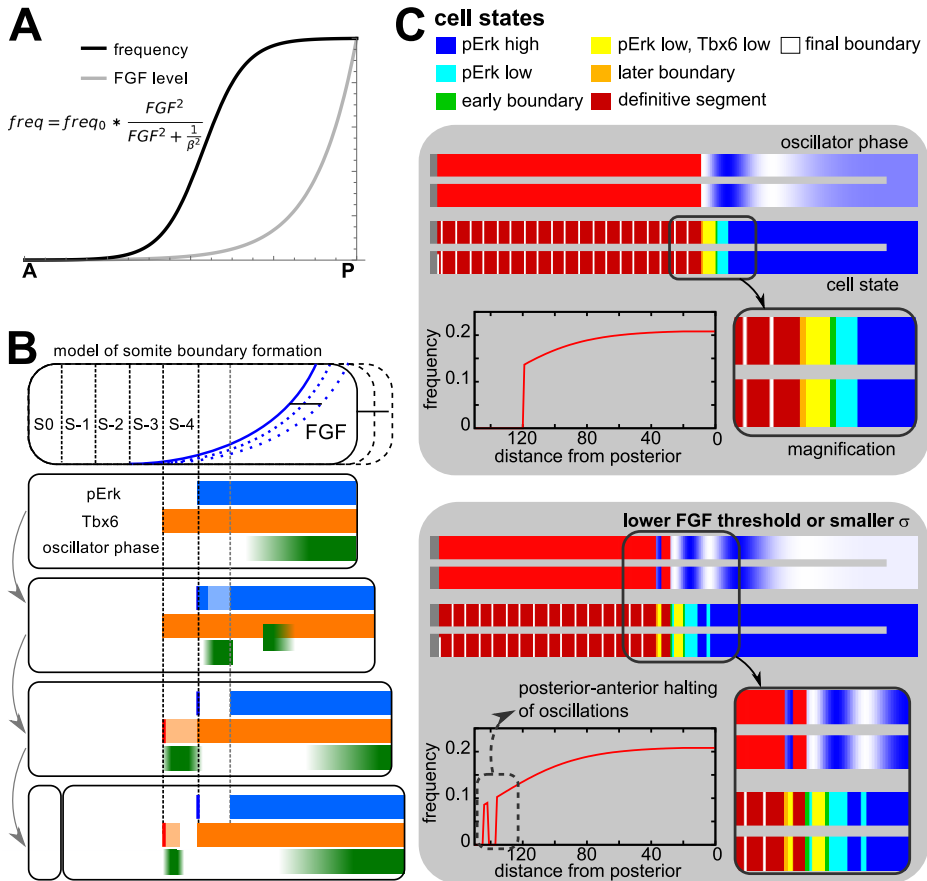


Figure 3.6. A new model for zebrafish somitogenesis. A) Cells in the model have an explicit FGF8 level, which is 1 in the tailbud and decays with a fixed rate in cells in the PSM, mimicking the mRNA gradient. From this level, we derive the frequency of oscillations with the given equation. B) Cells start in a so-called “pErk high, Tbx6 high” state in the posterior PSM. After their FGF8 level drops below a certain defined threshold (grey dotted line), they become competent to change to a “pErk low, Tbx6 high” state, which happens when they experience a peak in the oscillator phase. When another peak passes, they transition to a “Tbx6 low” state, which we for now consider to be the definitive segment state, upon which oscillations cease. For simplicity, boundary cells are defined as the last ones in a neighbourhood to transition to a new state, therefore switching on *Mespa/b*. Cartoon inspired by (Saga, 2012, Wanglar *et al.*, 2014, Yabe and Takada, 2016) C) Examples of simulations with different σ (which determines at which concentration of FGF8 the frequency declines), or the FGF8 level at which cells can change to a ‘pErk low’ state.

model In our extended zebrafish model, somite formation depends on both the morphogen (FGF8) gradient and on the cell state when a peak of the travelling wave (Notch/Her) oscillator passes (Fig. 3.6A,B). Specifically, cells start in the tailbud in a “pErk high, Tbx6 high” state. They remain in this state until they fall below a certain FGF8 threshold level and experience the passing by of a wave of the travelling wave oscillator. Cells then switch to a “pErk low, Tbx6 high” state. Similarly, the next wave of the travelling wave oscillator causes a switch to the “pErk low, Tbx6 low” state. Next, oscillations halt, leading to a fully determined somite (Fig. 3.6B). Cells at the somite boundaries obtain a state distinct from their neighbours; this is a simplification based on the observation that Tbx6 seems to switch on *Mespa/b* at the anterior boundary of its expression domain before disappearing (Wanglar *et al.*, 2014). Also note that in our model, oscillations cease earlier (around somite S-III) than has been observed experimentally (Shih *et al.*, 2015). We consider this to be a reasonable approximation given that we are primarily interested in boundary formation and not somite determination. For more details we refer to the methods section.

symmetric model results In figure 3.6C we show simulations with different parameters. The top pictures show the phase of the oscillator in the PSM, the bottom pictures show the cell states described above. We find that the number of travelling waves in the PSM can be increased by lowering the FGF8 threshold for pErk state switching or by decreasing σ (which determines for which FGF8 level half the maximum frequency occurs). Both result in a larger PSM and cells at the anterior end of the PSM having a lower oscillation frequency (figure 3.6C, bottom). Decreasing the decay rate of FGF8 simply extends the PSM and increases the wavelength of the waves, but not their number (not shown). For all parameter settings, oscillations cease in a posterior to anterior manner within a forming somite, which is in agreement with experimental observations (Shih *et al.*, 2015). Also note how segments form before the frequency of oscillations is zero or even close to zero; this is a consequence of the explicit state transitions shutting down oscillations before the FGF8 concentration (and hence frequency profile) is so low that oscillations cease by themselves. While the continuation of oscillations beyond the point of first boundary determination precludes the memorisation of oscillator phase to allow for somite polarity establishment, the observed posterior-to-anterior progression of the oscillation waves may provide time-dependent polarity information. Indeed, experimental data indicate that polarity establishment occurs downstream of boundary formation (Hubaud and Pourquié, 2014).

asymmetric model results: isolated frequency or FGF decay changes To simulate the increase of FGF8 in the anterior right PSM, we reduce the decay rate of the morphogen on the right after the formation of four somites, which leads to more FGF8 in the anterior PSM with some delay. We maintain symmetric oscillation frequencies and impose a return to symmetric decay values five cycles later. This yields a few smaller

somites about three cycles after the introduction of the asymmetry, which creates the illusion of a delay on the right because the PSM extends further anteriorly, while the number of somites is the same (Fig. 3.7A). Upon return to symmetric values, a few oversized somites are formed that cause a return to symmetric somitogenesis. When instead the morphogen decay remains symmetric and only the frequency of oscillations is decreased on the right, somite formation there is truly delayed compared to the left. After four delayed but equally sized somites have been formed, somites on the right become larger than those on the left, so that somite formation becomes asymmetric (Fig. 3.7B). When the frequency on the right is returned to normal, a few smaller somites are created after which symmetry is restored. We thus observe that frequency decrease and FGF8 have opposing effects on somite size. Also note that the delay between gradient or frequency modification and effect on somitogenesis naturally follows from the predetermined somite states in the model, which means that a cluster of cells posterior to the last-formed somite already has determined somite boundaries and are unaffected by the change.

asymmetric model results: combined frequency and FGF decay change When both frequency and decay rate are reduced on the right (and returned to normal after five cycles), there are combinations of decay rate and frequency change that yield delayed somite formation (lagging behind 2-3 somites) which subsequently becomes restored into symmetric somitogenesis, in agreement with experimental data (Kawakami *et al.*, 2005) (Fig. 3.7C, Video 3). With a larger difference in frequency, the change in decay rate should also be larger to ensure maintenance of symmetric somite sizes. Still, somitogenesis remains roughly symmetric even for decay rates diverging to about 10% from the “perfect” compensatory decay rate (Supp. Fig. S3.3).

conclusion Similar to the baseline clock-and-wavefront model, our extended zebrafish model predicts the need to coordinate changes in oscillation frequency and wavefront position to obtain delayed but symmetric somitogenesis. However, in the baseline model there was an additional need to tune the timing of frequency and wavefront changes. In the extended model the FGF8 decay dynamics automatically cause a delay in the onset of determination front displacement. Finally, it appears that less precise tuning is required in the extended model, suggesting a more robust mechanism. Furthermore, this model has a more natural explanation for the delay in somite formation on the right: a combination of delayed FGF8 gradient retraction and a slower clock, which induces an increase in the time needed to change the state of already-competent cells. This in contrast to the baseline clock-and-wavefront model, in which the determination front had to be kept in place artificially for a certain number of cycles to create a delay in somite formation.

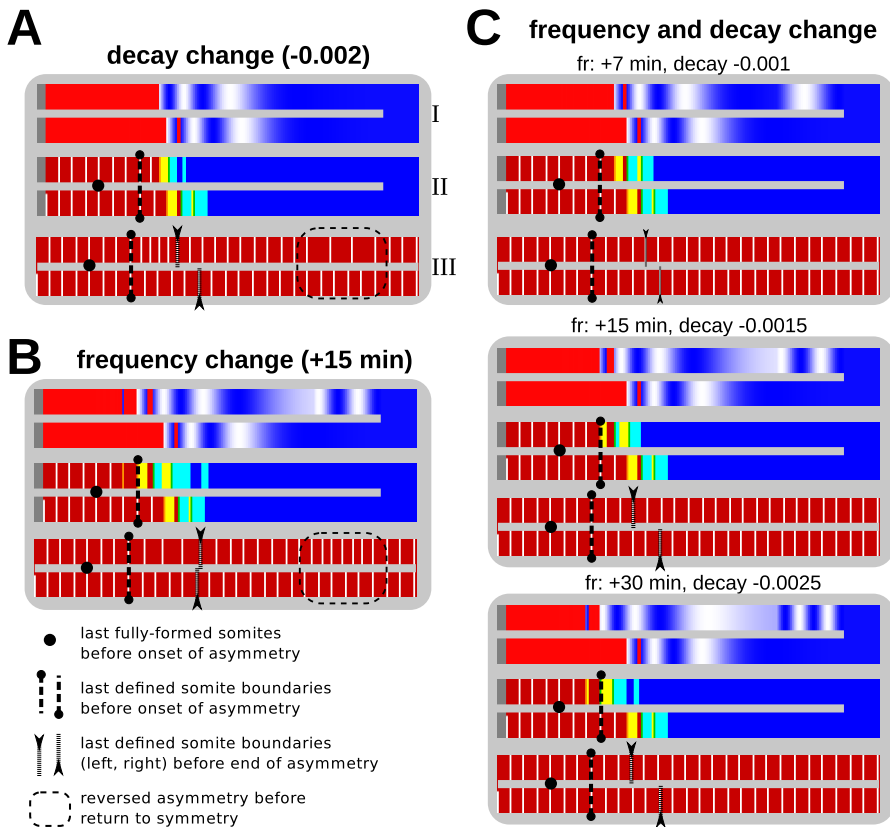


Figure 3.7. Frequency change and FGF8 increase can compensate each other. Pictures display oscillation state during asymmetry (I), cell states during asymmetry (II) and all boundaries formed during the asymmetry period (III). The frequency changes and decay rate changes are for the right side only, starting at the 4 somite stage and ending at the 9 somite stage. Note the considerable delay between the change in parameters and a change in somite formation, due to the pre-emptively formed somite boundaries. In the right column, the change in frequency and decay are chosen such that symmetric somitogenesis is obtained, with never more than a one-cell shift in boundary position.

Mouse somitogenesis

experimental data During mouse embryogenesis, many components of the Wnt, FGF and Notch signalling pathway oscillate (Krol *et al.*, 2011). For instance, the FGF8 signalling pathway effector pErk oscillates in the mouse PSM, and these oscillations are thought to determine the pace of somitogenesis (Harima and Kageyama, 2013, Niwa *et al.*, 2011). Rather than forming a travelling wave across the tissue like the oscillating genes of the Notch pathway, pErk oscillates roughly synchronously across the entire PSM, with a frequency similar to that of Notch in the tailbud (Niwa *et al.*, 2011). As a

consequence, in the anterior where Notch oscillations slow down, a frequency difference and hence a phase difference arises between the two oscillators. It was shown that somite boundary formation occurs in the anterior PSM when Tbx6 expression coincides with a wave of Notch expression and low expression of the pErk oscillator. Mesp2 expression represents the earliest sign of this boundary, and subsequently switches off Tbx6 and regulates somite rostro-caudal polarity (Oginuma *et al.*, 2008, Yabe and Takada, 2016).

model Experiments have shown that pErk oscillates due to inhibition by Dusp4 and Sprouty4, which are in turn regulated by Hes7, a component of the Notch pathway (Hayashi *et al.*, 2009, Niwa *et al.*, 2011). So also in mouse, there is a feedback from the Notch pathway on pErk. An intriguing question is how the pErk oscillations can maintain a constant frequency across the PSM despite the influence of the travelling wave. Here we simplify matters by ignoring this unresolved interdependence between the two oscillators, and instead incorporate pErk dynamics by adding a “global” oscillator to the model, with the same phase and frequency everywhere in the tissue, independent from the travelling wave (Notch) oscillator. Notch oscillates with the same frequency as the pErk oscillator in the tailbud, but its phase diverges from pErk in the anterior PSM as the Notch oscillator slows down due to declining morphogen levels (Fig. 3.8A). Somites form in the anterior when the expression of the pErk oscillator is low, at the position where the Notch oscillator is high (Fig. 3.8A; these states are defined in terms of particular phases of the two oscillators, for more details see methods section).

symmetric model results In figure 3.8C and D, simulations with this model are shown without left-right asymmetry. The top pictures only show PSM, somites and somite boundaries, while the bottom pictures show the phase of the Notch oscillator at the time of somite formation. Note that we do not assume that this reflects the final rostro-caudal polarity of the somites. Indeed, differences in the frozen Notch oscillator phase between different simulations easily arise due to different frequencies in the anterior-most cells at the time of somite formation, indicating that this would not represent a robust somite polarity patterning mechanism (compare the two simulations in figure 3.8C).

In principle, this model does not need a threshold morphogen level to determine where somitogenesis can occur like the zebrafish model does; the two oscillators suffice, since the frequency profile ensures that the peak of Notch only coincides with a low phase of the global oscillator in the anterior, where the Notch oscillator has slowed down (Fig. 3.8B). Under these conditions, the frequency profile of the Notch oscillator determines the size of somites. The behaviour of the model without such a threshold is shown in figure 3.8C. It can be observed that only a single peak of Notch travels across the tissue at any given time, while experiments suggest that there could be an additional wave of Notch initiating right before somite formation (Niwa *et al.*, 2011). Incorporating an FGF8 threshold (above which somitogenesis is inhibited), allows for more than one

peak to form because it can delay the moment at which somites can be formed, retaining multiple travelling waves in the PSM (Fig. 3.6C, Fig. 3.8C).

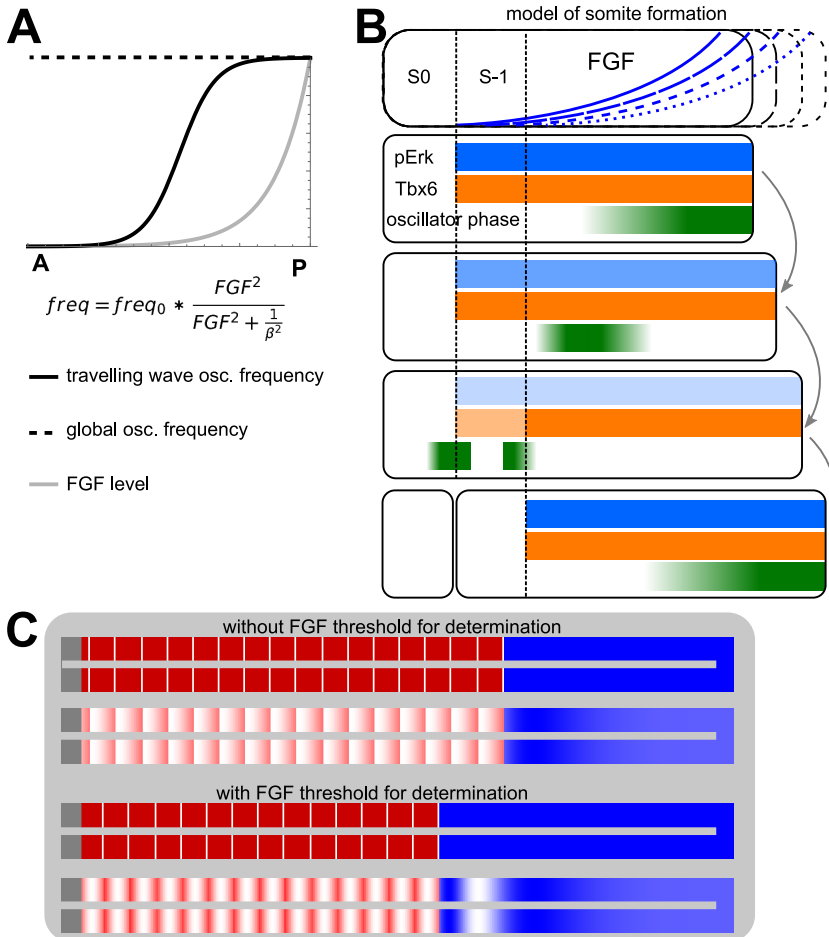


Figure 3.8. The two-oscillator model. A) The frequency profile of the “global” pErk oscillations and the travelling-wave Notch oscillations. B) The formation and freezing of a wave in the two-oscillator model. The Notch oscillator forms a travelling wave towards the anterior, which creates a new somite when the pErk oscillation is at its minimum. We ignore somite polarity formation. Cartoon inspired by (Saga, 2012, Yabe and Takada, 2016). C) Examples of somitogenesis with the two-oscillator model. First row: without an FGF level threshold at which somitogenesis can take place. tailbud oscillation period: 120 minutes, σ : 10, FGF decay: 0.002, frequency profile exponent: 2.5. Second row: same parameters, but now with with an FGF level threshold at which somitogenesis can take place. threshold: 0.1 a.u.

pErk dynamics in absence of RA Currently, we lack data on the pErk dynamics in absence of RA, and hence in the presence of left-right asymmetries. While we know that the domain of FGF expression extends further anterior in the right PSM (Vermot *et al.*, 2005) and that the frequency of Her oscillations is decreased, it is unclear whether or not pErk oscillates symmetrically and with what frequency. We tested two possibilities: either pErk oscillations remain symmetric with a frequency corresponding to the Notch oscillator frequency in the left tailbud, or pErk oscillations on the right become slower than on the left, with both sides having a frequency corresponding to the Notch oscillator in the tailbud at the respective sides. In both scenarios we incorporated a lower frequency for Notch oscillations on the right, and a lower morphogen decay.

asymmetric model results: symmetric pErk When the pErk oscillation frequency is assumed to remain symmetric (but the Notch oscillations are not), somite formation cannot be delayed on the right: the timing is determined by pErk, which still reaches low levels simultaneously on the left and right. The slower Notch oscillations lead to a number of very large somites on the right, followed by some very small somites after the symmetry in morphogen decay and Notch oscillation frequency is restored (Fig. 3.9A, top). A delay on the right can be created by adding the threshold morphogen level for somite determination mentioned before. When the asymmetry sets in, the lower morphogen decay on the right causes an increase in morphogen (FGF/Wnt), which can force somitogenesis to skip one or more cycles because the entire right PSM is still above the threshold. This however leads to a sustained lagging of somitogenesis on the right, even upon return to symmetric parameter values (Fig. 3.9A, middle). This means that symmetry is never restored, unlike what is observed experimentally. Addition of a second, lower morphogen threshold, which forces somitogenesis to happen regardless of the phase of either pErk or Notch oscillator, resolves this by allowing the right PSM to catch up with the left (Fig. 3.9A, bottom, Video 4).

Although both thresholds seem reasonable, the lower threshold does leave open the question of how polarity will be established in the somites formed by the passing of this threshold. The transition from undetermined to determined tissue is now created by a smoothly anterior-to-posterior retracting wavefront dictated by FGF8 dynamics, instead of blocks of tissue being sectioned off into a somite simultaneously due to the posterior-to-anterior passing of a wave of Notch expression. The phase of the Notch oscillator when this smooth wavefront passes may somehow still aid in the determination of rostral-caudal somite polarity and boundary formation (as is the case in the simple model). On the other hand, the second threshold could explain why aberrant tissue determination is observed in some mouse embryos that are mutant for RA production (Vermot *et al.*, 2005).

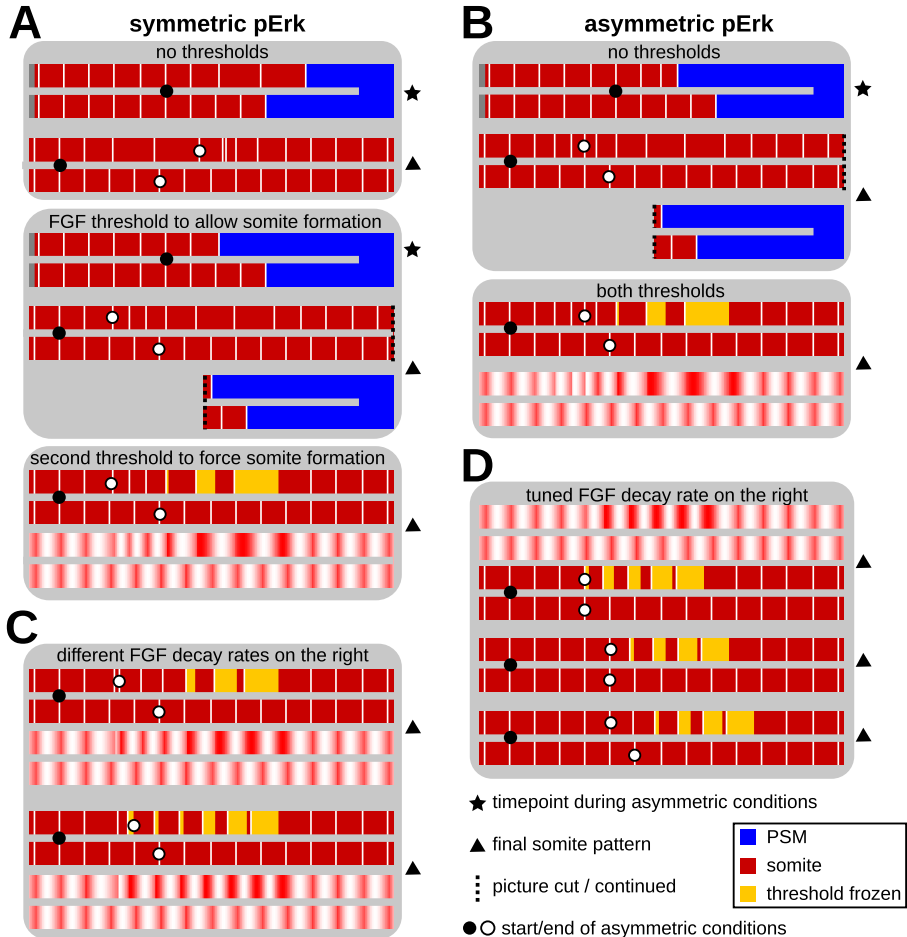


Figure 3.9. FGF levels determining somite formation are required for correct reproduction of LR phenotypes. Legend indicates the different cell types; “threshold frozen” are the cells that stopped oscillating due to the passing of the explicit FGF threshold rather than the somitogenesis mechanism. A) Simulations where pErk oscillations are symmetric. The first FGF threshold, restricting where somitogenesis can take place, is required to create a delay. The second, a minimum FGF level required to keep cells in a non-determined state, is required to restore somite symmetry after frequency and FGF decay on the right have gone back to normal. B) Simulations where pErk oscillations are asymmetric. No thresholds are required to create a delay, but the minimum FGF level is required to restore symmetry. The threshold for somitogenesis has no effect (not shown here, identical phenotype to the top. Used parameters: normal period: 120 minutes, asymmetric period on the right: 150 minutes. Normal FGF decay: 0.005/minute, right-hand FGF decay: 0.002/minute; first threshold: 0.2 a.u.; second threshold: 0.05 a.u.; σ : 10; exponent: 2.5. Start of asymmetry after 6 formed somites, end after 4 more somites are formed on the left.

Figure 3.9. C) Tuning the decay rate with symmetric pErk dynamics does not yield symmetric somite formation. FGF decay rates on the right: 0.003 and 0.004. D) Tuning the decay rate with asymmetric pErk dynamics can lead to delayed but symmetric formation of somites, except for altered polarity in the region where the catching-up mechanism has acted. The different rows have different durations of the asymmetric regime (3, 4 or 5 cycles). FGF decay is 0.004. See also Supp. Fig. S3.4 for tuning with larger frequency differences.

asymmetric model results: asymmetric pErk When the pErk oscillation frequency is instead assumed to be asymmetric, a delay forms naturally by the slower progression of somitogenesis on the right, as seen before with the zebrafish model (compare Fig. 3.7C to Fig. 3.9B). Adding the first morphogen threshold for somite determination does not change this phenotype (not shown). This delay persists upon return to symmetric oscillations, and the lower morphogen threshold which forces somite formation, is still required to catch up and restore symmetry. As in zebrafish, we can tune the morphogen decay rate such that somite formation is fairly symmetric, but only up to the point when the parameters are restored to symmetric values. Then, either the somites become irregular, or the passing of the lower threshold causes cells to stop oscillating regardless of the pErk and Notch oscillator phases, which again makes somite boundaries and polarity unclear (Fig. 3.9D, Supp. Fig. S3.4, Video 5). Note that such tuning of FGF8 decay and oscillator frequency asymmetry is impossible in the simulations with symmetric pErk, where some somites are always somewhat larger or smaller than on the left. Furthermore this cannot be improved by an additional tuning of the onset of morphogen decay asymmetry with respect to Notch frequency asymmetry, like we did with the simple model (Fig. 3.9C).

conclusion Experiments are needed to determine which of the possible behaviours of pErk actually occurs. For now, the asymmetric option appears more likely, as it implicitly incorporates the feedback between the Notch and pErk oscillations that is also hinted at by experimental evidence (Hayashi *et al.*, 2009, Niwa *et al.*, 2011): slower Notch oscillations lead to slower pErk oscillations. Another advantage is that the delay between left and right arises naturally and does not require the skipping of a full round of somite formation, which can only be obtained by substantially increased FGF levels. In contrast to the zebrafish model, the mouse model does not allow the formation of fully symmetric catch-up somites. In zebrafish, the multiple Notch waves travelling in the PSM naturally allow formation of "catch-up" somites when the FGF wavefront shifts back to a normal position on the right. In mouse instead, normal somite patterning depends on both the Notch and pErk oscillations and this catch-up mechanism is absent. A secondary FGF threshold is needed to induce formation of catch-up somites in a manner distinct from normal somite formation.

Table 3.2. Outcome of different somitogenesis models for left-right asymmetry in different organisms

| organism | simple model | zebrafish model | 2 oscillator model |
|-----------|---|--|--|
| chick | correct phenotype | n.a. | n.a. |
| zebrafish | yes, but hard to obtain symmetric formation of all somites -> have to tweak onset of different asymmetries. | yes; can obtain symmetric somites, natural origin of delay | n.a. |
| mouse | yes, but have to hard-code the delay on the right | n.a. | yes; natural origin of delay have to assume presence of two thresholds for somitogenesis. Asymmetric pErk dynamics more likely |

3.2.3 Somitogenesis and left-right signalling interactions

In the models above, we imposed the asymmetric conditions of reduced oscillation frequency and an extended morphogen gradient. To improve our understanding of asymmetric somitogenesis, we will need to develop more detailed, mechanistic models for the interaction between the left-right signalling pathways and somitogenesis. Here, we make a start by characterising the left-right pathway using a dynamical systems perspective (François and Siggia, 2012, Verd *et al.*, 2014). We propose the following simple model: there is a regulatory sidedness module that (under normal conditions) endows mesoderm cells with either a left or a right sided identity. We assume that the module functions such that each identity locally enhances itself while repressing the opposite identity, (generating bistability), while at a distance each identity induces the other identity and/or represses its own identity (ensuring lateral inhibition) (Fig. 3.10A). The sidedness module can be interpreted as a symmetrical extrapolation of known Nodal-Lefty interactions, in which Nodal locally reinforces its own expression while, through Lefty, repressing its own expression at the other side of the midline and hence ensuring opposite identity there. We assume that left-right signalling occurs upstream of this sidedness module, biasing the module such that left-identity always arises on the left (Fig. 3.10B).

The sidedness module uses, and may therefore bias, signalling components involved in somitogenesis such as FGF, Wnt and Notch (Meyers and Martin, 1999, Nakaya *et al.*, 2005, Raya *et al.*, 2003). For instance, FGF8 is required to confer right-sided identity in chick, and left-sided identity in mouse (Meyers and Martin, 1999); Notch is required for left-sided identity, and is asymmetrically expressed on the left side of the node in chick embryos (Raya *et al.*, 2003). RA is required to buffer these effects, and it has been shown that a protein involved in RA signalling (Nr2f2) is biased to the left side in chick embryos, and to the right in mouse embryos (Vilhais-Neto *et al.*, 2010). It thus seems likely that the left-right pathway is also responsible for the biased expression of Nr2f2, thus taking care of both the bias and the buffering. If the RA buffering machinery were also downstream of the sidedness module, we would expect that somitogenesis remains symmetric even when the left-right signal upstream of the module is disturbed. However, it has been shown that in about 46% of embryos without this upstream signal (H^+/K^+ ATPase), somitogenesis is asymmetric, with no clear left-right bias (Kawakami *et al.*, 2005). From this, we infer that RA signalling is more directly downstream of the left-right signal, not the sidedness module (Fig. 3.10B).

Using this model, we can explain a loss of left-right asymmetry from a malfunctioning of the sidedness module, causing both sides to obtain a similar identity that is either all-left, all-right or none/mixed. If RA is absent, the biases in FGF, Wnt and Notch introduced by the module are not compensated and somitogenesis is asymmetric, always in the same direction. When the upstream left-right signal is disturbed, the sidedness

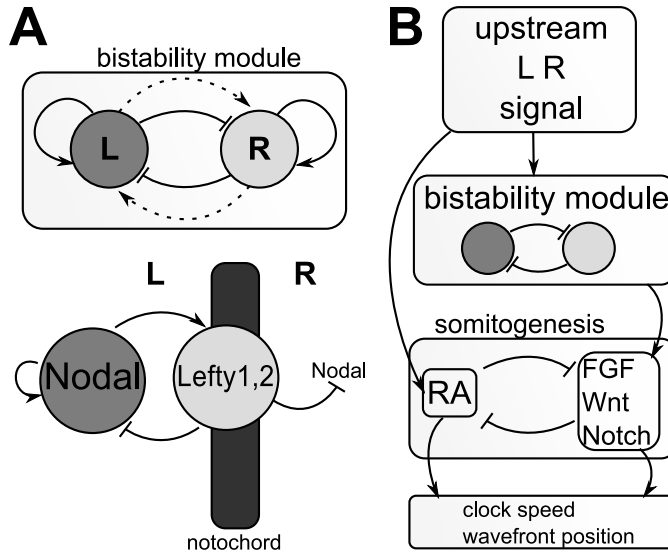


Figure 3.10. A simplified model for the interactions between the left-right pathway and somitogenesis. A) Our model includes a sidedness module which is able to generate asymmetry spontaneously. A typical example contains two components that exert a short-range inhibition on each other, but activate each other over a longer range. In the left-right pathway in particular, this module could be made up of Nodal and Lefty. Wherever Nodal is upregulated, it amplifies itself and upregulates Lefty genes both in the lateral plate mesoderm (LPM) and the notochord. Lefty then inhibits Nodal in the right LPM, thus conferring a right-handed identity through a lack of Nodal. B) Normally, the sidedness module receives input from an upstream left-right pathway, involving H^+/K^+ ATPase, nodal flow etc. This pathway upregulates Nodal on the left, thus forcing the balance of the module. We hypothesise that RA is regulated separately by this upstream pathway. The other signalling pathways of somitogenesis influence, and are influenced by, the sidedness module.

module confers random left-right identity and RA is also biased randomly, but independent from this module. By chance, the RA bias and the sidedness module can end up aligned so that RA can buffer the effects of the module: somitogenesis then remains symmetric. In other cases RA ends up on the wrong side, not buffering but exacerbating the random asymmetries caused by the independent action of the sidedness module. This biased asymmetry due to blocking RA signalling, and randomly oriented asymmetry with low penetrance due to blocking the left-right signalling pathway, agrees with experimental observations (Kawakami *et al.*, 2005). Furthermore, this model predicts that the asymmetric phenotype arising from a perturbed upstream signal is more severe than the asymmetric phenotype upon blocking of RA signalling.

While the proposed model thus can already explain some important aspects of asymmetric somitogenesis, it clearly represents just a starting point. To firmly tie this model to our models of somitogenesis, we need to have a more mechanistic understanding of how the left-right pathway influences oscillation frequency and anterior extent of the determination wavefront and how these effects are being buffered by RA. An interesting candidate for affecting both oscillator frequency and wavefront is sonic hedgehog (SHH), an important player in the left-right pathway that has been shown to influence FGF8 expression as well as influence somitogenesis clock speed via inhibition of Gli2/3 (Resende *et al.*, 2010). Since RA was also shown to inhibit Gli2/3 (Resende *et al.*, 2010), this suggests a means for RA to buffer the oscillation frequency impact of the left-right pathway. Similarly, since FGF8 and RA are mutually antagonistic, this would also offer a potential explanation for RA mediated buffering of the influence of the left-right pathway on wavefront position. However, the story is likely to be more complicated. First, it is clear that SHH and FGF play different roles in left-right signalling in different species (Meyers and Martin, 1999). Second, while in mouse and zebrafish both clock speed and wavefront are affected by the left-right pathway, these changes are likely to arise at least partly independently given the fact that in chicken only clock speed is affected by the left-right pathway (Vermot and Pourquié, 2005).

3.3 Discussion

A precise, reproducible and symmetric progression of somitogenesis is of crucial importance for vertebrate fitness, as evidenced by the severely disabling effects of conditions such as scoliosis. Still, although there is a rich tradition of developing models aimed at obtaining a better understanding of vertebrate somitogenesis, thus far models have neither been applied nor developed for understanding left-right asymmetry. In the current paper we undertook the first steps in developing such models. In addition, we investigated the relevance of somite determination, the mechanism by which oscillations start to cease and somite boundaries are pre-patterned, for explaining the different left-right asymmetry phenotypes observed for different vertebrate model species.

In the classical clock-and-wavefront model formulated by Cooke and Zeeman, oscillations fully terminate upon encountering the wavefront, simultaneously across all cells forming the future somite. Specification of somite polarity is not considered in this model. In later clock-and-wavefront models, oscillations cease in a cell by cell manner upon encountering the wavefront, and cells are assumed to memorise their oscillator phase. The memorised phase differences are assumed to specify both somite boundaries and within somite polarity. In some models this ceasing of oscillations and memorising of phase is simply superimposed (Jaeger and Goodwin, 2001, Morelli *et al.*, 2009), while in others the wavefront induces a transition from oscillatory to bistable behaviour (François

et al., 2007, ten Tusscher and Hogeweg, 2011), or arises emergently from a Hopf-Turing bifurcation (Cotterell *et al.*, 2015, Meinhardt, 1982) or as a result of coupling to non-oscillating cells (Murray *et al.*, 2011). In all these models, oscillations cease right at the onset of somite boundary determination.

In the current study we started out with a model in which the superimposed wavefront leads to oscillator stopping and phase memorisation (Morelli *et al.*, 2009). We also developed 2 new models. In the first, intended to mimic zebrafish somitogenesis, we incorporated the experimental observation that the early somite boundary marker pErk decreases its expression in discrete, somite wide jumps in a Notch oscillation dependent manner (Akiyama *et al.*, 2014). With this model we show how a progressive delay can arise in somitogenesis in the right PSM following a decrease in oscillation frequency and an increase in the anterior extent of the wavefront, and how symmetry can be restored once left-right signalling terminates. This model reproduces the experimental observation that within an individual pre-somite, oscillations halt in a posterior-to-anterior manner (Shih *et al.*, 2015); it also predicts such a P-A progression for removal of pErk (and Tbx6 (Wanglar *et al.*, 2014)) at the future somite boundaries, thus offering an explanation for the step-wise posterior shifts of the pErk domain boundary.

In the second model, representing mouse conditions, we incorporated the experimental observation that pErk displays oscillations which do not slow with distance from the tailbud, and that somite boundary determination occurs when pErk levels are low, again in a Notch oscillation dependent manner (Niwa *et al.*, 2011). Using this model, we show how the developing asymmetric phenotype depends on both the asymmetry in FGF levels as well as the assumed pErk dynamics in the delayed, right-hand PSM. If, similar to notch oscillations the pErk dynamics are also assumed to be slower in the right PSM, this model behaves similar to the zebrafish model, generating a progressive delay in right-hand somite formation. However, here this delay arises from slowing of pErk rather than Notch oscillations. A difference between the zebrafish and mouse model is that while in the zebrafish model "catch-up" somites are formed with normal polarity information, this is not the case in the mouse model, in agreement with experimental observations (Vermot *et al.*, 2005).

In the two new models, somite determination occurs without a memorisation of oscillator phase. Since waves of Notch signalling set the pace of somite formation, somite determination occurs in a posterior to anterior manner, thus potentially providing alternative somite polarity information. Indeed, recently a two somite periodicity was observed to result from oscillator slowing and was proposed to contribute to the formation of sharply delineated somite boundaries and anterior-posterior polarity (Shih *et al.*, 2015). Note that our second model resembles another two-oscillator somitogenesis model, that was

recently proposed to explain somite size scaling (Beaupeux and François, 2016). A notable difference between this model and the model we propose here is that rather than boundary determination arising from a particular size of the phase difference between the two oscillators, in our model determination occurs only if the two oscillators are simultaneously in a specific phase, as suggested by experimental data (Harima and Kageyama, 2013).

In vertebrates, the interactions between Wnt,FGF and RA determine the position of the determination front (Aulehla and Pourquié, 2010, Diez del Corral *et al.*, 2003), and RA is involved in maintaining somite symmetry. Still, RA is not necessary for somite formation to occur. In amphioxus, the model species representative of the cephalochordate sister group of the vertebrates, somites form close to the tailbud and somitogenesis is asymmetric (Schubert *et al.*, 2001). It was shown that FGF8 is not required for the formation of the posterior somites in amphioxus (Bertrand *et al.*, 2011), FGF and RA do not interact (Bertrand *et al.*, 2015) and RA is not able to generate symmetric somitogenesis. It thus appears that the FGF-RA antagonism evolved to ensure symmetric somitogenesis in vertebrates (Bertrand *et al.*, 2011, Brent, 2005). Another striking difference between amphioxus and vertebrate somite formation is that in amphioxus somite determination occurs relatively close to the tailbud, whereas in vertebrates there is a large PSM between the tailbud and the determination front. This extended PSM may have arisen as a side effect of the evolution of the FGF-RA antagonism. Alternatively, the extended PSM may be essential to allow sufficient time and space for buffering small asymmetries and have been directly selected for. Clearly, much remains to be discovered on the function of the extended PSM, the PSM spanning oscillator frequency gradient and the resulting travelling waves for somite determination and symmetry.

Our study shows how differences in somite determination dynamics between the different vertebrate species may contribute to their diverse asymmetric phenotypes. Thus, the asymmetric phenotype arising in absence of RA provide additional information that can be used to further decode the underlying developmental mechanism. Indeed, our results suggest that rather than focusing on a catch-all mechanism in all vertebrate species and assuming that species differences merely reflect neutral developmental systems drift, we should keep an open mind for the possibility of functionally significant species differences.

Acknowledgements

We thank Paulien Hogeweg and Enrico Sandro Colizzi for discussion and comments on the manuscript.

3.4 Methods

3.4.1 Clock and wavefront model

We model the presomitic mesoderm (PSM) as a 2D strip of cells. In the posterior the cells form a single coherent tissue representing the tailbud, more anteriorly the cells form two strips of tissue flanking the notochord (Fig. 3.2A). Each individual cell is endowed with an internal oscillation clock that is represented by a simple sinusoidal phase oscillator, as described in (Ares *et al.*, 2012, Jaeger and Goodwin, 2001, Morelli *et al.*, 2009, Murray *et al.*, 2011) (Fig. 3.2A). At the tissue level a frequency profile is assumed to dictate oscillation frequency as a function of position in the PSM. In contrast to previous work, we do not investigate the interaction between intrinsic oscillations and cell-cell coupling, nor do we apply noise (Ares *et al.*, 2012, Herrgen *et al.*, 2010, Morelli *et al.*, 2009, Murray *et al.*, 2011); we therefore do not include cell-cell coupling of oscillations for simplicity. The angular frequency of the oscillator in each cell thus only depends on its position in the PSM, with the following relation:

$$\omega(x) = \omega_0 * \left(1 - \frac{1}{\sigma^n} * x^n\right) \quad (3.1)$$

where $\omega(x)$ is the angular frequency at a certain distance x away from the posterior end of the PSM (Fig. 3.2B). ω_0 is the oscillation frequency of cells at the posterior end of the PSM, and σ is the length over which the frequency will drop to 0. Usually σ is taken to be the PSM length, unless otherwise indicated. Finally, n is the exponent that determines the nonlinearity of the frequency profile: the higher the exponent, the further anterior in the PSM the frequency will start decreasing and the steeper the slope will be. When cells become incorporated into a (pre)somite at the anterior end of the PSM, they stop oscillating while memorising their phase.

Cells are continuously added at the posterior end of the PSM, and the oscillators of these new cells are assumed to obtain the phase and frequency of the cells already present there (Fig. 3.2A). The anterior wavefront of somite determination travels toward the posterior at the same speed as cells are added, so that the PSM maintains a constant size (Morelli *et al.*, 2009). The frequency profile shifts along, so that cells experience a progressively lower oscillation frequency, until the wavefront passes and their phase becomes frozen (Fig. 3.2B). We adapt this model of somite formation as we go on to account for differences between animals in the next section.

left-right differences When we implement left and right differences, we change the frequency ω_0 , and/or the extent of the frequency profile σ differently in the left and right PSM, which results in different behaviour for the left and the right somites (Fig. 3.2C). Often, the left side is kept the same as the starting conditions, for reference. We always

start the simulations with the same ω_0 and σ on the left and the right, then “switch on” the difference after a few somites have been formed. The asymmetry is then maintained until the end of the simulation, unless otherwise indicated.

3.4.2 Extended models

As a first step in extending the above described model, we incorporate an explicit description of FGF/Wnt (morphogen) decay driven gradient dynamics (Aulehla and Pourquié, 2010) and how this subsequently influences the frequency profile. Individual model cells i contain a specific level of morphogen M . This level is set to a constant value of 1. in the tailbud and slowly decays in all other cells in the PSM, yielding an exponential posterior-to-anterior gradient:

$$\frac{dM_i}{dt} = -decay * M_i \quad (3.2)$$

We do not include diffusion of morphogen in our model.

Next, we describe how the frequency gradient f depends on the morphogen level using the following relation:

$$f = f_0 * \frac{M^2}{M^2 + 1/\beta^2} \quad (3.3)$$

where β determines the morphogen concentration at which the frequency has decreased to half its value in the tailbud. The combination of the exponential morphogen gradient with the above non-linear function for the oscillator frequency, generates similarly shaped frequency profiles as the one used in our earlier simpler model (figure 3.6A).

We implement the observed asymmetry in anterior extent of the FGF gradient by adjusting FGF/Wnt decay rates. We assume that the decrease in oscillation frequency in the right PSM arises independent of FGF/Wnt and model it through a decrease in the maximum of the frequency profile. The rationale for this assumption is that an increase in FGF on the right as observed in zebrafish and mouse would be expected to have a speeding up, rather than slowing down, effect, which contradicts the asymmetric phenotype. Furthermore, that other factors besides FGF/Wnt are known to affect oscillation frequency more strongly (Resende *et al.*, 2010).

Zebrafish model

We extended our model to explicitly model zebrafish somite patterning dynamics based on experimental data (Akiyama *et al.*, 2014, Wanglar *et al.*, 2014). These experimental data suggest that cells transition through a sequence of discrete states before transforming into a fully determined somite. Oscillating cells start out in a “pErk high,

Tbx6 high” state in the posterior PSM. As FGF levels drop beyond a certain level a first wave of Her expression causes cells to transition to a “pErk low, Tbx6 high” state, pre-patterning the S-IV-S-V somite boundary. A next wave of Her expression causes cells to transition to a “pErk low, Tbx6 low”. Subsequent waves of Her expression are likely to induce further, not yet characterised, transitions in gene expression demarcating further differentiation until the S0 stage is reached and a fully determined somite forms. It is only in this final stage that oscillations are observed to cease (Shih *et al.*, 2015).

In the model, FGF dropping below a certain threshold value can easily be accommodated with the incorporation of explicit FGF (morphogen) dynamics. We simulate the passing of a wave of Her expression as the passing of a particular oscillator phase ($\sin(\omega_t) > 0.95$), and termination of this wave is simulated with a contrasting oscillator phase ($\sin(\omega_t) < -0.95$). For simplicity, we only model the transition from the “pErk high, Tbx6 high” state to the “pErk low, Tbx6 high” state and subsequently to the “pErk low, Tbx6 low”, ignoring additional transitions to less well known states. In our model, oscillations cease and full somite determination occurs upon transition cells from the “pErk low, Tbx6 high” state to the “pErk low, Tbx6 low” state and termination of the Her expression wave ($\sin(\omega_t) < -0.95$). Note that by skipping the less well known transitions in our model, oscillator termination and somite determination occur 3 clock cycles earlier than in reality. Still, the model at least captures the early boundary formation that also occurs in reality, as cells transition from the “pErk high, Tbx6 high” to the “pErk low, Tbx6 high” state.

Mouse model

Our formulation of a two-oscillator model for mouse somitogenesis is based on experimental observations from Niwa *et al.* and Harima and Kageyama (Harima and Kageyama, 2013, Niwa *et al.*, 2011). We incorporate a “global oscillator” which oscillates with the same phase and frequency across the entire PSM, as an approximation of the pErk dynamics observed in mouse. In the tailbud region, we assume that the global oscillator cycles at the same frequency as the travelling wave oscillator, while setting it a quarter phase (-0.5π) behind to allow the first wave to travel for a full oscillation cycle. Experimental observations indicate that somite boundary patterning occurs when pErk levels are low and Notch levels high. In more abstract terms this means that the two oscillators should each be in a specific phase and this should occur simultaneously. We simulate this in our model by imposing that when the global oscillator is in the “low” ($\sin(\omega_t) < -0.9$) part of its cycle, the region where the travelling wave oscillator is high ($\sin(\omega_t) > 0.9$) – and all the tissue anterior to this region – stops oscillating and are transformed into a somite.

In addition to the above requirements for somite patterning, we can incorporate a threshold FGF (morphogen) level above which somites cannot form despite pErk levels being low and Notch levels being high. This additional threshold effectively delays somite formation and can lead to the formation of multiple travelling waves in the PSM. On top of this we can add a second, lower FGF threshold which works in the opposite direction: When the level of FGF in cells drops below this level, somite patterning ensues regardless of the phase of the global and travelling wave oscillators.

Table 3.3. Parameter values

| parameter | values | remarks |
|---|-------------------------|---|
| clock-and-wavefront model | | |
| chick ω_0 | 0.070 min ⁻¹ | maximum oscillation frequency in tailbud |
| mouse ω_0 | 0.052 min ⁻¹ | |
| zebrafish ω_0 | 0.21 min ⁻¹ | |
| chick n | 2 | exponent of quadratic equation governing frequency profile |
| mouse n | 3 | |
| zebrafish n | 2 | |
| σ | 60 cells | length of frequency profile |
| zebrafish model | | |
| normal ω_0 | 0.21 min ⁻¹ | asymmetric frequency values indicated in figure legends |
| normal FGF decay | 0.005 min ⁻¹ | |
| FGF threshold for somite formation | 0.15 a.u. | unless otherwise indicated |
| β | 10 | determines at which FGF concentration the frequency decreases |
| two-oscillator model | | |
| mouse ω_0 | 0.052 min ⁻¹ | |
| chick ω_0 | 0.070 min ⁻¹ | |
| mouse n | 2.5 | |
| chick n | 2 | |
| normal FGF decay | 0.005 min ⁻¹ | |
| β | 10 | determines at which FGF concentration the frequency decreases |
| FGF inhibitory threshold for somite formation | 0.2 a.u. | |
| FGF forcing threshold for somite formation | 0.05 or 0.075 a.u. | Slightly higher threshold required for chick simulations |

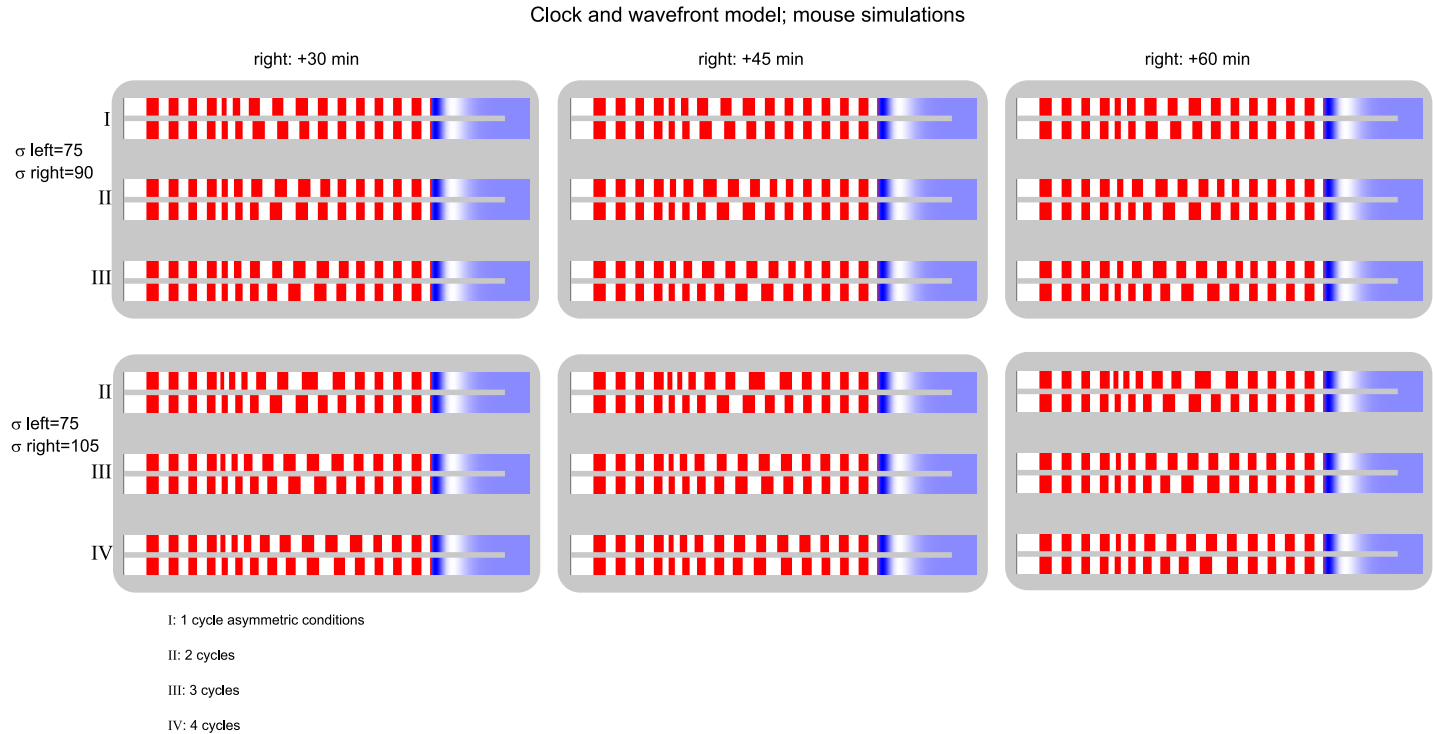


Figure S3.1. The clock and wavefront model for mouse asymmetric somitogenesis. We varied the anterior shift of the wavefront on the right side, the reduction in frequency, and the number of cycles for which the asymmetric conditions last (including the period during which the wavefront shifts).

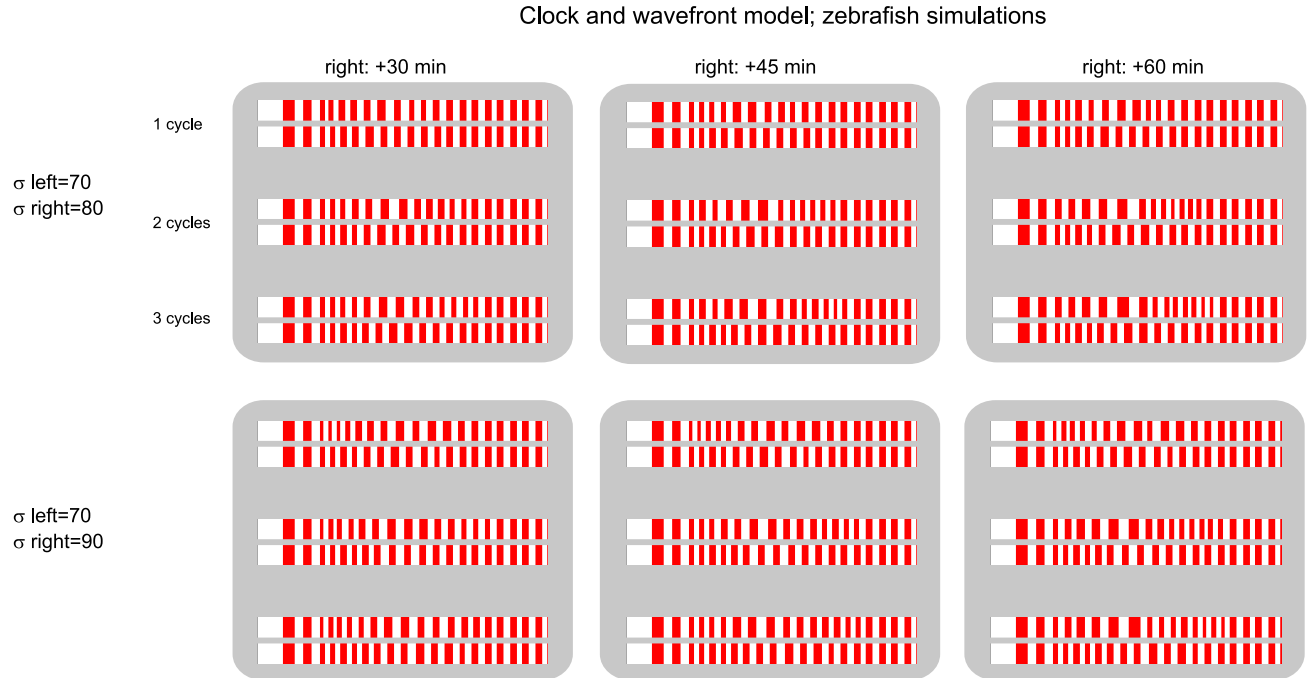


Figure S3.2. The clock and wavefront model for zebrafish asymmetric somitogenesis. We varied the anterior shift of the wavefront on the right side, the reduction in frequency, and the number of cycles for which the asymmetric conditions last (including the period during which the wavefront shifts).

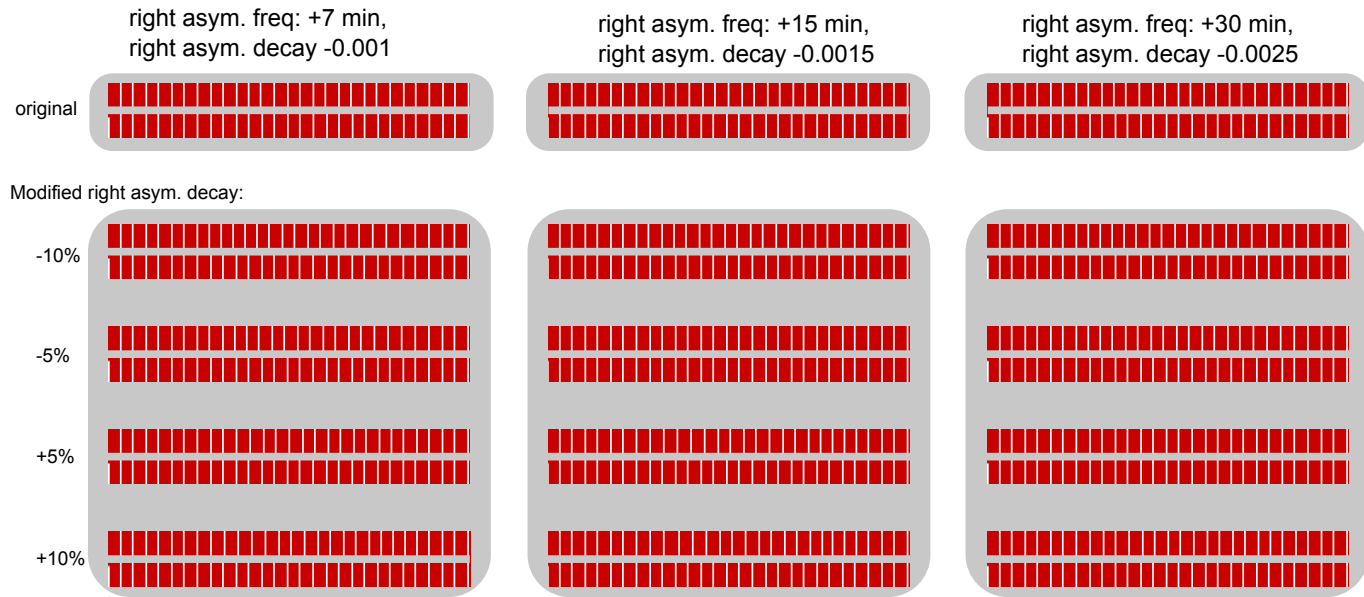


Figure S3.3. The extended model for zebrafish asymmetric somitogenesis is robust to small changes in the decay rate in the right PSM. During asymmetric conditions, there are combinations of lower right-hand frequency and FGF decay rate which lead to symmetrically formed somites. We varied this “perfect” asymmetric decay rate to test the robustness of the symmetric somites. Here we see that often, a 10% deviation still yields relatively symmetric somites. Occasionally, the “modified” decay rate is actually more symmetric.

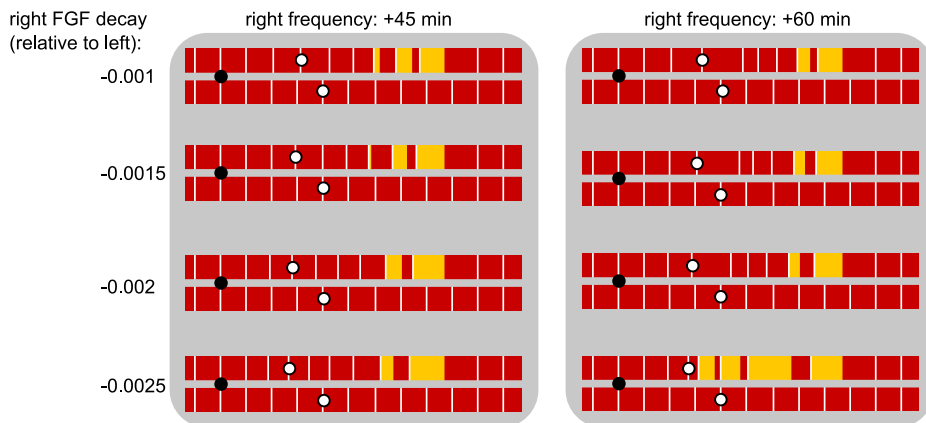


Figure S3.4. The extended mouse model has difficulty obtaining fully symmetrically formed, delayed somites. Here we show asymmetric somitogenesis with different asymmetric values for the right-hand FGF decay rate, relative to the left value (0.005/min). Left oscillation period is 120 min. Black denotes the start of asymmetric conditions, the white dot the end. Asymmetric conditions start after 6 cycles and end after another 4.

3.5 Movies

Movies

Video 1: Simulation of asymmetric chick somitogenesis with basic model. After 5 oscillation cycles, the position of the determination front is shifted anteriorly by 2 somites on both sides. The oscillation period on the right is increased from 90 min. to 105 min. After 6 additional cycles (of 90 min.), the frequency and determination front position are shifted back to normal values. Anterior to the left.

Video 2: Simulation of asymmetric mouse somitogenesis with basic model. After 4 oscillation cycles, oscillation period on the right is increased from 120 min. to 150 min. After another cycle (of 120 min.) the position of the determination front is shifted anteriorly by two somites on the right, and one somite on the left. When the wavefront shift is complete, the parameters return to symmetric values (after two cycles).

Video 3: Simulation of asymmetric zebrafish somitogenesis with extended model. After 20 oscillation cycles, the oscillation period on the right is increased from 30 min. to 45 min and the morphogen decay is decreased from $0.005 \text{ a.u. min}^{-1}$ to $0.003325 \text{ a.u. min}^{-1}$. After 5 additional cycles these parameters are restored to normal values.

Video 4: Simulation of asymmetric mouse somitogenesis with extended model; symmetric pErk. After the formation of 1 small and 6 normal somites (10 cycles from the start of the simulation), the oscillation period on the right is increased from 120 min. to 180 min., and the morphogen decay is decreased from 0.005 to 0.002. After 4 cycles, these parameters are restored to normal values. Other parameters as in figure 3.9A, bottom.

Video 5: Simulation of asymmetric mouse somitogenesis with extended model; asymmetric pErk. After the formation of 1 small and 6 normal somites (10 cycles from the start of the simulation), the oscillation period on the right is increased from 120 min. to 150 min., and the morphogen decay is decreased from 0.005 to 0.004. After 5 cycles, these parameters are restored to normal values. Other parameters as in figure 3.9D, bottom.

Part II

Clocks, stripes and shapes

4

In silico evo-devo: reconstructing stages in the evolution of animal segmentation

RENSKE M.A. VROOMANS, PAULIEN HOGEWEG AND KIRSTEN H.W.J. TEN
TUSSCHER (2016)

EvoDevo, 7:14

Abstract

The evolution of animal segmentation is a major research focus within the field of evolutionary-developmental biology. Most studied segmented animals generate their segments in a repetitive, anterior-to-posterior fashion coordinated with the extension of the body axis from a posterior growth zone. In the current study we ask which selection pressures and ordering of evolutionary events may have contributed to the evolution of this specific segmentation mode. To answer this question we extend a previous *in-silico* simulation model of the evolution of segmentation by allowing the tissue growth pattern to freely evolve. We then determine the likelihood of evolving oscillatory sequential segmentation combined with posterior growth under various conditions, such as the presence or absence of a posterior morphogen gradient or selection for determinate growth. We find that posterior growth with sequential segmentation is the predominant outcome of our simulations only if a posterior morphogen gradient is assumed to have already evolved and selection for determinate growth occurs secondarily. Otherwise, an alternative segmentation mechanism dominates, in which divisions occur in large bursts through the entire tissue and all segments are created simultaneously. Our study suggests that the ancestry of a posterior signalling centre has played an important role in the evolution of sequential segmentation. In addition, it suggests that determinate growth evolved secondarily, after the evolution of posterior growth. More generally, we demonstrate the potential of evo-devo simulation models that allow us to vary conditions as well as the onset of selection pressures to infer a likely order of evolutionary innovations.

4.1 Introduction

Segmentation, the division of the animal body plan into multiple, repeating units, has fascinated evolutionary and developmental biologists alike. Only the vertebrates, arthropods and annelids display overt body segmentation while several other clades show intermediate levels of segmentation in only a subset of tissues or organs, a property called metamerism (Balavoine and Adoutte, 2003, Davis and Patel, 1999, Peel and Akam, 2003). Repetitive patterning is studied in most detail in overtly segmented animals. In these clades, segments are typically laid down in a regular anterior-posterior sequence, via a process involving posterior growth (also called terminal addition) and periodic, sequential generation of segments (Couso, 2009, Peel and Akam, 2003). A famous exception is the fruitfly *Drosophila* in which segments are laid down simultaneously across a preformed body axis.

It is currently unresolved why segmented animals mostly display this superficially similar, sequential mode of segmentation. This issue is partly related to the question of whether segmentation was present in the bilaterian ancestor, either as overt segmentation or as metamerism, or rather that it evolved multiple times in parallel in the different lineages (Balavoine and Adoutte, 2003, Blair, 2008, Budd, 2001, Chipman, 2010, Couso, 2009, Davis and Patel, 1999, Jacobs *et al.*, 2005, Minelli and Fusco, 2004, Richmond and Oates, 2012, Seaver, 2003, Tautz, 2004, Wanninger *et al.*, 2009). Arguments in favour of a single origin of segmentation include the prevalence of sequential segmentation (Couso, 2009, Peel and Akam, 2003). Studies using ancestral state reconstruction suggest that this mode of segment addition via posterior outgrowth represents ancestral bilaterian properties (Gold *et al.*, 2015, Jacobs *et al.*, 2005). In addition, the three segmented lineages have shared genes involved in segmentation, such as Notch, Engrailed, and Wnt (Couso, 2009, Rivera and Weisblat, 2009, Tautz, 2004, Williams *et al.*, 2012). Arguments in favour of parallel evolution of segmentation instead put forward that there are also large differences in the genes responsible for segmentation and that the limited observed overlap in gene usage could be explained by parallel recruitment from the limited developmental genetic toolbox (Chipman, 2010). The precise mechanisms of cell division, axial elongation and sequential segmentation also differ substantially between these lineages, ranging from teloblastic growth and stereotyped cell divisions in annelids and some crustaceans (Balavoine, 2014, Shankland and Seaver, 2000), to posterior growth zones in most insects and vertebrates (Bénazéraf and Pourquié, 2013, Peel, 2004) with variable roles of cell division versus cell rearrangement (Mayer *et al.*, 2010, Nakamoto *et al.*, 2015). Furthermore, multiple segmentation processes can take place in different body regions or tissue types even within a single organism, each with their own evolutionary origin (Graham *et al.*, 2014), which further supports (partial) parallel evolution.

Still, independent of whether sequential segmentation evolved once or multiple times, we can ask whether certain factors or conditions may have contributed to this particular evolutionary outcome. Earlier evo-devo simulation studies have demonstrated that sequential segmentation represents a robust evolutionary outcome with high future evolutionary potential (François *et al.*, 2007, Fujimoto *et al.*, 2008, ten Tusscher and Hogeweg, 2011), suggesting evolutionary advantages of this particular segmentation mode. In addition, prior evolutionary events may have generated biases or constraints that influenced the likelihood of the evolution of sequential segmentation. For example, evolutionary comparisons show that a posterior signalling region characterised by caudal, Wnt and FGF signalling predates the origin of the bilaterians (Niehrs, 2010). Furthermore, it was recently suggested that posterior growth through terminal addition was already present in the bilaterian ancestor (Gold *et al.*, 2015). Thus, we may ask whether these properties have played a decisive role in sending evolution down the path of evolving sequential segmentation.

In the current study we aim to answer these questions. For this we substantially extended a previously used *in silico* model (ten Tusscher and Hogeweg, 2011). Rather than superimposing a particular growth pattern, we incorporate a gene controlling cell division and let evolution determine the type of tissue growth dynamics that arises. Then, by varying whether or not a stable posterior signalling centre is present in simulations, we can investigate the role of such a signalling centre on the type of growth and segmentation that evolves. We thus substantially expanded the degrees of freedom available to the evolutionary process, allowing us to investigate under which conditions sequential segmentation is the most likely evolutionary outcome.

We observe two predominant evolutionary outcomes: sequential segmentation with posterior growth, and simultaneous segmentation involving tissue-wide bursts of divisions. We find that the likelihood with which the strategies evolve, depends on the type of imposed morphogen dynamics and the strength and timing of an evolutionary pressure for determinate growth. We show that a self-organised posterior signal is more difficult to evolve than a developmental strategy which does not rely on such a posterior centre. From this we conclude that the prior evolution of a posterior signalling centre has played a decisive role in determining the evolution of sequential segmentation. Furthermore, we demonstrate that an evolutionary pressure for determinate growth reduces the likelihood of evolving sequential segmentation. When we apply this evolutionary pressure after sequential growth and segmentation have evolved, a mechanism to stop growing can evolve which is coordinated with the pre-existing sequential segmentation. We therefore propose that the order of evolutionary events is key to inferring the likelihood of particular developmental strategies. Reversing the argument, our work strongly suggests that a posterior signalling zone evolved prior to segmentation, and that sequential growth and segmentation evolved prior to determinate growth.

4.2 Methods

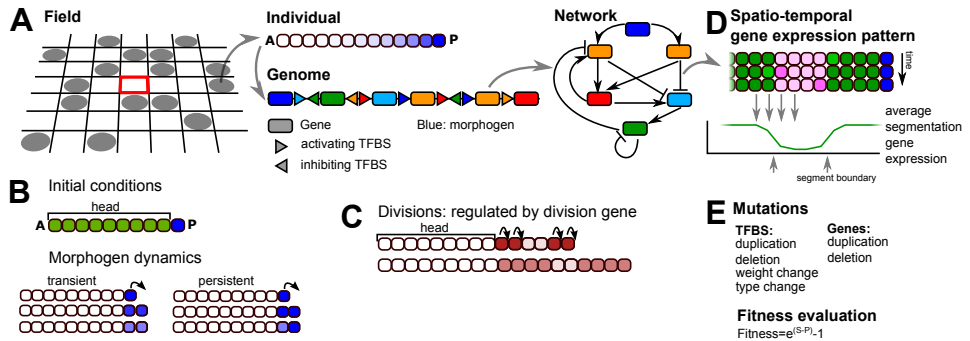


Figure 4.1. Overview of the model. A) The developing individuals live on a 2D lattice. Each individual consists of a row of cells. The genome of the individual codes for a network of regulatory interactions, which determines the spatio-temporal dynamics of the proteins within each cell (see D). B) The initial conditions for each new individual at the start of its development. There is a growth zone with high morphogen, and a 'head' region without morphogen. The morphogen dynamics may vary. Either they are imposed, yielding persistent posterior morphogen (left): the morphogen is kept at a high level in the posterior-most cell while decaying in the other cells; or the morphogen can become regulated by the network, so that only the initial conditions are specified (right). C) Divisions are regulated by a division protein; when its level passes a threshold, the cell can divide. Upon division, the level of the division protein in both daughter cells is halved, but not the level of the other proteins. D) At the end of development, the expression of the segmentation gene is averaged over a number of time steps, and from this the segment boundaries are determined. E) The mutational operators acting on the genome.

4.2.1 The model

General setup

We extend an existing individual based model of a population of organisms evolving on a lattice (ten Tusscher and Hogeweg, 2011) (Fig. 4.1A). Each organism possesses a so-called “pearls-on-a-string” genome consisting of genes encoding transcription factors (TFs) and upstream regulatory regions with transcription factor binding sites (TFBS) (Crombach and Hogeweg, 2008). At birth, organisms consist of a short one-dimensional row of cells which grows through the course of the individual’s development. An individual’s probability of reproduction (fitness) depends on the number of segments present in its gene expression pattern after a predefined amount of developmental time.

Individuals

Genome, network The genome codes for a regulatory network with the genes representing the nodes, and the TFBS as the regulatory links (edges) between nodes (Fig. 4.1A). These regulatory interactions can be repressive or activating. This network governs gene expression dynamics and hence protein levels. Gene expression dynamics are modelled with ordinary differential equations as shown in eq. (4.1):

$$\begin{aligned} dG_i/dt &= \frac{input^2}{input^2 + 1} * E_i - d_i * G_i \\ input &= \text{MAX}(0., \sum_j \frac{w_j * G_j^2}{G_j^2 + H_j^2}) \end{aligned} \quad (4.1)$$

Transcription of a gene is determined by the summed input of all activating and repressing TFBS regulating this gene, where the influence of each individual TFBS is assumed to depend on TF concentration in a saturating manner. E_i is the maximum expression level of gene i and d_i is the degradation rate of the resulting protein; both values can evolve. w_j is the weight determining the strength with which TF j influences the expression of gene i ; this weight is negative (-1) for repressing TF and positive (+1) for activating TF, the sign of these weights is subject to evolution. H_j represent the evolvable Hill constants of the TFBS, where the Hill constant corresponds to the level of the TF at which half-maximal activation or repression occurs. The expression of all genes of the same type (see below) are summed into a single protein level.

Developmental tool kit and initial conditions There are 16 types of genes, indicated with a number from 0 to 15.

Gene 0 is the morphogen: unless indicated differently, it is not regulated by any of the other genes, thus corresponding to a maternal input. It is kept at a high expression level in the posterior-most cell, while decaying with a predefined rate in the rest of the embryo (Fig. 4.1B). In a subset of simulations instead, this high posterior expression is only used as an initial condition and is thus not automatically maintained in the posterior-most cell, and the gene may become regulated by other genes.

Gene 1 and gene 2 are signalling genes, responsible for direct, membrane bound cell-cell signalling (similar to e.g. Delta-Notch signalling). This direct cell-cell signalling is implemented as follows: if a gene has TFBS of type 1 or 2 in its upstream region, the expression of that gene in a particular cell is regulated by the levels of protein type 1 or 2 in its directly neighbouring cells, while its own intracellular levels of these proteins have no impact on the expression of that gene but only on that of genes in neighbouring cells. If cell-cell signalling is switched off in a simulation, gene 1 and 2 function as normal genes.

Gene 4 is the division gene: when it is highly expressed (protein level > 80. a.u.), the

cell may divide with high probability ($p=0.975$). Upon division, the level of only the division gene is halved in the resulting two daughter cells (Fig. 4.1C).

Gene 5 is the segmentation gene, whose final pattern of gene expression along the body axis determines the fitness of the organism.

Individuals start their development with a short row of 10 cells, where the posterior cell forms the primordial “growth zone” in which the morphogen level is high; in the remaining 9 cells (the “head” that is assumed to have evolved prior), the morphogen is absent (Fig. 4.1B). At the start of development genes 6 and 7 are uniformly expressed in the zygote; while other genes have an initial expression level of 0. Throughout development, the protein levels are updated according to the network (Eq. 4.1).

Fitness evaluation At the end of development (after a fixed number of time steps), the number of well-formed segments determines an individual’s fitness. A segment is defined by the average expression pattern of the segmentation gene over the last 20 or 40 developmental steps (Fig. 4.1D). This averaging helps ensure the evolution of temporally stable segmental patterning, as it will not reward oscillatory segmentation that fails to converge on a constant spatial pattern (as occurred in (Fujimoto *et al.*, 2008)). Segments should be at least 7 cells wide, and boundaries between segments should consist of a clear transition of the expression of the segmentation gene from a high to a low level, or vice versa, within 5 cells (similar to earlier definitions (François *et al.*, 2007, ten Tusscher and Hogeweg, 2011)). The number of too-narrow segments is subtracted from the number of well-formed ones, reducing the fitness. To further ensure stability of the final developmental pattern we apply an additional fitness penalty for the amount of variance of the pattern from the average (pattern instability) within the final 20 developmental steps.

In a subset of simulations, some fitness can also be obtained by reaching a target tissue size. This fitness bonus is independent from the number of segments, enabling sequential as well as simultaneous evolution of tissue size and segmentation. We also apply some penalties unrelated to the segments. First, we require that at least one gene of each type is present in the genome; if this requirement is not met, the individual is not allowed to reproduce. Second, a penalty is applied when the individual grows larger than the target final tissue size. Finally, small fitness penalties are used for gene and TFBS numbers in order to prevent excessive genome growth. The fitness then becomes $e^{\max(0,F)} - 1$.

$$\begin{aligned}
 F = & \text{nr good segments} \\
 & - \text{nr narrow segments} \\
 & + \text{proximity to target size} \\
 & - G * \text{gene nr} \\
 & - T * \text{TFBS nr} \\
 & - U * \text{nr unstable cells}
 \end{aligned} \tag{4.2}$$

See table 4.2 for all parameter values.

Evolution

Initial conditions, mutations and simulations The population is initialised with 50 genetically identical individuals. The population resides in a grid of size 30x30, imposing a maximum to the population size of 900 individuals. The genome of the initial individuals contains a single copy of each gene, in randomised order and with an average of 2 TFBS of random type upstream. Individuals compete for reproduction into a neighbouring empty spot. Those with a higher fitness have a larger probability of being selected. Specifically, an individual's chance to reproduce is proportional to its fitness divided by the sum over the fitness of itself and the other individuals neighbouring the empty position. Death occurs with a constant probability. Upon reproduction, the genome is mutated via duplications and deletions of both genes and TFBS (Fig. 4.1D). TFBS may also mutate their type (which protein binds), weight (activating or repressing) and Hill constant, and new TFBS may appear *de novo* as an innovation. Genes may mutate their maximum activation level E and decay rate d . Gene duplication also copies the associated TFBS, and results in multiple genes of the same type. The expression of all genes of the same type therefore contributes to the expression level of a single protein. Note that since there are no mutations that change the gene type, gene duplication can not be followed by subsequent divergence in our model.

Analysis

For each set of model settings and parameter values we run 50 simulations. Each simulation yields one particular growth and segmentation strategy with only minor variations within the population. Therefore, we only assess one fit individual per simulation. We consider a simulation successful when the fittest individuals at the end of the simulation can generate more than 3 segments.

space-time plots We use space-time plots as a first impression of the developmental mechanisms that evolved. Because we simulate 1D tissues, we can simply place snapshots of the tissue at many consecutive time points below each other, while keeping the position of the head fixed at the anterior. We display two types of space-time plots; in one, we denote the cell type of each cell with a colour, which represents a unique combination of gene expression values; in the other, cell divisions are depicted by making the newest cells white (in a division, we always consider the posterior daughter cell as the newest, so the anterior daughter stays black). See for example Fig. 4.2.

pruning Because an evolved genome consists partly of redundant interactions, we prune the genomes via a repeated process of removing genes and binding sites in the genome, while keeping the tissue size and final spatial expression pattern of the segmentation gene the same (ten Tusscher and Hogeweg, 2011). This makes it easier to analyse the network by eye and identify the role of the special genes such as the division gene and the segmentation gene. All networks depicted are pruned.

assessment of phenotypic variability We have only one source of noise in our simulations, namely the small probability that a cell with high division gene expression does not immediately divide. Still, this noise may cause some variability in the number of segments that are formed, even if the genome remains exactly the same. We call this phenotypic variability, and assess this by repeating the development of an evolved individual 50 times and counting the number of good and malformed segments formed each time. We display the results of this repeated development in a histogram, where we indicate the number of malformed segments with lighter bars. When an individual evolved a mechanism that makes the same number of good segments in 40 or more times out of 50 repeated developments, we call it “developmentally robust”. Note that this differs from the probability of breeding true (mutational robustness): we do not assess how mutations influence the likelihood of producing the same number of segments, as done in (Salazar-Ciudad *et al.*, 2001a).

In table 4.1 we display the number of times certain segmentation strategies evolve, and the average number of segments made per strategy. We take the phenotypic variability of an individual into account by averaging number of segments it makes during the repeated development mentioned above. These average numbers for single simulations are then averaged over all simulations in which the same developmental strategy evolved.

division gene dynamics In some figures, we display the dynamics of the division gene in the posterior-most cell (containing a high morphogen level). These dynamics result from regulation of the division gene by other genes, and from halving of the division gene upon division. In order to see the effect of regulation more clearly, we also show the network dynamics when the halving of the division gene is left out.

Modelling choices

Body axis segmentation is, like most developmental patterning, a complex phenomenon involving processes ranging from the subcellular to the organism level scale. In this study we simulate the evolution of body axis segmentation and therefore need to simulate development in a population of individuals across many generations. To keep our model tractable both in terms of simulation time and for analysing the evolutionary trajectories of the developmental processes, we substantially simplify the developmental

processes in our model relative to real world developmental processes, but incorporate those properties we deem necessary for studying the evolution of segmentation. Below we detail the three major simplifications, their potential consequences, and why we think these simplifications are justified.

The most obvious simplification is the 1D nature of the model, so that cell divisions automatically lead to an elongated body axis regardless of where they occur in the tissue - basically assuming that the axiality is already defined. In reality, developmental patterning occurs in a 2D or 3D tissue, where complex symmetry breaking events during early development are essential for setting up the anterior-posterior and dorsal-ventral axis. Furthermore, animal axis elongation often involves cellular motility and adhesion properties that are not included in this model. This limits the self-organisation potential of the developmental processes evolving in our model. However, since symmetry breaking is an ancestral property preceding bilaterian evolution (Niehrs, 2010), we can safely assume that it already existed before segmentation evolved.

We model a cellularised environment in which morphogen gradients are set up through decay dynamics, and signalling is limited to direct receptor-ligand type cell-cell signalling. Neither the morphogen nor any of the other gene products are subject to diffusion. Diffusion plays an important role in Turing-type patterning (Meinhardt, 1982), and in setting up the morphogen gradients dictating early *Drosophila* segmentation (Pankratz and Jäckle, 1990). Still, in *Drosophila* genes downstream of the morphogen gradients do not require diffusion, and so far no Turing patterns have been found to underlie animal segmentation. Segmentation usually takes place in a cellularised environment, in which the role of diffusion is necessarily restricted to short distances or combined with other gradient establishing mechanisms such as slow mRNA and protein decay (Aulehla and Pourquié, 2010). Thus, we are confident that we do not exclude any major real world segmentation mechanisms from evolving in our model, and that leaving out diffusion is justified.

Finally, our model only incorporates gene expression regulation through combinatorial TF regulation on a single regulatory region, ignoring several other factors that may influence gene expression. In vertebrates for instance, the timing of Hox gene mediated axial patterning, and its coordination with segmentation, involves chromosome looping, epigenetic histone and DNA modifications, and cluster level gene regulation (Mallo and Alonso, 2013). Still, the goal of our study is to explore the evolution of gene expression dynamics, rather than to mimic how these dynamics are precisely regulated. The transcription factor based regulatory network has sufficient degrees of freedom to allow the evolution of the diverse gene expression dynamics (such as oscillations) that underlie real world segmentation processes, while supporting computational efficiency and analytical tractability.

Glossary

model input:

transient posterior signal: At the start of development, the morphogen gene (0) is present in the posterior-most cell; there it decays, as in its daughter cells inheriting morphogen through divisions. See figure 1B.

persistent posterior morphogen: the persistent presence of the morphogen is superimposed in the posterior-most cell where it is neither subject to regulation nor decay; it decays in non-posterior cells inheriting the morphogen through division.

posterior signalling centre: persistent morphogen signal emanating from the posterior-most cells; in our simulations this may arise from a persistent posterior signal or from a transient posterior signal if regulatory interactions have evolved leading to a persistent signal demarcating the posterior end of the tissue.

simulation results:

tissue-wide growth: divisions occur throughout the tissue.

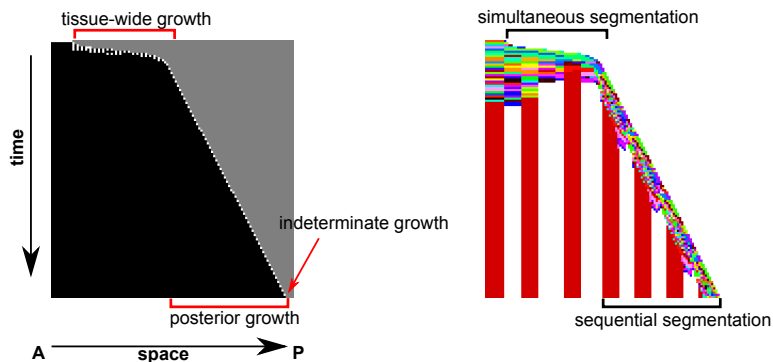
polar growth: divisions only occur in cells close to the anterior or posterior end of the tissue, in which case we also speak of anterior or posterior growth, respectively.

indeterminate growth: When divisions are not arrested by the end of the predefined development period, but could (in principle) continue indefinitely.

determinate growth: When divisions cease before the end of development, yielding a fixed tissue size.

simultaneous segmentation: multiple segments are formed at (roughly) the same time.

sequential segmentation: segments are formed successively in an anterior-to-posterior progression.



4.3 Results

4.3.1 Evolutionary strategies with transient posterior signal

To investigate the relevance of the prior existence of a stable posterior signalling centre and the morphogen gradients emanating from it for the evolution of posterior growth and sequential segmentation we performed simulations that do and simulations that do not superimpose the existence of such a signalling centre. We start with an exploration of evolving segmentation strategies in the absence of a superimposed morphogen gradient (Fig. 4.2A, inset). Instead, we assume transient expression of gene 0 (the “morphogen”), restricted to the posterior-most cell of the embryo, and subject to decay in all cells. As a consequence, this gene will have the same level in the posterior cell as in all cells that descended from it and information on tissue polarity becomes quickly diluted. Under these conditions, a stable posterior signalling centre would have to evolve from scratch by evolving regulation of this transient signal to generate a stable posterior morphogen gradient (rather than being automatically present (François *et al.*, 2007, ten Tusscher and Hogeweg, 2011)). Alternatively, a segmentation mechanism could evolve which does not rely on a persistent posterior signal.

We perform four sets of 50 simulations: with/without cell-cell signalling (CCS), and with/without division noise (first four rows in Table 4.1). The simulations without CCS and without noise form a negative control group which does not have any symmetry-breaking mechanism: indeed, segmentation never evolves. In the remaining sets with either noise, CCS or both, segmentation does evolve. The vast majority of successful simulations (yielding more than three segments) evolves a segmentation strategy in which the tissue grows via one or more short-lived tissue-wide burst of divisions (Fig. 4.2A). The segments all appear roughly at the same time after the burst of divisions; we call this simultaneous segmentation.

In absence of CCS and presence of division noise, this simultaneous segmentation mechanism typically yields high phenotypic variability, often generating few segments and only occasionally producing many segments (e.g. Fig. 4.2B, see Methods). The segments are often irregular in size, with some much wider than others. The segmentation mechanism uses the stochastic delay of division in a few cells early in development, which changes the dynamics in those cells sufficiently to differentiate them from their neighbours (Fig. 4.2C and D). This mechanism therefore does not resemble *Drosophila*-type simultaneous segmentation but rather reflects the fact that the evolutionary process is free to evolve any possible growth and patterning modes.

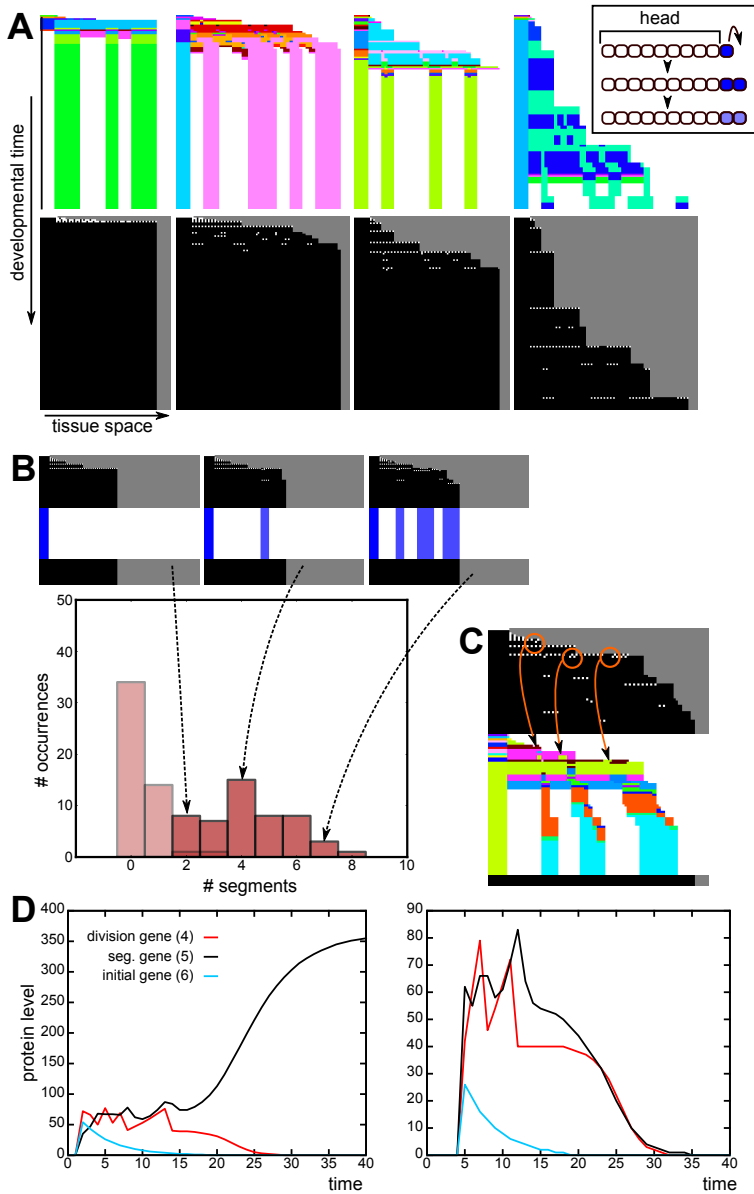


Figure 4.2. Transient posterior signal without CCS yields only simultaneous segmentation. A) Space-time plots of successfully evolved individuals, who mainly differ in the timing and number of tissue-wide division bursts. The right-most case only occurred once. The colour coding in the top row indicates cell type (based on the levels of all proteins), the white dots in the bottom row indicate new (just-divided) cells (see also Methods). Inset: The initial conditions of the morphogen dynamics used in these simulations. The head cells do not divide. The posterior-most cell has high morphogen concentration, which is inherited by its daughters. The morphogen gene can be regulated by the evolving network, just like any other gene, but is not regulated initially. (...) 109

Figure 4.2. (...) B) The development of evolved individuals is not robust. The histogram depicts the number of bands generated when the development of a single evolved individual is repeated 50 times (see Methods). Lighter bars indicate the number of too-short segments. Examples of the resulting development shown above with space-time plots of the divisions, with as inset the expression pattern of the segmentation gene. C) Division timing plays a role in determining segment position. Cells which by chance happened to divide a bit later (circled in orange) form a lineage with high expression of the segmentation gene. D) time plots that show the dynamics in a high-segmentation gene stripe (left) and a low stripe (right). Note the changes in concentration of the division gene due to divisions.

In the presence of CCS instead, simultaneous segmentation does not require cell division noise: 37 out of 50 simulations with CCS and without noise evolve segments, while 39 simulations with both CCS and noise yielded segmentation (see Table 4.1). In simulations with CCS, the average number of segments is doubled compared to the simulations without CCS. Moreover, 13 out of the 39 successful simulations with noise and CCS yield low phenotypic variability, meaning that they are able to make the same number of segments in more than 40 out of 50 repeated developments; we call this developmentally robust (Fig. 4.3A). Six of the simulations with CCS evolve simultaneous segmentation which uses cell-cell signalling to split developing segments in two (Fig. 4.3A). This is an alternative to the wave splitting mechanism observed in Turing pattern systems (Crampin *et al.*, 2002), as the evolved segment-splitting mechanism only relies on signals from direct neighbours rather than feedbacks between diffusive substances. In the presence of CCS, we also find the rare evolution of polarised growth: in two cases the head region is used as a signalling centre for divisions and gene expression oscillations (Fig. 4.3B). One simulation with CCS and without division noise evolves divisions that are restricted to a broad posterior zone, from which a number of segments appears sequentially (Supp. Fig. S4.2). This mechanism uses signalling from the formed segments to an undifferentiated zone to initiate localised division bursts which then yield new segments.

Altogether, without a superimposed posterior morphogen gradient we obtain a nearly 100% bias towards simultaneous growth and patterning, with the rare appearance of polarised growth. We therefore next test whether a polarised growth dynamics evolves more frequently if we select for tissue size but not for segmentation, thereby reducing the complexity of the selection target. By only selecting for tissue size, the majority of simulations still evolves tissue-wide division bursts as with simultaneous segmentation. We observe anterior growth (with or without initial division burst) in 6 out of 50 simulations, posterior growth in 4 cases (in 2 of which posterior growth is combined with a large initial tissue-wide division burst) and a combination of anterior and posterior growth in 2 out of 50 simulations (Fig. 4.4). In these cases, divisions are restricted to the posterior cell because it has only one neighbouring cell and thus receives less inhibitory signal, and a morphogen gradient never evolves. Polarised growth on one end of the tissue thus

seems a rare evolutionary outcome, given its low frequency even for a simpler selection target. Our results show that without a superimposed posterior morphogen gradient, evolution is unlikely to result in posterior growth and sequential segmentation.

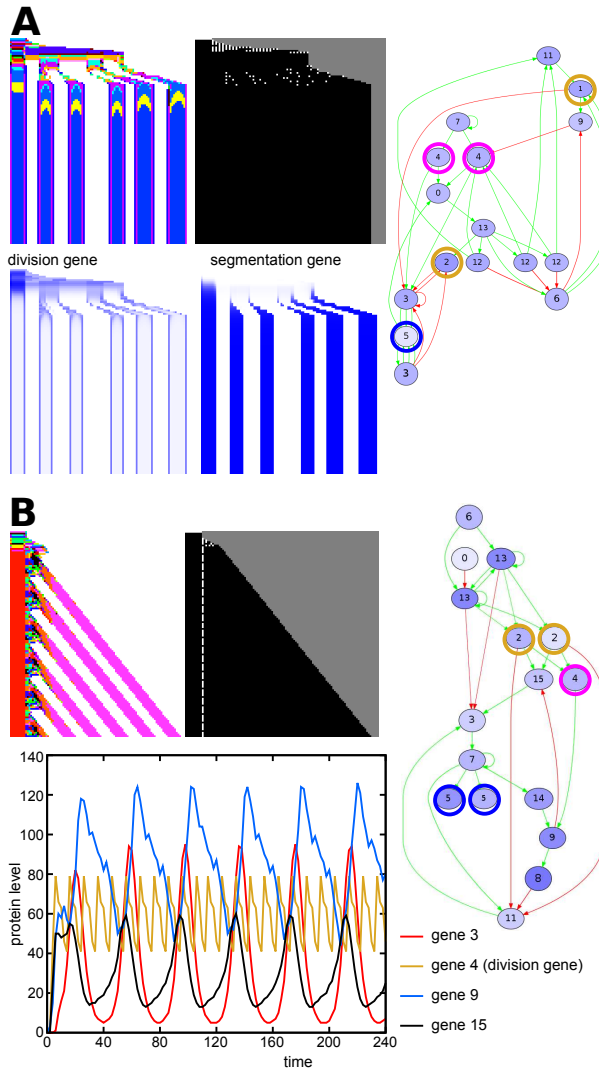


Figure 4.3. With CCS, different segmentation mechanisms can evolve. (transient signal, CCS) A) This individual uses cell-cell signalling at the boundaries of an emerging segment to split the segments in two. The segmentation gene and the division gene are maintained only in the two boundary cells of this primordial segment, because they receive different input from their neighbours. The division gene then generates a new burst of divisions in that cell, expanding the single cell into a new segment. (...)

Figure 4.3. (...) This mechanism piqued our interest because it superficially resembles the splitting of the *odd* stripes in *Tribolium* (Sarrazin *et al.*, 2012). It provides an alternative to Turing-like wave splitting in growing media (Crampin *et al.*, 2002), which uses diffusive signalling over longer distances, while this segment-splitting mechanism uses only direct CCS. Another difference is that in the Turing mechanism, the wave splitting results from growth, while here segment splitting directs divisions. Although cell divisions are thought to play a minor role in the axis extension of *Tribolium*, tissue wide divisions have been observed that could support the segment-splitting mechanism we find here (Nakamoto *et al.*, 2015). In *Tribolium* however, segment addition happens sequentially, while segment splitting here occurs in simultaneously generated segments. Furthermore, the mechanism in *Tribolium* is distinctly asymmetric: the secondary stripe that splits off is considerably narrower than the primary stripe. It thus remains an open question which mechanism causes segment doubling in *Tribolium*: Turing-like, the mechanism described here, or an as yet unidentified mechanism. B) This individual uses signalling cues emanating from the static head to stimulate divisions in the cell adjacent to the head. The graph depicts the gene expression oscillations that occur in this cell, which subsequently pattern the tissue. In the networks, the division gene is circled in magenta, the segmentation gene in blue and the signalling genes in yellow.

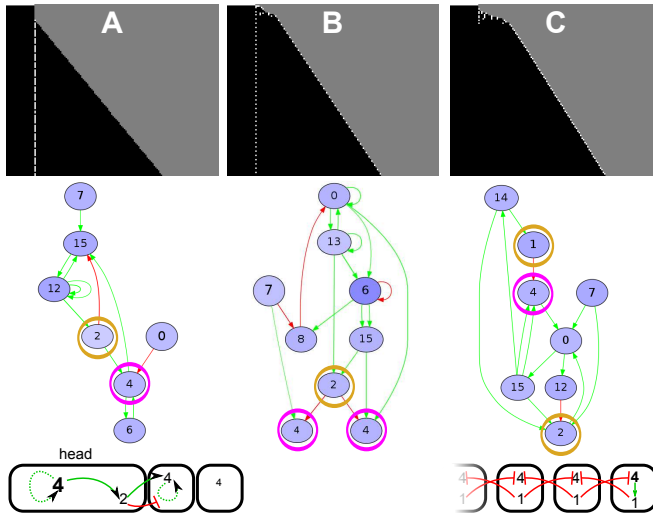


Figure 4.4. Selection only for tissue size occasionally yields anterior or posterior growth (transient signal, CCS) Examples of individuals which evolved anterior growth (A), posterior growth (C) or a combination of the two (B) in the absence of selection for segments (as described in the main text). A) Anterior growth exploits the fact that the head does not express gene 0 (morphogen) and does not divide; therefore it accumulates division protein (4). The head thus functions as a signalling centre. C) Posterior growth uses the fact that the posterior-most cell has only one neighbouring cell and thus receives less cell-cell signalling. The signalling genes 1 and 2 are circled in yellow, and the division gene in magenta.

4.3.2 Evolutionary strategies with persistent posterior signal

Next, we performed two sets of simulations with a persistent posterior signal, in the form of a superimposed posterior morphogen gradient: one set with and one without cell-cell signalling, and both with division noise (Table 4.1). To achieve this, the posterior-most cell now receive a morphogen that is subject to decay in all cells except this posterior-most cell. In these simulations, we find two qualitatively different strategies. The majority of simulations (53 out of 100) evolves a posterior growth zone combined with sequential segmentation, while the tissue-wide burst with simultaneous segmentation observed in the previous section is now less common (30 out of 100).

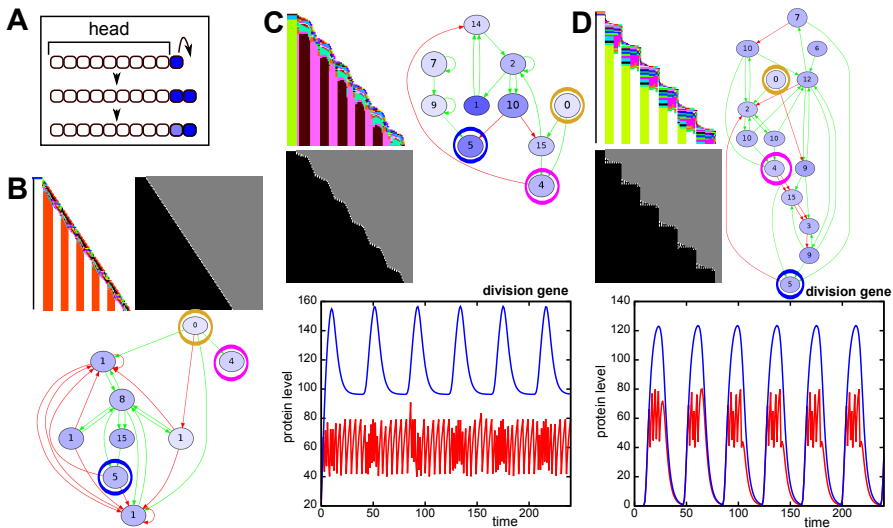


Figure 4.5. Three different types of sequential segmentation can evolve. (persistent signal, no CCS) A) In these simulations, the morphogen is highly expressed in the posterior-most cell. If that cell divides, morphogen expression is maintained in the posterior daughter, and its level decays in the other cells. B-D) Space-time plots, networks and division-gene dynamics of different types of sequential segmentation. In the networks, the morphogen is circled in yellow, the division gene in magenta and the segmentation gene in blue. In the graphs, the division gene dynamics are depicted only for the posterior-most cell, with high morphogen level. The red line shows the network dynamics if halving of the division protein due to divisions is taken into account. The blue line depict the dynamics if the network is run without halving the division protein once it reaches the division threshold. B) a smoothly growing individual. Note how the division gene is only regulated by the morphogen. C) “wavy” posterior growth. The growth zone keeps dividing, but sometimes its daughters also divide. Note the oscillating expression of the division gene in the posterior-most cell. D) “stair-like” posterior growth. The division gene strongly oscillates, and is therefore regularly low even in the posterior cell.

In the simulations with CCS we also find combinations of the simultaneous and sequential mechanism, where the first few segments are generated using a simultaneous mechanism, and the remaining segments arise through posterior growth combined with gene expression oscillations (5 simulations).

Often, the evolved sequential segmentation mechanisms are very developmentally robust with only small phenotypic variations, and they yield a large number of segments with a regular pattern (Table 4.1 and Fig. 4.5). Typically, gene expression oscillations in the growth zone are used to pattern the segments, resembling the mechanism in vertebrates and arthropods. We distinguish three common variations, differing in the distribution of divisions in space and time (Fig. 4.5B-D). When growth proceeds smoothly, the division gene is only regulated (directly or indirectly) by the morphogen (Fig. 4.5B). Other variants of posterior growth and sequential segmentation show a wavy or even stair-like growth pattern, reflecting non-continuous, burst-like division dynamics of the posterior growth zone. In these cases the division gene itself oscillates, with low amplitude in the wavy pattern or high amplitude in the stair-like pattern (Fig. 4.5C,D). These oscillations are caused by regulation of the division gene by other genes in the network that are part of the segmentation oscillator. In Supp. Fig. S4.4 we discuss some non-robust cases of sequential segmentation; there, the division gene is itself a part of the oscillator, making the oscillator sensitive to the stochastic nature of the divisions.

In absence of CCS, a posterior morphogen gradient improves the developmental robustness of the evolved simultaneous patterning mechanisms (Supp. Fig. S4.3, table 4.1). Cells may use differences in morphogen concentration rather than the differences arising through stochastic cell division for segmental patterning (in fact, division noise is now only a source of phenotypic variability). While adding CCS to simulations with transient posterior morphogen decreased phenotypic variability, in the presence of persistent posterior morphogen no further improvement was observed upon addition of CCS.

So far, we selected for increasing numbers of segments, thus only implicitly selecting for tissue growth. However, it can be hypothesised that body axis elongation -even in the absence of subdivision into segments- already confers an evolutionary advantage. We therefore compare the previously described simulations in which we selected only for segmentation (the set without CCS), to simulations in which we independently select for both axial elongation (to a particular target size) and segmentation. While the number of simulations yielding sequential mechanisms is the same (31), we now find that 8 simulations yield the combined simultaneous and sequential strategy, and only 6 simulations yield fully simultaneous segmentation. Thus, the bias towards sequential segmentation has become somewhat stronger.

The evolved developmental mechanisms look similar between the set with and the set without selection for axial elongation, but we find that the evolutionary trajectories that lead to these strategies differ markedly between the sets (Fig. 4.6). When we select for both axial extension and segmentation, in all simulations we first see the evolution of body axis extension to obtain a tissue close to the target size, and subsequently the evolution of a subdivision of the body axis into segments (Fig. 4.6A).

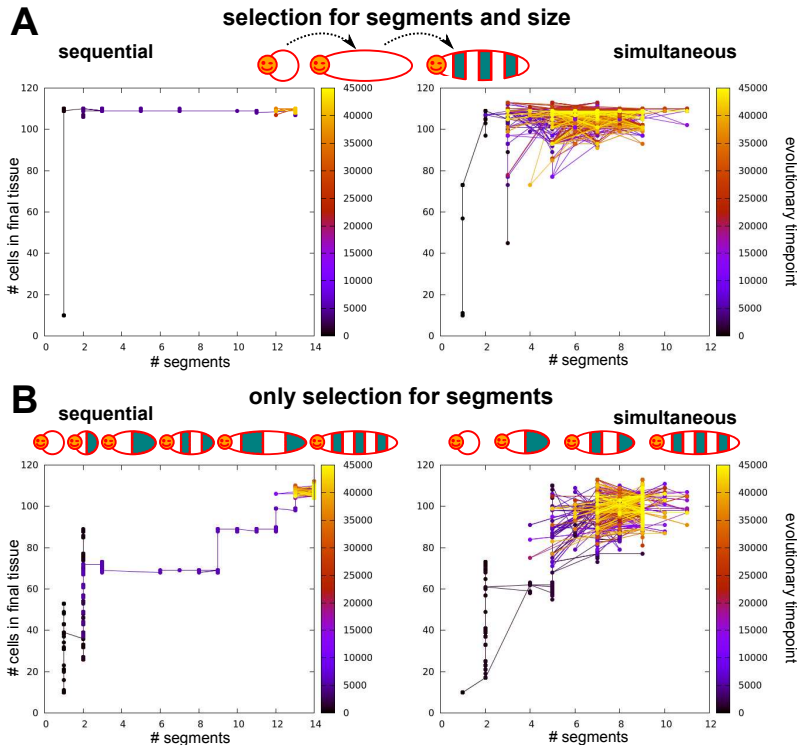


Figure 4.6. The order of evolutionary events differs between simulations with and without selection for axial extension. (persistent signal, no CCS) Graphs depict the evolution of final tissue size and segment number for one simulation. Left column: sequential segmentation, right: simultaneous segmentation. The colour of the nodes indicates the evolutionary timepoint. A) Evolutionary trajectories when selecting for axial extension and segmentation: tissue size evolves first to target size before segmentation evolves. B) Only selection for segments. For sequential segmentation, growth happens in phases. First the available tissue evolves to be filled with segments before tissue size increases further. For simultaneous segmentation, tissue size and number of segments evolve concurrently, but the process is noisy. Note the increased robustness towards the end of the simulation.

If instead selection is only on segment number, we see differences in the evolution of tissue size between simultaneous and sequential segmenters (Fig. 4.6B). In the simultaneous case, tissue size increases concurrently with segment number, although the evolutionary sequence is erratic due to the high phenotypic variability of the segmentation mechanism. At later evolutionary time points, we observe a decrease in the variation in tissue size and segment number (Fig. 4.6B). In the case of sequential segmentation, we instead observe a repeated sequence of first evolving a certain tissue size, and subsequently evolving the subdivision of this tissue with an increasing numbers of segments (Fig. 4.6).

We conclude that, of the possible evolutionary options to segment a tissue, posterior growth coupled to sequential segmentation has a higher potential to be a robust patterning mechanism, and is capable of generating more and more regularly shaped segments. In addition, when incorporating an ancestral posterior signalling centre involved in body axis polarity, it is also the most likely evolutionary outcome. This likelihood slightly increases when body axis extension evolves prior to segmentation.

4.3.3 Evolving determinate growth

From the previous section, it is clear that posterior growth with sequential segmentation is the most successful of the possible developmental strategies: it evolves more often, it is able to form many and regularly shaped segments, and has the potential to be developmentally robust. So far, we did not take into account that the evolved sequential segmentation mechanisms do not terminate growth at the end of development. Instead they evolve a growth rate that is tuned to allow them to grow to the target size within the constant, superimposed duration of development. If this duration of development were to be extended, larger individuals with a larger number of segments would automatically arise. While there are indeed bilaterian animals (like many annelids) which do continue growing indefinitely (Balavoine, 2014), most animals stop growing and making segments, for instance vertebrates and insects have a determinate number of segments and roughly determinate growth. We therefore decided to include selection for determinate growth, by applying a fitness penalty for division during the last 20 time steps (no CCS). (Note that the definition applied here for determinate and indeterminate growth is somewhat different from definitions used elsewhere (Harper *et al.*, 1986). See also the glossary.)

With increasing strength of this evolutionary pressure, a larger fraction of simulations yields simultaneous growth and segmentation, until the bias is completely reversed (Fig. 4.7A). In a subset of simulations stair-like sequential growth evolves, which allows for sequential growth while circumventing the fitness cost of late-stage divisions. Only very rarely (max 4 out of 50) does a simulation yield sequential segmentation with a mechanism that leads to the controlled halting of growth (Fig. 4.7A, space-time plots).

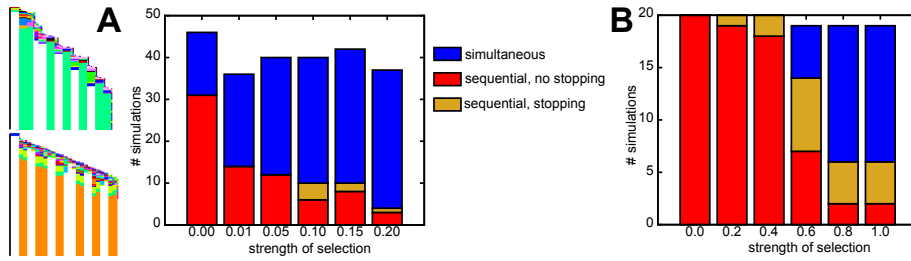


Figure 4.7. The time of onset of selection for stopping growth influences evolutionary outcome. (persistent signal, no CCS) A) The frequency with which simulations evolve simultaneous segmentation, increases with stronger selection pressures to stop growing. Only rarely do individuals evolve with determinate segmentation. B) When the selection pressure to stop growing is added after segmentation has evolved, sequential segmentation with determinate growth more frequently evolves. The number of simulations which switch to simultaneous growth does still increase with increasing selection pressure. 20 sequentially growing individuals were allowed to continue evolution with the added pressure. The selection pressure required to effect a change is higher; the period over which divisions are penalised is now also 40 steps instead of 20.

We reason that by applying the selection for determinate growth from the start of the simulation, we implicitly select for simultaneous growth, which is indeed determinate. To prevent this bias, sequential segmentation would have to have evolved before the appearance of this selection pressure. It also seems biologically reasonable to assume that determinate growth is a secondary trait: when comparing segmented animals with indeterminate and determinate growth, it seems that at least in arthropods the clades with determinate growth have more evolutionary derived, complex body plans. We speculate that determinate growth becomes more important upon evolution of segment specification, where e.g. locomotive appendages are limited to trunk segments and the abdomen is unsupported. To test the idea of secondary selection for determinate growth, we extract individuals from 20 earlier simulations in which sequential segmentation evolved without the pressure to stop growing; then we continue their evolution in the presence of this pressure. The outcome of these continued simulations depends on the strength of the evolutionary pressure to stop growing. If the pressure is too low, determinate growth does not evolve often (Fig. 4.7B). If instead the pressure is too high, the potential for growth and segmentation is often transiently lost after which a simultaneous mechanism evolves instead; we do not observe smooth transitions from sequential to simultaneous segmentation (Fig. 4.7B). However, between these two extremes lies a parameter region in which one-third of simulations evolve the capacity to stop growing while maintaining posterior growth and sequential segmentation (Fig. 4.7B). Thus, our hypothesis is confirmed, delaying selection for determinate tissue growth to a later evolutionary stage does indeed more often yield the evolution of sequential growth and segmentation combined with determinate growth.

In some of the cases where determinate growth evolves, the functional gene regulatory network expands to include a control gene. The expression of this control gene slowly increases over time until it passes a threshold and shuts down the division gene (Fig. 4.8A). This becomes evident in the different gene expression pattern in the last segment (Fig. 4.8A). More often however, the stopping mechanism relies on an oscillating gene that passes a threshold due to the slight stochasticity in divisions and shuts down the division gene (Fig. 4.8B). This latter mechanism yields large phenotypic variability, while the former mechanism is more reliable (Fig. 4.8C).

Developmental programs incorporating the first stopping mechanism (gene expression build-up) become slightly less robust compared to the original sequential segmentation mechanism without stopping growth, as stochastic divisions may influence the time at which the growth stopping protein level is being reached (Fig. 4.8C). Developmental programs applying the second mechanism (stochastic threshold passing) become significantly less robust, which logically follows from the fact that they rely on the stochasticity of divisions to determine when to stop (Fig. 4.8C). Note that in both cases segment size does remain regular.

Table 4.1. Evolved developmental strategies.

Left number: number of simulations yielding this mechanism (*figure nr.*)

Right number: average number of segments generated with this mechanism

| | simulation set | simultaneous | sequential | other | failed |
|---------------------------------------|-----------------------|-------------------|--------------------|--------|--------|
| Transient posterior morphogen | no CCS | 39 (4.2,S4.3) 3.2 | 0 | 0 | 11 |
| | no CCS, no noise | 0 | 0 | 0 | 50 |
| | with CCS | 34 (4.3,S4.1) 6.8 | 2 (4.3) 10.0 | 3 9.1 | 11 |
| | with CCS, no noise | 32 8.4 | 1 (S4.2) 13 | 4 9.3 | 13 |
| Persistent posterior morphogen | no CCS | 15 (S4.3) 6.4 | 31 (4.5,S4.4) 12.1 | 0 | 4 |
| | with CCS | 15 6.9 | 22 12.3 | 7 9.1 | 6 |
| | tissue size selection | 6 6.4 | 31 12.7 | 8 10.9 | 5 |

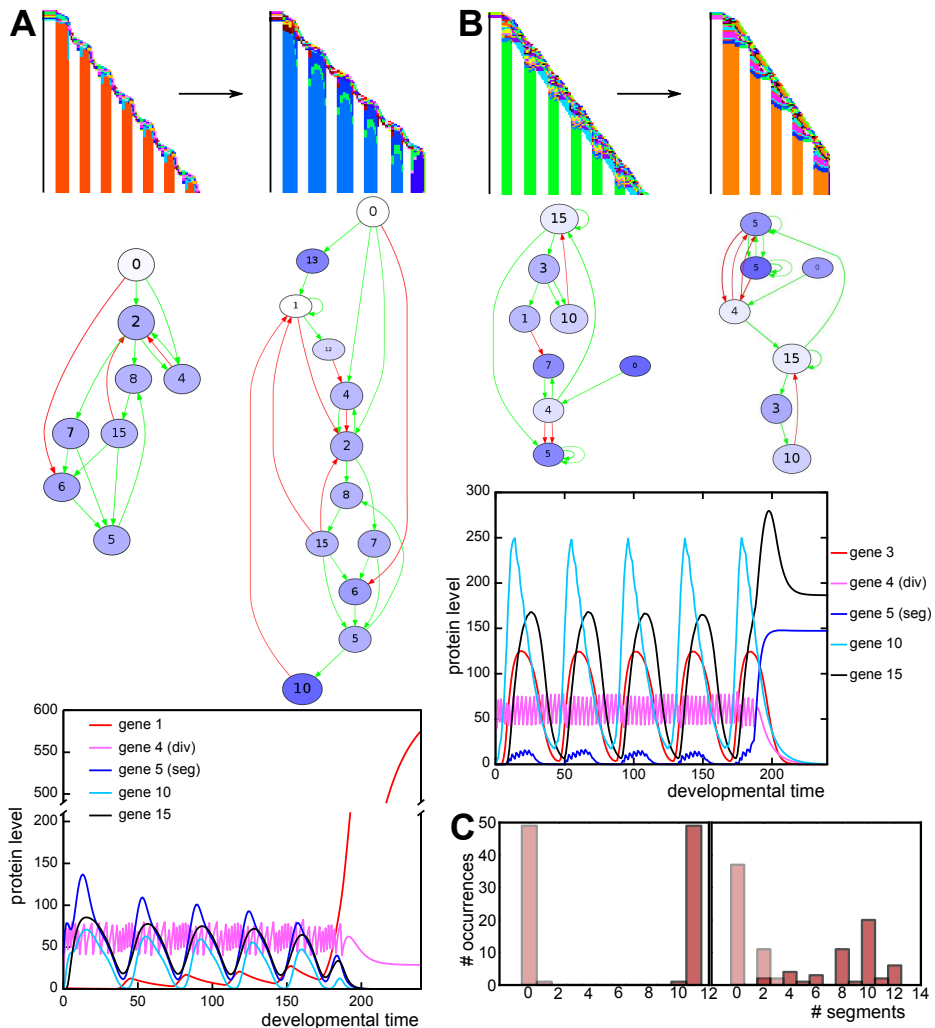


Figure 4.8. Mechanisms to stop growing differ in phenotypic variability (persistent signal, no CCS) A-B) Individuals capable of sequential segmentation, that are subsequently subjected to a pressure to stop growing. A) This individual stops growing by increasing the expression of gene 1 over time. Note the addition of the extra module for stopping growth in the network. B) This individual stops growing because gene 5 stochastically passes a threshold, above which its expression stabilises and switches off the division gene. C) The individual of (A) maintains its developmental robustness (left histogram), the one in (B) doesn't (right histogram). Both had very low phenotypic variation before the addition of the extra selection pressure, looking like the left histogram.

4.4 Discussion

A number of previous modelling studies has looked into the evolution of segmentation (see (ten Tusscher, 2013) for review) (François *et al.*, 2007, Fujimoto *et al.*, 2008, Salazar-Ciudad *et al.*, 2001a). These studies mainly focused on the evolution of sequential segmentation (as in vertebrates or short-germ insects) versus simultaneous segmentation (long-germ insect, *Drosophila*-like), investigating their similarities and suggesting potential evolutionary transitions. In the current study, we took a somewhat different approach, focusing on factors that may have contributed to the likelihood of evolving sequential segmentation. We aimed to explain its dominance as a segmentation mechanism and the order of events through which it arose. Taking a “worse-case approach”, we maximally allowed alternative mechanisms to evolve, and then determined under which conditions posterior growth and sequential segmentation predominates.

We found one main alternative developmental strategy besides sequential segmentation: simultaneous segmentation, in which after a short tissue-wide burst of divisions all segments appear roughly at the same time. This simultaneous mechanism is not similar to *Drosophila*-like segmentation, where a hierarchy of gene regulation robustly creates regular-sized segments. Note that the *Drosophila* strategy likely evolved secondarily, from an initial sequential segmentation mode, so one should neither aim nor expect a *Drosophila*-like segmentation to evolve from scratch in our simulations. Rather, the evolution of a quite different type of simultaneous segmentation in our simulations is a result of the freedom of the evolutionary process, which we use to distinguish evolutionary scenarios. The simultaneous strategy that evolved in our simulations generates irregularly sized segments, and the number of segments in genotypically identical individuals tends to be variable. In contrast, the evolved sequential segmentation generates a large number of regularly sized segments in a robust, reproducible manner, thus leading to larger fitness values at the end of evolutionary simulations. A number of subtypes of posterior growth and sequential segmentation evolved; the most notable mechanism involves regular, segmental oscillation dependent bursts of cell division in which two segments are down simultaneously - another benefit of the larger degrees of freedom of the model.

Stable posterior signalling is a prerequisite for sequential segmentation We showed that evolution of terminal addition type posterior growth is highly unlikely in the absence of persistent posterior signalling, independent of whether we selected for segmentation or body axis elongation. Under these conditions, the potential for symmetry breaking is restricted to the early phases of development, generating a bias in favour of an early tissue wide burst of divisions and against posterior growth. In absence of cell-cell signalling, simulations relied on the stochasticity of divisions to generate segments. In the presence of cell-cell signalling, lateral inhibition type patterning is used to pattern segments

during the tissue wide division burst. Our results thus suggest that the evolution of a posterior signalling centre is a crucial prerequisite for the evolution of posterior growth and sequential segmentation. Given the presence of a posterior signalling centre in all bilaterians as well as cnidarians, it can be safely assumed to represent an ancestral property (Martin and Kimelman, 2009, Niehrs, 2010, Petersen and Reddien, 2009). Thus, we can reformulate our findings and state instead that the prior evolution of a posterior signalling centre provided a strong bias towards the evolution of posterior growth and sequential segmentation.

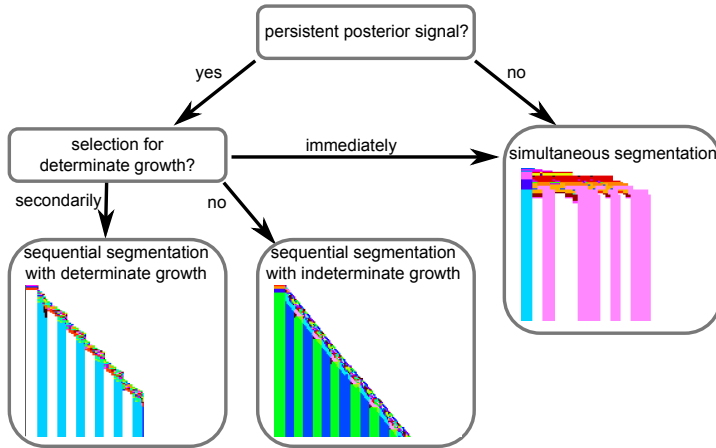


Figure 4.9. Summary of conclusions Flowchart summarising the results of different simulations. Note that the arrows only indicate the majority of simulations in a set, e.g. persistent posterior signal without determinate-growth selection occasionally yields simultaneous growth as well.

Determinate growth as a secondary trait In simulations incorporating a persistent posterior morphogen signal, selection for determinate growth completely reversed the evolutionary bias from sequential to simultaneous segmentation. We found that to evolve posterior yet determinate growth, the selection for determinate growth had to occur secondarily, after the evolution of posterior growth and sequential segmentation. Simple segmented animals such as millipedes and annelids contain large numbers of highly similar segments and many annelids appear to keep adding segments throughout their life (Balavoine, 2014). In contrast, insects and vertebrates develop a smaller, constant number of highly specialised segments after which posterior growth is terminated. We thus hypothesise that a constant segment number evolved secondarily, and was only selected for once segment specialisation arose and locomotive capabilities became restricted to a limited number of segments. Consistent with this, HOX genes, which are crucial

in segment specialisation, appear to be involved in terminating posterior growth (Young *et al.*, 2009). As an intermediate form, myriapods and the extinct trilobites stop adding segments when reaching maturity, but the final number of segments is variable (Jacobs *et al.*, 2005). This has been linked to their limited segment specialisation, where the exact number of segments is not that important. This is reminiscent of the form of determinate growth that evolves in our simulations, which is not robust and yields variable segment numbers.

Still, how to explain the fact that many unsegmented and metameric animals display determinate growth, for instance *C. elegans*? Assuming an unsegmented bilaterian ancestor, determinate growth may have evolved prior to sequential segmentation. In the current study we did not explicitly test for this, however we expect that even if we selected for only axial tissue growth without segmentation, simultaneous growth would arise if we also immediately selected for determinate growth. Given that ancestral state reconstructions suggest that terminal addition is an ancestral bilaterian trait (Gold *et al.*, 2015, Jacobs *et al.*, 2005) while the evidence is less conclusive for sequential segmentation, we would expect that determinate growth at least evolved secondary to posterior growth. Alternatively, the presence of determinate growth in unsegmented and metameric organisms can be explained by the presence of a bilaterian ancestor displaying terminal addition, sequential segmentation and determinate growth with many lineages subsequently completely or partly losing segmentation. Finally, a less-parsimonious scenario involves an unsegmented, indeterminately growing bilaterian ancestor, with parallel evolution of either determinate growth alone, or following sequential segmentation in several lineages.

In our current model we observed two mechanisms to stop posterior growth: one depending on stochastic changes in an oscillating gene, making it very non-robust, the other depending on the gradual build-up of a slowly decaying gene, yielding lower phenotypic variability. This latter strategy resembles a hypothetical mechanism proposed by Meinhardt for the sequential activation of HOX genes (Meinhardt, 1982, 2015). Considering the origin of the HOX cluster from tandem duplication of an early HOX gene (Kappen *et al.*, 1989), it can be envisioned that an early gene involved in regulating growth, became transformed through duplication into a sequentially activated HOX cluster in which gene order (i.e. posterior most HOX gene active) rather than level of a single gene now can act as a robust growth termination criterion. An interesting subject for future studies would thus be to investigate whether under combined selection for both segmentation and HOX-like specialisation domains, a HOX-type control of growth evolves. Such evolutionary outcomes may provide important answers in the puzzle of how the complex hierarchical network of *Drosophila* evolved from a sequentially segmenting predecessor; given the relatedness of gap and HOX genes and the suggested ancestral role for gap genes in growth control.

Conclusions In summary, we proposed an order of evolutionary events and selection pressures involved in generating posterior growth, sequential segmentation, and determinate growth. First, we provided evidence that the prior evolution of a stable posterior signalling centre has played a decisive role in evolving terminal addition and sequential segmentation. Then, we showed that the evolution of sequential segmentation combined with determinate growth can only take place by adding the selection pressure for determinate growth secondarily. Our study demonstrates that varying the onset of selection pressures can be a powerful tool in investigating the likely order of evolutionary events.

4.5 Acknowledgements

We thank Enrico Sandro Colizzi, Hilje Doekes, Bram van Dijk, Rutger Hermsen and Thomas Cuypers for helpful suggestions and discussion of the manuscript.

Table 4.2. parameter values

| parameter | values | remarks |
|-----------------------------------|------------|---|
| general | | |
| grid size | 30x30 | |
| evolutionary time steps | 50000 | |
| death rate | 0.5 | |
| initial # agents | 50 | |
| Development | | |
| developmental time steps | 240 | the number of integration steps |
| integration step size | 1. | fourth-order Runge Kutta integration |
| Morphogen decay rate | 0.2 | Only for persistent posterior morphogen |
| initial tissue size | 10 cells | of which 9 form the head |
| Gene and protein dynamics | | |
| gene product decay rate | 0.05 - 0.9 | |
| Hill constant of the TFBS | 10. - 400. | |
| gene transcription | 10. - 100. | |
| Mutational dynamics | | |
| Nr of gene types | 16 | |
| gene duplication | 0.006 | Note that with the gene, also its TFBS are duplicated. |
| gene deletion | 0.009 | |
| TFBS weight change | 0.001 | |
| TFBS type change | 0.001 | |
| TFBS duplication | 0.0015 | |
| TFBS deletion | 0.004 | |
| TFBS innovation | 0.001 | spontaneous emergence of new TFBS |
| Fitness | | |
| G: penalty per gene | 0.0001 | |
| T: penalty per TFBS | 0.00001 | |
| bonus for final tissue size | 0 or 0.1 | per cell added by division |
| target size | 110 cells | |
| penalty for exceeding target size | 1. | for each cell more than target size |
| control period | 20 steps | Period over which gene expression stability and sometimes number of late-stage divisions is measured. |
| U: expression variance penalty | 0.1 | Penalty per cell that has a variance in segmentation gene level > 5.0 during the control period. |

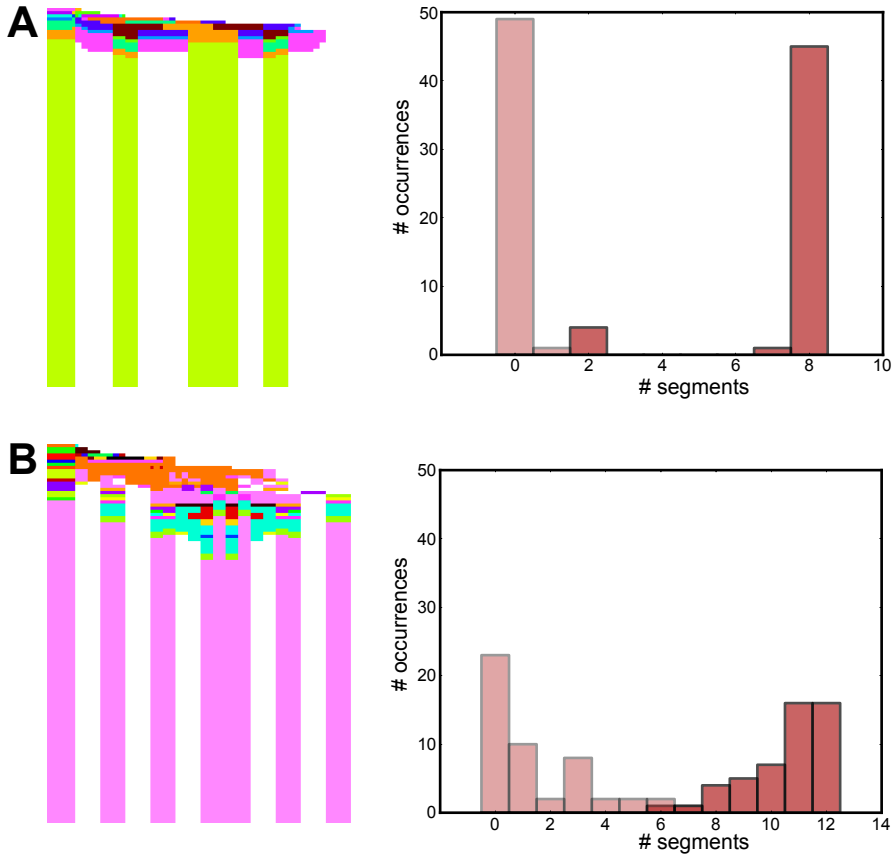


Figure S4.1. Cell-cell signalling allows for robust development in absence of posterior morphogen (but not always) (transient signal, CCS) A) This individual evolved to be very robust. It usually makes 8 segments and only in rare cases 7 good segments and one short segment. B) This individual did not evolve to become very robust despite the presence of cell-cell signalling genes. Nevertheless, it is usually able to make many more segments than the individuals evolved without cell-cell signalling. The histograms represent the variation in phenotypic outcome when an individual's development is repeated 50 times. In the histograms, the dark bars represent the good segments, and the lighter bars the too-short segments. See Methods for further explanation.

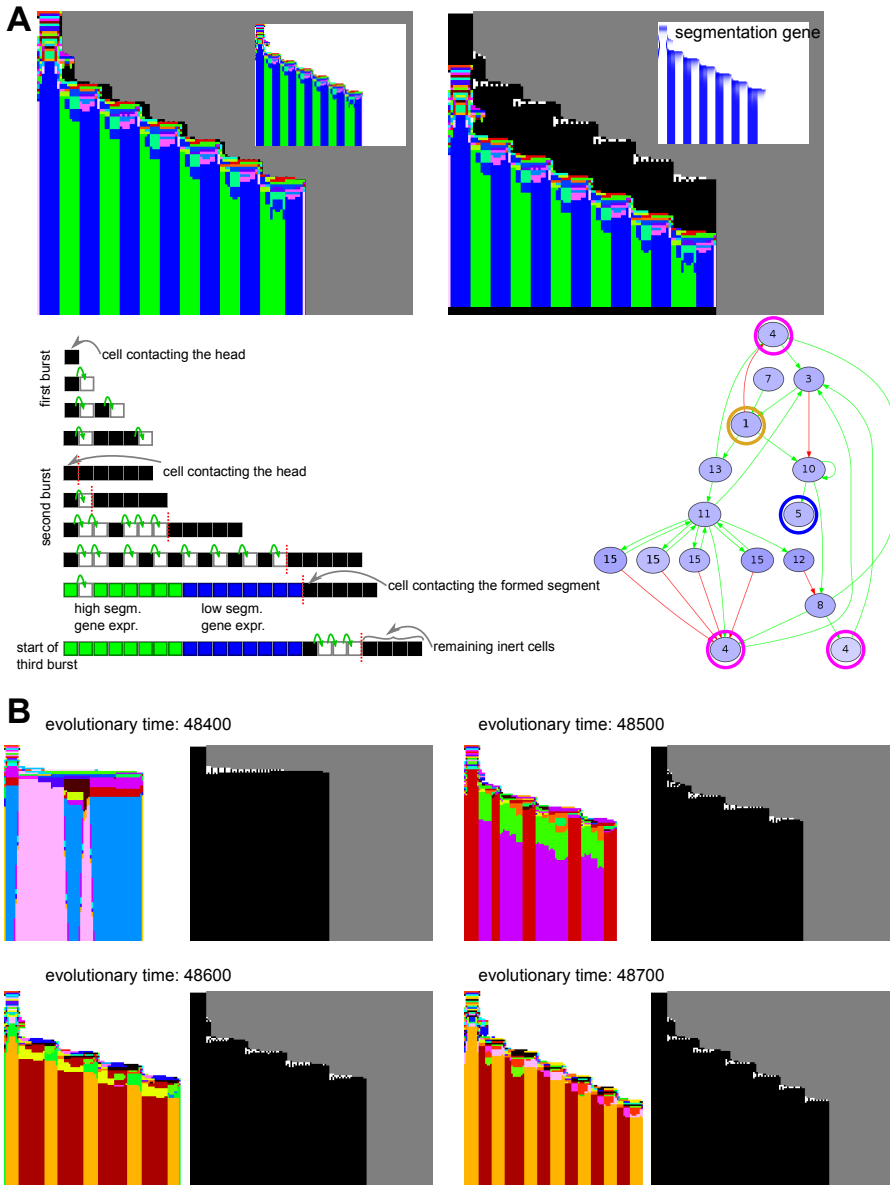


Figure S4.2. Very rare evolution of posterior growth in absence of persistent posterior signal (CCS, no noise) A) This individual first creates a small pool of 5 cells in an initial burst, which subsequently become completely inert (the black cells in the left space-time plot). The sixth cell, which is still in contact with the head, gets induced via CCS (gene 1, circled in yellow in the network) to initiate another burst of divisions, and these cells form the first two segments (high -> green; and low -> blue) next to the head (see cartoon). When these new segments mature, there is a short time window in which the most anterior of the 5 posterior inert cells gets induced to initiate a new burst. (...)

Figure S4.2. (...) Thus, the posterior pool gets depleted by one cell with each burst, putting a stop to the growth process. (Note the shortening of the dark, non dividing region in the space-time plots) In the network, the division gene is circled in magenta, and the segmentation gene in blue. B) This mechanism arises very late in evolution, from a simultaneously segmenting individual.

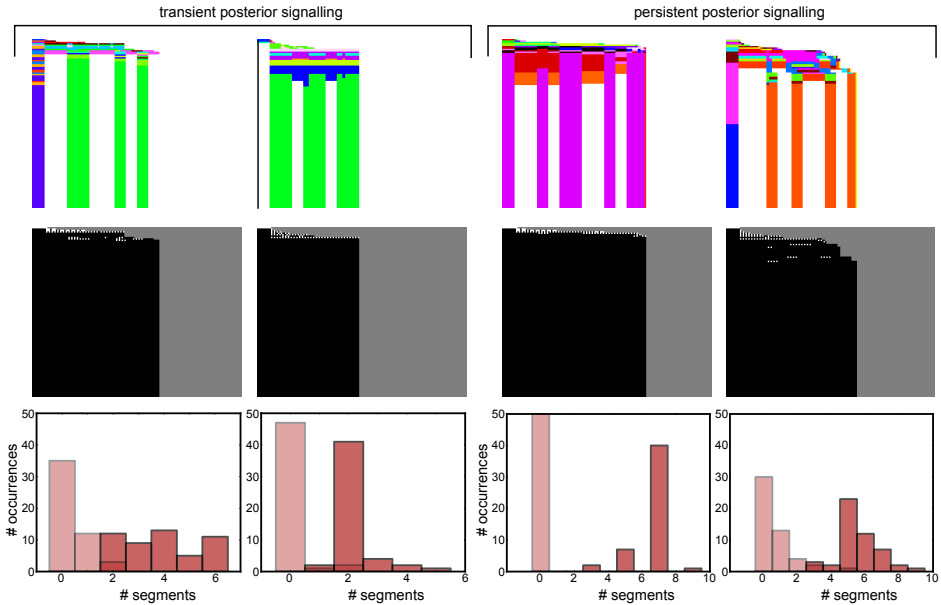


Figure S4.3. Simultaneous segmentation with persistent posterior morphogen can be more robust than with transient signal (no CCS) A comparison between transient and persistent posterior morphogen on the evolved simultaneous mechanisms. Those evolved with persistent signalling are capable of making more segments, and are sometimes very robust. When they are not robust, they still manage to make more segments on average. Note that the second individual with transient signalling may be robust, but this means it usually makes just two segments.

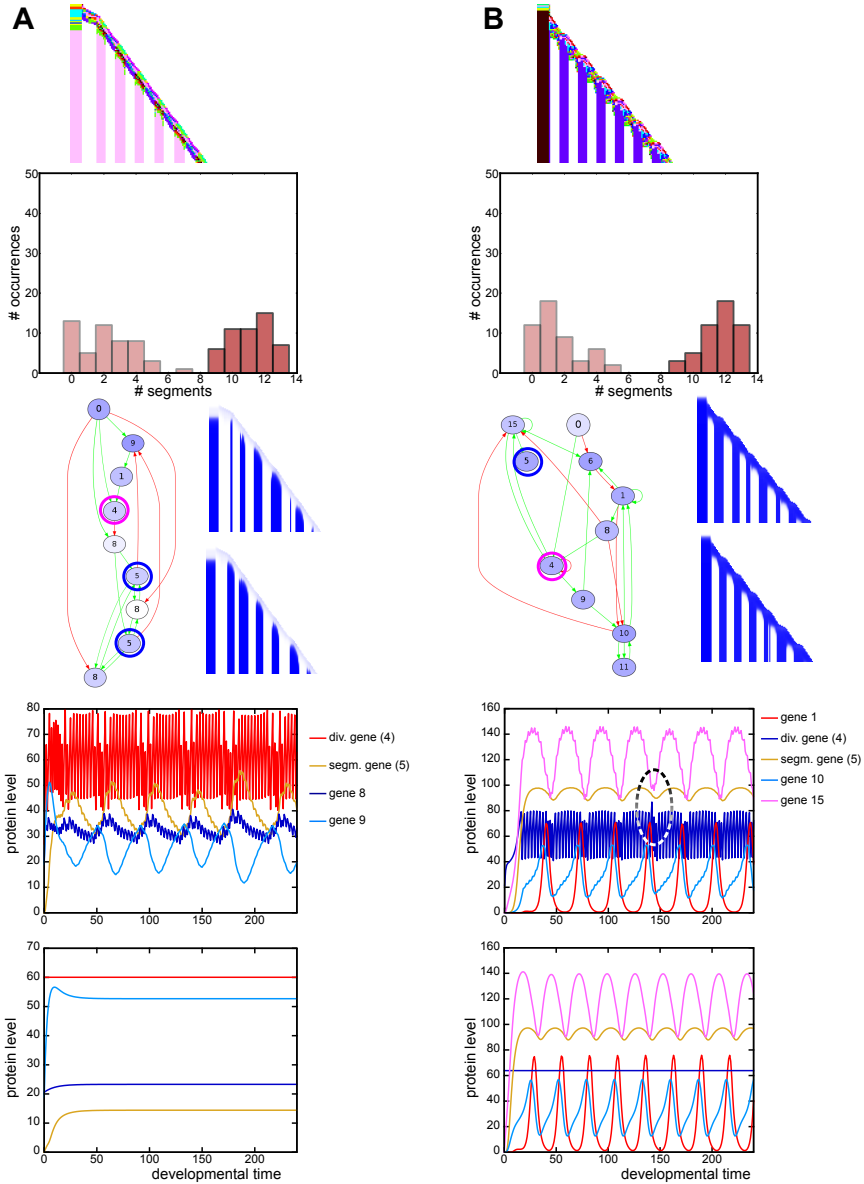


Figure S4.4. Two examples of non-robust development with sequential segmentation (persistent signal, no CCS) Top row, developmental space-time plots; second row, histogram of the outcome of 50 repeated developments; third row, evolved gene interaction networks and examples of variation in segmentation gene expression; fourth row, gene expression in the posterior cell with normal divisions; bottom row, gene expression in the posterior cell with averaged division gene expression instead of divisions. (...)

Figure S4.4. (...) A) The gene expression oscillations of this individual are entirely dependent on the regular divisions (oscillations are absent when the division gene is averaged), and are therefore very sensitive to division noise. B) While the oscillations in this individual do not depend on the divisions themselves, they are influenced by the level of the division gene. Stochastic changes in the timing of division result in changes in the level of the division protein (see circled point in the graph) which may alter the fate of the daughter cell emanating from the growth zone.

5

Segment-specific adhesion as a driver of convergent extension

RENSKE M.A. VROOMANS, PAULIEN HOGEWEG AND KIRSTEN H.W.J. TEN
TUSSCHER (2015)

PLoS Computational Biology, 11(2): e1004092

Abstract

Convergent extension, the simultaneous extension and narrowing of tissues, is a crucial event in the formation of the main body axis during embryonic development. It involves processes on multiple scales: the sub-cellular, cellular and tissue level, which interact via explicit or intrinsic feedback mechanisms. Computational modelling studies play an important role in unravelling the multiscale feedbacks underlying convergent extension. Convergent extension usually operates in tissue which has been patterned or is currently being patterned into distinct domains of gene expression. How such tissue patterns are maintained during the large scale tissue movements of convergent extension has thus far not been investigated. Intriguingly, experimental data indicate that in certain cases these tissue patterns may drive convergent extension rather than requiring safeguarding against convergent extension. Here we use a 2D Cellular Potts Model (CPM) of a tissue prepatterned into segments, to show that convergent extension tends to disrupt this pre-existing segmental pattern. However, when cells preferentially adhere to cells of the same segment type, segment integrity is maintained without any reduction in tissue extension. Strikingly, we demonstrate that this segment-specific adhesion is by itself sufficient to drive convergent extension. Convergent extension is enhanced when we endow our *in silico* cells with persistence of motion, which *in vivo* would naturally follow from cytoskeletal dynamics. Finally, we extend our model to confirm the generality of our results. We demonstrate a similar effect of differential adhesion on convergent extension in tissues that can only extend in a single direction (as often occurs due to the inertia of the head region of the embryo), and in tissues prepatterned into a sequence of domains resulting in two opposing adhesive gradients, rather than alternating segments.

5.1 Introduction

Convergent extension refers to the simultaneous narrowing and extension of tissues due to extensive cell rearrangements, and is a key morphogenetic event during formation of the bilaterian body plan. In bilaterian animals, convergent extension first occurs when the main body axis forms and extends, pushing the head and tail further away from each other. Although this axis extension is universal in bilaterians, the cell and tissue behaviour causing it differs widely between species (for reviews see (Keller *et al.*, 2000, Tada and Heisenberg, 2012, Wallingford *et al.*, 2002, Yin *et al.*, 2009)). In *Xenopus* for example, dorsal mesodermal cells polarise and change their adhesive properties (reviewed by (Skoglund and Keller, 2010)); cells then crawl between each other in a zipper-like process called intercalation (Keller *et al.*, 2000, Wallingford *et al.*, 2002). In contrast, convergent extension of zebrafish mesoderm consists of two processes: directed migration to the dorsal axis and intercalation (Sepich *et al.*, 2005, Wallingford *et al.*, 2002, Yin *et al.*, 2009). Finally, *Drosophila* germband extension occurs in a tightly connected epithelium, where cells intercalate by contracting those parts of the membrane that have a dorsal-ventral orientation (Bertet *et al.*, 2004, Rauzi *et al.*, 2008, 2010).

Convergent extension is an inherently multiscale process, in which subcellular contractility and adhesion, cell level polarity and migration, and tissue level deformations are involved. Models incorporating this multiscale nature are of key importance to study the feedback interactions that give rise to tissue extension. Thus far, models are largely conceptual in nature, testing whether an experimentally observed (sub)cellular process or hypothetical mechanism can indeed drive convergent extension (Backes *et al.*, 2009, Brodland and Veldhuis, 2012, Honda *et al.*, 2008, Weliky *et al.*, 1991).

Among the identified mechanisms capable of driving convergent extension confirmed by these models are lamellipodia formation (Brodland and Veldhuis, 2012, Brodland, 2006), directed mitosis (Brodland and Veldhuis, 2012), oriented membrane contraction (Honda *et al.*, 2008, Rauzi *et al.*, 2008), cell extension or protrusions (Backes *et al.*, 2009, Honda *et al.*, 2008, Weliky *et al.*, 1991) and anisotropic differential adhesion (Zajac *et al.*, 2003). These different mechanisms also differ in the origin of the directional signal, the cue which informs cells into which direction to move. In the models including either directed mitosis, lamellipodia or oriented membrane contractions (Brodland and Veldhuis, 2012, Brodland, 2006, Honda *et al.*, 2008), this direction is explicitly imposed in the model by telling cells in which direction to extend. In contrast, in the models with anisotropic differential adhesion and cell elongation (Backes *et al.*, 2009, Zajac *et al.*, 2003), there is no global information: cells have internal polarity and through cell-cell interactions the cells align. Other models are somewhere in between; Rauzi *et al.* (Rauzi *et al.*, 2008) use experimental data on the polar distribution of actomyosin, resulting in a coordinated contraction of only dorso-ventrally oriented membranes. The model by Weliky

et al. (Weliky *et al.*, 1991) does not impose the direction in which cells extend, but includes two boundaries enclosing the tissue which inhibit cell extensions, thus providing an overall bias.

Regardless of how cell/tissue polarity is incorporated in these models, convergent extension has so far always been studied in homogeneous tissues consisting of cells with identical fates. However, axis extension usually does not occur in homogeneous tissues, but rather in tissues that have been or progressively become patterned into regions of different cell fate. In *Tribolium* for instance, segments are formed by an oscillating gene clock, shortly after which the newly segmented part of the tissue starts to narrow and extend (Benton *et al.*, 2013, Choe *et al.*, 2006, El-Sherif *et al.*, 2012, Sarrazin *et al.*, 2012). Therefore, an interesting question is how patterns are maintained under convergent extension, which leads to extensive cell rearrangements and therefore potentially mixes up cells of different fate. Considering this, it is striking that in *Xenopus*, the antero-posterior patterning of the mesoderm is crucial for convergent extension (Ninomiya *et al.*, 2004); also in *Drosophila*, a segmented body pattern is essential for germband extension (Irvine and Wieschaus, 1994, Zallen and Wieschaus, 2004). This leads to the intriguing suggestion that rather than segments becoming lost due to convergent extension, these segments may actively drive convergent extension.

Since the interplay between tissue patterns and convergent extension has so far received little attention, we use a computational model to investigate how a segmented tissue pattern can be maintained during convergent extension, and whether and how such a pattern may itself drive convergent extension. We use the Cellular Potts model (CPM) formalism (Glazier and Graner, 1993, Graner and Glazier, 1992), which has been successfully used to model different mechanisms of convergent extension in a homogeneous tissue (Backes *et al.*, 2009, Zajac *et al.*, 2003), as well as several other morphogenetic processes like somitogenesis (Hester *et al.*, 2011), ommatidia formation in *Drosophila* (Käfer *et al.*, 2007), and *Dictyostelium* culmination into a fruiting body (Marée and Hogeweg, 2001). CPM is particularly suitable for performing the type of multiscale simulations necessary to investigate convergent extension since it endows cells with an explicit size and shape, allowing for both subcellular resolution and deformation, as well as cell level properties such as adhesion and migration (Marée *et al.*, 2007).

In this work, we show that convergent extension by itself tends to disrupt a segmented gene expression pattern that was previously formed. We demonstrate that this disruption may be counteracted by letting cells adhere preferentially to cells of the same segment type. Furthermore, we find that such segment-specific adhesion by itself can both provide the directional signal and serve as a driving force for convergent extension. When we add a simple form of directional persistence (representing inertia in the cell's direction of movement due to the delay caused by cytoskeleton-recycling dynamics) this substantially

increases the efficacy of convergent extension through segment-specific adhesion. The latter is especially true in larger and stiffer tissues, where segment-specific adhesion alone is insufficient to cause a significant tissue shape change.

5.2 Results

5.2.1 The model

For all of our simulations we used a 2D CPM model with two different cell types (red and green) which represent segments with different identities. These *in silico* cell types can either have no segment-specific adhesion -a green cell will then adhere equally strongly to a red cell as to a green cell- or have segment-specific adhesion, meaning that cells prefer to stick to cells of the same type. In both cases, cells adhere more to other cells than to the surrounding medium, such that the tissue does not fall apart into separate cells or tissue types. The medium itself has no other properties than its adhesion with cells.

In CPM, adhesion is regulated by J values, which represent the surface energy (per amount of contact surface) between cells of the same type, between cells of a different type, or a cell and medium. The CPM tries to minimise the total energy of the system, so contacts with lower J values are preferred. The strength of adhesion or repulsion between cells depends on the difference between J values, which can be conveniently represented by the surface tension (γ) (Glazier and Graner, 1993). The γ values are calculated from the J values as follows: $\gamma_{i,j} = J_{i,j} - \frac{J_{i,i} + J_{j,j}}{2}$, where i and j represent different cell types (m=medium, r=red, g=green). Note that $J_{m,m} = 0$. We will refer to γ values throughout this paper, J values are mentioned in the figure legends. A positive $\gamma_{i,j}$ value means that cells prefer to adhere to cells of the same type, whereas negative values indicate that cells of different types prefer to mix. We will only use positive or 0 values for γ .

For a subset of simulations, we added a so-called persistence mechanism to our model. Persistence is the tendency of cells to maintain their previous direction of movement (memory or inertia), due to the non-instantaneous turnover of the cytoskeleton (Ridley *et al.*, 2003). When persistence is strong, cells are able to migrate rapidly in a consistent direction, as observed for example in lymphocytes and in gastrulating cells in zebrafish. We implemented persistence by giving cells a favoured (target) direction of movement. Despite this bias, the cell is not always able to move exactly in this direction due to hindrance by other cells or simply random fluctuations. Therefore, this target direction was regularly (after a fixed number of simulation steps) updated with the cell's actual direction of displacement, representing the eventual remodelling of the cytoskeleton (Fig. 5.1C)(Beltman *et al.*, 2007). A cell moving this way performs a persistent random walk.

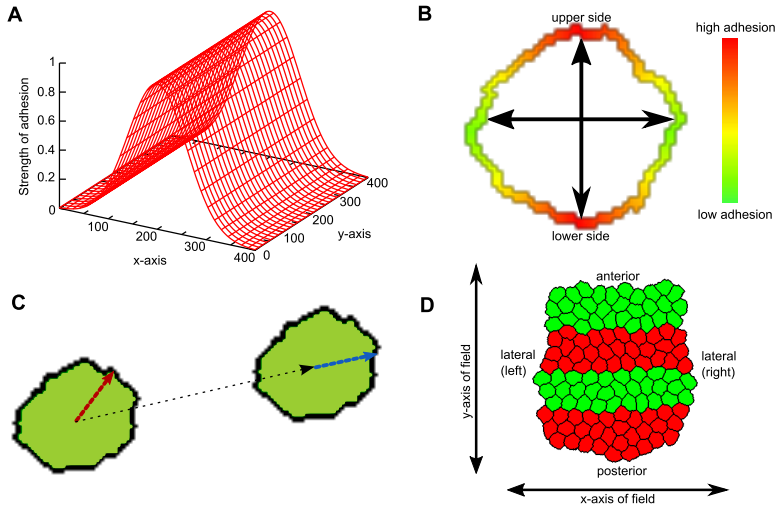


Figure 5.1. Model setup. (A) Graded adhesion. In the centre of the field, adhesion between cells is higher than at the boundaries. This difference is smoothly graded according to a Gaussian distribution. (B) Axial adhesion. The parts of the membrane that are coloured red (and indicated by the vertical axis) adhere more to neighbours than the sides, which are coloured green (and indicated by the horizontal axis). The strength of adhesion is graded smoothly across the membrane, and the orientation of the axes is fixed and identical for all cells. (C) Persistence mechanism. The cell has a higher propensity to move approximately in its target direction (red vector) than the opposite direction. Every s Monte Carlo Steps (MCS), this vector is updated according to the actual displacement of the cell (black vector; new vector: blue). (D) Initial state of the tissue. Cells are placed closely together, so that they form a coherent tissue at the start of the simulation. The tissue is already subdivided into regular segments of identical widths and cell numbers.

Initially, we also tested two explicit mechanisms for convergent extension, which both used global information to direct the cells. The first mechanism, called graded adhesion, was based on the observation that mesoderm cells in zebrafish follow a gradient in cadherin activity towards the central axis ((von der Hardt *et al.*, 2007), reviewed in (Tada and Heisenberg, 2012)). In our model, we implemented this by imposing a static gradient of cell adhesion, where the location of the cell in the field determined how strongly it adhered to neighbouring cells; cell contacts that were closer to the centre of the x-axis adhered more strongly than those that were farther away (Fig. 5.1A). The second mechanism, called axial adhesion, was an adapted version of the mechanism presented by Zajac *et al.* (Zajac *et al.*, 2003). This mechanism was based on the observation that intercalating cells in *Xenopus* are polarised and elongated, and the conjecture that these cells may also have a polarised distribution of adhesive molecules along their membrane. Figure 5.1B shows the basic idea: the upper and lower sides of a cell (defined by the

y-axis of the field) have a higher density of adhesion molecules than the left and right sides. A cell's adhesion to a neighbouring cell is then a product of the local density of adhesion proteins on both cells, which is approximated by adjusting the J value (we don't explicitly model adhesion proteins). We chose the axes for adhesive density such that the tissue should extend perpendicularly to the segments.

Later on, these two mechanisms were no longer applied.

For a detailed description of the implementation of all mechanisms, we refer to the Methods section.

We initiated the *in silico* tissues with a regular, segmented pattern of red and green cells (Fig. 5.1D). For convenience, we use the terms anterior-posterior (A-P) axis and medio-lateral (m-l) axis when we talk about the major and minor body axes of the tissue (which may have any orientation in the field). When we refer to the axes of the field (which are fixed), we simply use x-axis and y-axis. In most simulations however, the y-axis and A-P axis had the same orientation, meaning that the tissue extended in the direction of the y-axis of the field.

5.2.2 Segment-specific adhesion required to maintain segments during convergent extension

To study the effect of convergent extension on a pre-segmented tissue pattern, we started with the incorporation of either of the two explicit global mechanisms (graded adhesion or axial adhesion) without including segment-specific adhesion or persistence. We observed for both convergent extension mechanisms that the tissues extended and narrowed, but that cells at the boundaries invaded other segments, with some losing all contact with their designated segment (graded adhesion, Fig. 5.2A, S1 Video; axial adhesion, Fig. 5.2D, S2 Video). Note that the strength of both mechanisms was relatively low in these tissues, and that the loss of segment integrity became more pronounced when the strength of the mechanisms was increased (S1 Figure).

Next, we added a small positive surface tension between the red and green celltype ($\gamma_{r,g} = 4$, $\gamma_{c,m} = 4$), causing preferential adhesion to same-segment-type cells. This sufficed to prevent cells from leaving their segment during convergent extension (Fig. 5.2B and E; S3, S4 Videos). Moreover, the boundaries of the segments were much straighter. To determine whether this differential adhesion caused any additional differences, we tracked the total direction of movement of each cell over the whole course of a simulation (Fig. 5.2, vector plots). In all cases we see the typical pattern of convergent extension: vectors directed inwards on the lateral sides, and outwards at the anterior and posterior ends. Despite the preservation of segments when segment-specific adhesion was present, there was very little difference in the appearance of the vectors. We colour-coded the

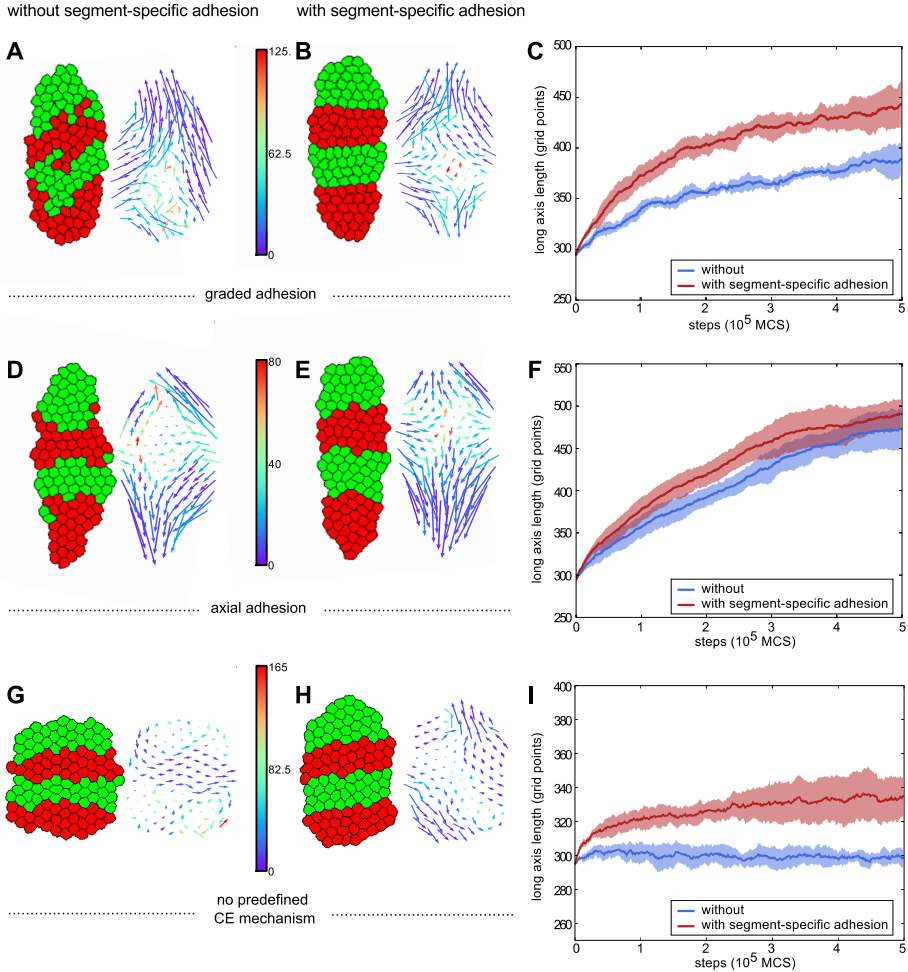


Figure 5.2. Maintenance of a segmented pattern during convergent extension requires segment-specific adhesion. Left images display tissue at the final step of the simulation (at 500,000 MCS). Right images contain the displacement vector of each cell in the simulation. The tail of a vector is located at the start position of the corresponding cell, the head at the end. The colour denotes the average angle of the vector with its neighbouring vectors. (A,B,C: row 1) Simulations with graded adhesion, strength $w = 11$. (D,E,F: row 2) Simulations with axial adhesion, strength $\beta = 2$. (G,H,I: row 3) Simulations without a predefined convergent extension mechanism. (A,D,G: col 1) Simulations without segment-specific adhesion, $J_{c,m} = 10$, $J_{c,c} = 16$. (B,E,H: col 2) simulations with segment-specific adhesion ($J_{c,m} = 10$, $J_{r,g} = 16$, $J_{r,r} = 12$, $\gamma_{r,g} = 4$) (C,F,i: col 3) Length of the long axis of the tissue as a function of simulation steps (MCS). Blue is without and red is with segment-specific adhesion. The curves are averaged over 5 runs of the model, shading indicates standard deviation.

displacement vectors according to the average angle with their neighbours (Fig. 5.2A, B, D, E). It seems that in the presence of differential adhesion cell migratory dynamics are slightly more coherent.

The vector plots in the cases with segment-specific adhesion (Fig. 5.2B, E) suggest that there was a considerable amount of A-P movement, which was unexpected given that cells remained restricted to their own segment. We checked whether this restriction led to a limitation of axis extension (Fig. 5.2C, F). Strikingly, segment-specific adhesion did not limit axis extension, but in fact enhanced it.

5.2.3 Segment-specific adhesion as a driver of convergent extension

The fact that segment-specific adhesion seemed to enhance axis extension, (Fig. 5.2C, F) prompted us to investigate the effect of segment-specific adhesion without any additional mechanism for convergent extension (Fig. 5.2G-I). Compared to tissue without segment-specific adhesion (Fig. 5.2G), tissue that had a small amount of segment-specific adhesion (the minimum amount needed to maintain segments in the presence of an explicit convergent extension mechanism), elongated significantly (Fig. 5.2H). Furthermore, convergent extension occurred without the cells or tissue having an explicit notion of their A-P axis (as opposed to the simulations in Fig. 5.2A-F, where the direction was imposed). This directionality now arose automatically, from the orientation of the interface between segments. The vector plot of the tissue with segment-specific adhesion (Fig. 5.2H) resembled the pattern generated by the graded and axial adhesion mechanisms, albeit with less extensive movement. This pattern was absent in the tissue without segment-specific adhesion (Fig. 5.2G). These results showed that segment-specific adhesion may not only be able to maintain segments, but could also be a driving force of convergent extension on its own.

To further investigate this possibility, we varied the surface tension between red and green segments (difference in adhesion between like and unlike cells, $\gamma_{r,g}$), and the tension of cells with the medium ($\gamma_{c,m}$). With increasing intersegment tension, having contact surface between segments becomes less energetically favourable, creating the tendency to reduce the segment interface. This caused the segments to round up more and so become narrower and thicker, so also the entire tissue extended and narrowed more strongly for increasing ($\gamma_{r,g}$) (Fig. 5.3A). Moreover, the more the tissues extended, the more the vector plots in Fig. 5.3A resembled the typical pattern of convergent extension.

Tissue extension is counteracted by increasing the tension with the medium, because extension and narrowing leads to a larger contact surface with the medium, which becomes less favourable with larger $\gamma_{c,m}$. Put differently: if cells prefer not to be in contact with the medium, the tissue as a whole will remain more rounded (minimal surface with

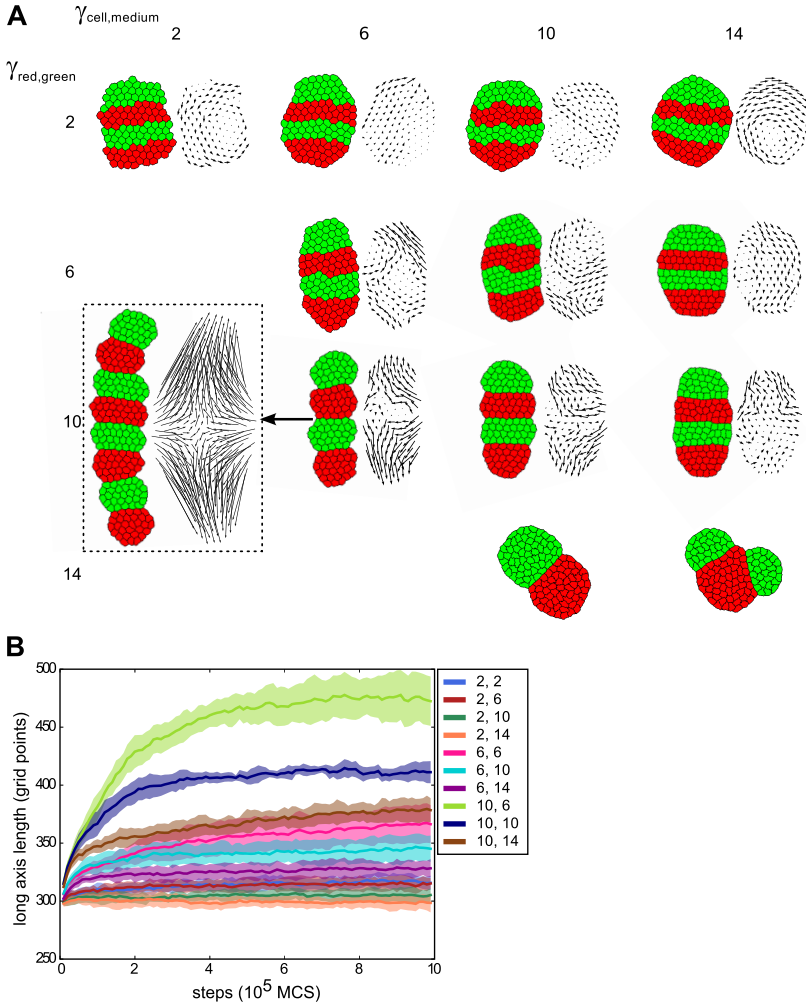


Figure 5.3. Segment-specific adhesion leads to convergent extension for a wide range of parameters. (A) Parameter space of a tissue of four segments with varying values for $\gamma_{c,m}$ and $\gamma_{r,g}$ (J values can be found in Supplementary Table 1). Initial segment width: 3, length: 10. For each set of parameters, 10 simulations were run over 1,000,000 MCS, representative final states are displayed. Only for the bottom two parameter sets we observed merging of segments, in resp. 9 and 8 out of 10 simulations. The dashed box shows a tissue initiated with 8 segments but with otherwise equal parameter settings to (10,6), to show the relative independence of simulation outcome from the number of segments. Vector plots were corrected for whole-tissue rotation. (B) For the same simulations, the length of the long axis of the tissue as a function of simulation steps (shading indicates standard deviation). The two parameter sets for which segments merged were not included, nor the simulation with 8 segments.

the medium) and therefore extend less. The final amount of extension therefore depended on the balance between the two opposing tensions. For the case in the parameter space with the most extreme extension ($\gamma_{r,g}=10$, $\gamma_{c,m}=6$), the tissue extended to about $1.5\times$ its original length (Fig. 5.3B, S5 Video). When we included more and thinner segments, the tissue extended even further (to more than $2\times$ the original length); otherwise, the results were qualitatively similar (box in Fig. 5.3, S2 Figure).

Occasionally, we observed that two segments of the same celltype contacted each other and merged, thus reducing the number of segments (Fig. 5.3A, bottom right; S2 Figure, bottom row; S6 Video). This biologically unrealistic behaviour only occurred for very strong differential adhesion, while biologically relevant behaviour prevailed in the remaining, considerably larger part of the parameter space that we explored.

5.2.4 Convergent extension by segment-specific adhesion enhanced with a persistence mechanism

So far, the *in silico* tissues with segment-specific adhesion reached their final length within about the same time scale as the explicit mechanisms. So far however, we used relatively small and loosely connected tissue. Therefore, we decided to investigate the efficacy of segment-specific adhesion in both larger and stiffer tissues.

In larger tissues, cells would need to travel greater distances to achieve the same degree of extension; this could potentially mean that the same process takes much longer in a larger tissue. It has indeed been suggested that if surface tension alone had to drive large changes in tissue shape, the process would take unrealistically long (Grima and Schnell, 2007). In Fig. 5.4A we compare two *in silico* tissues with the same surface tensions and the same ratio between the length and width of a segment, but one consisted of four times more cells (the number of cells in both the length and width of the segments was doubled). Because of the difference in total size, we used the ratio of the long axis over the short axis of the tissue to compare the extent of axis extension. It can be derived from first principles that for tissues with the same surface tensions and the same axis ratios at the start, the final axis ratio should be the same as well (S1 Text). As expected, the larger tissue extended at a much slower pace than the small tissue and did not reach the same axis ratio within the span of the simulation.

In the model we used so far, cells retained no memory of the direction in which they previously moved, and could change their direction of movement instantaneously. However, biological cells are to some extent persistent: due to polarisation and turnover dynamics of the cytoskeleton they tend to move for some time in a straight line before changing direction (Ridley *et al.*, 2003). We hypothesised that endowing our *in silico* cells with some persistence in their movement might enhance the effectiveness of the cell

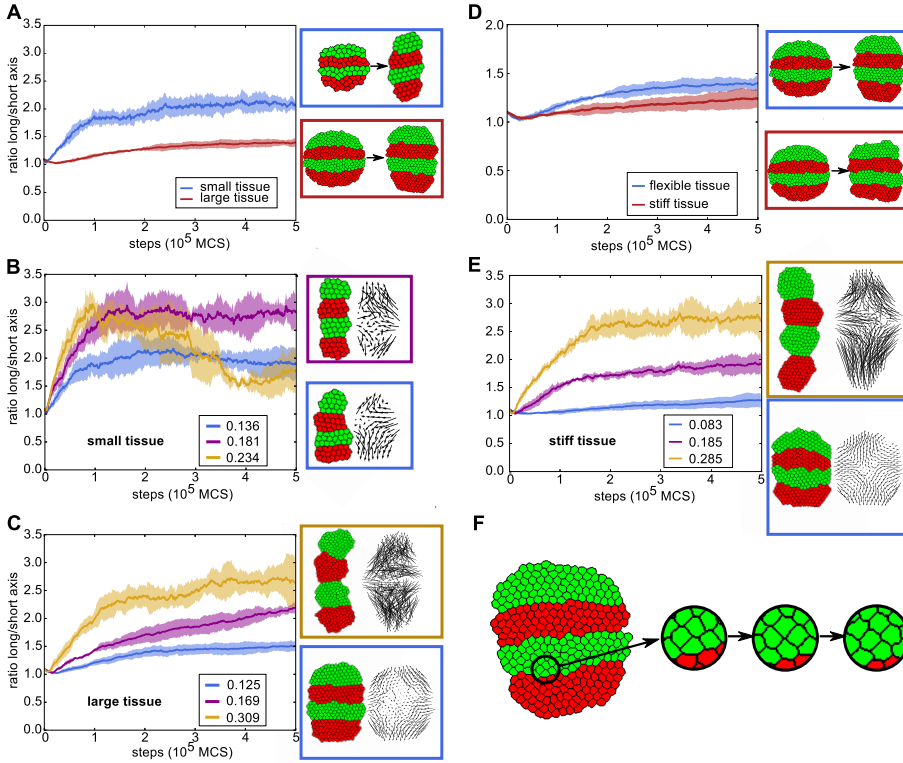


Figure 5.4. Convergent extension through segment-specific adhesion is enhanced by a persistent migration mechanism. (A) The ratio of the length of the long and short axis over simulation steps (MCS) without a persistence mechanism for a small and a large tissue (4x more cells). The large tissue (red curve) never reaches the same ratio as the small tissue. Next to the graph are examples of the state of both small and large tissue at the beginning and end of the simulation. Note that the cells in the small tissue actually are of the same size as those in the large tissue, but that the images are resized. J values: $J_{c,m} = 12$, $J_{r,g} = 22$, $J_{r,r} = 16$, $\gamma_{r,g} = 6$ (B,C) The long/short-axis ratio for the small tissue (B) or large tissue (C) with different cell speeds (indicated in the legends) resulting from different strengths of the persistence mechanism. Curves are averaged over 5 simulations, shading indicates standard deviation. Note that the speeds are emergent quantities measured from simulation output, and have dimension (lattice sites)/MCS (where the lattice sites have arbitrary length). Next to the graph are examples of the state at the end of the simulation (also with vector plots of cell displacement). The colour of the box indicates the parameter set the example belongs to. Vector plots were corrected for whole-tissue rotation, and in the plots for the large tissues, 60% of vectors were sampled for greater clarity. (B) Blue curve: $\mu = 0.5$, $s = 5$; purple curve: $\mu = 2.0$, $s = 5$; yellow curve: $\mu = 2.0$, $s = 10$. (C) Blue curve: $\mu = 0.5$, $s = 5$; purple curve: $\mu = 1.0$, $s = 10$; yellow curve: $\mu = 1.5$, $s = 30$. (D) The ratio of the length of the long and short axis for simulations with stiffer tissue, compared to the previous settings (large tissue). J values for stiffer tissue: $J_{c,m} = 24$, $J_{r,g} = 44$, $J_{r,r} = 32$, $\gamma_{r,g} = 12$. (E) Improvement of extension of the stiff tissue in the presence of persistence. Blue curve: $\mu = 0.5$, $s = 5$; purple curve: $\mu = 2.0$, $s = 40$; yellow curve: $\mu = 2.5$, $s = 40$. (F) Example of T1 transition in a tissue with intermediate strength persistence. $\mu = 2.5$, $s = 20$, speed 0.150, final axis extension 1.6x

motion resulting from segment-specific adhesion. Therefore, we implemented a simple persistence mechanism which has been used before in CPM for migrating lymphocytes (Beltman *et al.*, 2007) (see section “the model” and Methods for details). Note that we did not impose a tissue-level bias on the direction of persistence beforehand to favour convergent extension: the cells started each with their own random target direction.

Endowing cells with a limited tendency for persistence slightly increased the speed of cell displacement, yielding more rapid convergent extension and a more elongated tissue shape at the end of the simulation (Fig. 5.4B, C). In the large tissue, further increasing the level of persistence allowed the tissue to reach almost the same axis ratio as the small tissue without persistence in a comparable amount of simulation steps (Fig. 5.4C). The smaller tissue also gained extension speed and a larger axis ratio from increased cell speeds; however, because the tissue already extended quite rapidly, the contribution of persistence was substantially smaller (Fig. 5.4B). From the vector plots it can be seen that the overall cell displacement pattern still generated the typical convergent extension pattern.

Note that without differential adhesion between segments (S3 Figure), persistent cell motion only mixed up the segmentation pattern without yielding any tissue extension. This indicates that segment-specific adhesion provided the directional signal for axis extension; aligning the initially random direction of persistent cell motion and thus allowing it to enhance tissue extension.

When the strength of the persistence mechanism was strongly increased, the probability of segments merging suddenly increased (Fig. 5.4B, yellow curve). Interestingly, the large tissue seemed capable of sustaining larger cell speeds before segment collapse occurred. In both cases this extreme behaviour only occurred for rather strong persistence, while in a large part of the parameter space convergent extension was significantly enhanced without the risk of tissue collapse (see also S4 Figure).

Next, we tested the efficacy of segment-specific adhesion in stiffer, more epithelium-like tissues. For this, we doubled the J values, which reduces the amount of membrane fluctuations; as can be seen in Fig. 5.4D and F, the cells move considerably less and have a more distinct hexagonal shape than in the more flexible mesenchyme-like tissue we studied earlier. In these stiff tissues, segment-specific adhesion alone generates hardly any tissue extension, because the higher J values present an energy barrier to tissue shape change, much like tight junctions (Fig. 5.4D). However, combined with increasing levels of persistence, beyond the range of parameters used before, significant tissue extension arose (Fig. 5.4E). Thus, in a stiff, tightly connected tissue an active cell motility process is required to drive convergent extension, while differential adhesion still provides

the directional signal. Interestingly, for intermediate persistence levels cells maintain the hexagonal shape typical of stiffer epithelial tissues and T1 transitions can be frequently observed throughout the extension process (Fig. 5.4F). In contrast, for the highest persistence levels tested, cells display considerable more membrane fluctuations, causing them to lose the hexagonal shape imposed by the higher tissue tension. Therefore, under these settings the tissue can no longer be considered as epithelium-like.

5.2.5 Extensions to the model

The above simulations were all done with an unconstrained, fully segmented tissue. To further examine the relevance of differential adhesion as a driver of convergent extension in real-life morphogenesis, we modified our simulations in two ways corresponding to observed *in vivo* conditions: with a constrained anterior end and with a gradual instead of discretely segmented differential adhesion pattern (Fig. 5.5).

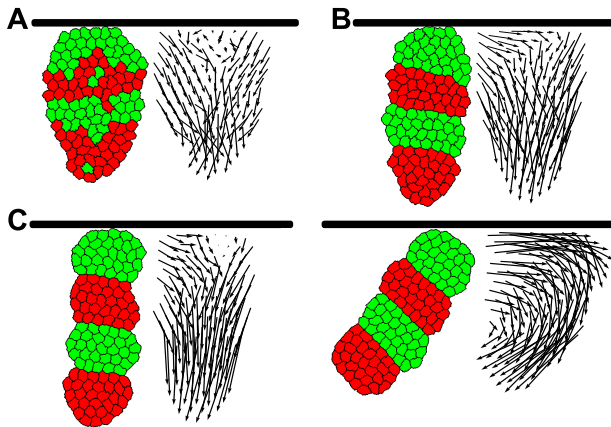


Figure 5.5. Extension to the model: a blocked anterior end does not alter results. (A,B) Simulations with graded adhesion, without (A) or with (B) segment-specific adhesion. The black bar represents the top boundary of the field (not an actual simulated object). 5 simulations were run over 500,000 MCS with parameters as in Fig. 5.2, representative final states are displayed. (C) Two simulations with only segment-specific adhesion, ($J_{c,m} = 10$, $J_{r,g} = 18$, $J_{r,r} = 8$, $\gamma_{c,m} = 6$ and $\gamma_{r,g} = 10$), run for 1,000,000 MCS as in Fig. 5.3. Note how the tissue pushed away from the boundary in the right figure; the vector field here is not corrected for tissue rotation.

In many cases, the tissue undergoing convergent extension is attached on one or more sides to adjacent tissue, and is therefore restricted in its movements in that direction. For instance, in *Tribolium* the converging tissue is attached to the head, which moves very little and does not change shape. We tested the influence of such a restriction on convergent extension and cell mixing by placing the anterior end of the tissue against

the border of the field, to constrain tissue movement at the anterior tissue boundary. We then applied the explicit graded adhesion mechanism (as in Fig. 5.2), and observed that cell mixing still occurred in the absence of segment-specific adhesion (Fig. 5.5A). Note that the anterior end of the tissue converged less because the tissue could not extend in the anterior direction, which becomes obvious in the vector plot of Fig. 5.5A where all arrows point either inward or to the posterior. This caused the tissue to become a bit ‘carrot-shaped’, which is indeed typical for extending tissues attached to non-extending tissues (see e.g., *Tribolium*). Again, segment-specific adhesion prevented mixing at the segment boundaries (Fig. 5.5B), and was by itself able to drive convergent extension (Fig. 5.5C) for larger surface tensions. Note that for strong segment-specific adhesion, the tissue tended to rotate and push away from the boundary to escape the restriction (see also vector plot), allowing it to elongate more in the same amount of simulation steps. This is an artefact of the way we modelled the restriction as only an impenetrable boundary into which no extension can occur; had the extending tissue also been attached to this boundary it would likely rotate less.

Convergent extension also occurs in non-segmented tissues. In *Xenopus* it was shown that when cells from the axial mesoderm were mixed, they quickly sorted out according to their original position on the antero-posterior axis, implicating that a position-based differentiation gradient rather than discrete segments yielded differential adhesion (Ninomiya *et al.*, 2004). Strikingly, the amount of convergent extension occurring depended on the degree of sorting out that had already occurred. This suggested that differential adhesion, besides getting and keeping cells at the right position, also played an instructional role in convergent extension. Furthermore, it was found that the differential adhesion mechanism acted both upstream of and in parallel to the PCP pathway to drive convergent extension (Ninomiya *et al.*, 2004).

Here, we tested whether a gradient in adhesion proteins could cause correct anterior-posterior sorting and whether it could bring about convergent extension in a similar manner to the *in silico* segmented tissue. A tissue with graded expression of a single protein would not display anterior-posterior, but radial cell sorting without any convergent extension, according to both experiments and computational models (Ninomiya *et al.*, 2012, Zhang *et al.*, 2011). We therefore generated a tissue with two adhesion proteins that formed opposite gradients. This meant that a cell with a high concentration of protein A had a low concentration of protein B and vice versa (S5 Figure, see Methods for details). Cells with high A adhere more strongly to other cells with high A (and vice versa). Furthermore, cells with intermediate concentrations of both proteins adhere more strongly to each other than to cells with a high concentration of just one protein (S5 Figure, explanation in Methods).

When cells were placed randomly in the tissue (as in the experiment with mixed tissue), they sorted out with cells with similar protein concentrations clustering together. However, tissues in which cells had no persistence sorted out only partially: they became stuck in local optima where multiple clusters of similar protein concentrations were present, which was also observed for large tissues with a single protein gradient (Zhang *et al.*, 2011) (S5 Figure). The partially sorted state was reached more quickly when the gradients of A and B concentrations were steeper, although this still did not lead to complete sorting. The tissue did sort completely when cells were endowed with high persistence, creating a rather turbulent tissue which could sort quite rapidly, with high-A cells on one end and high-B cells on the other (Fig. 5.6A).

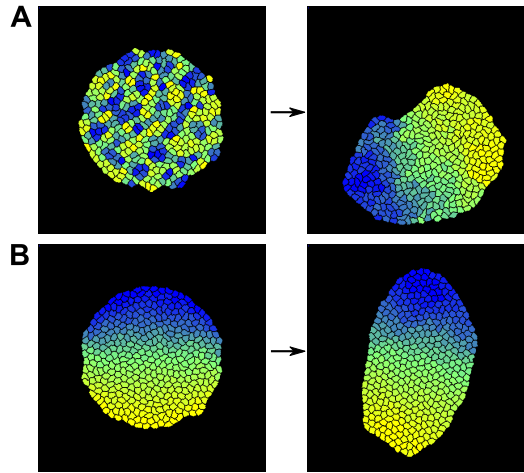


Figure 5.6. Extensions to the model: two opposing adhesion gradients lead to a-p sorting and convergent extension. (A) Random initial configuration, then run over 2,000,000 MCS. Maximum difference in adhesion strength between like and unlike cells, $mm=18$. $J_{c,m} = 15$, maximum $J_{i,j} = 28$. Persistence mechanism at high strength, $\mu = 2.0$, $s = 30$. (B) Simulation with two opposing gradients of adhesion proteins, sorted initial configuration (2,000,000 MCS). Persistence mechanism at low strength, $\mu = 0.5$, $s = 5$.

When the simulation started with a tissue in which cells were already sorted, it elongated, with the extent of elongation depending on the maximum difference in adhesion (Fig. 5.6B, S6 Figure). Modest persistence could enhance this process (S6 Figure), but strong persistence reduced the extension again (see the fully sorted, but unelongated tissues in S5 Figure). Therefore, if extension should follow after sorting of a fully mixed tissue, cell motility needs to be regulated together with the degree of sorting. However, in naturally occurring situations, AP patterning occurs prior to convergent extension, so complete mixing and hence the need for complete sorting are unlikely to occur. Rather, robustness to developmental noise will require limited sorting to optimise AP patterning, for which lower persistence levels are sufficient.

Thus, our results show that besides a segmented tissue pattern, graded distributions of adhesion proteins are also capable of driving a modest form of convergent extension.

5.3 Discussion

Interplay of tissue patterning and convergent extension

During formation of the bilaterian body axis, cells converge and intercalate to form a tissue that is longer and narrower. Convergent extension usually occurs in tissues which have undergone prior gene expression patterning such that cells have distinct fates at different positions in the tissue. Arguably, convergent extension, which often causes extensive cell rearrangements, should be tightly regulated to prevent it from interfering with this tissue pattern. An example where this is relevant is *Tribolium*, in which convergent extension follows shortly after segmentation (Sarrazin *et al.*, 2012). Paradoxically, it has been shown in both *Drosophila* and *Xenopus* that a segmented or other antero-posterior tissue pattern is required for convergent extension (Irvine and Wieschaus, 1994, Ninomiya *et al.*, 2004, Zallen and Wieschaus, 2004) suggesting that it is instructive for rather than compromised by tissue remodelling. It is therefore important to know how convergent extension may interact with a prepatterned tissue.

Here, we investigated the potential role of segment-specific adhesion in convergent extension of a fully segmented tissue. We applied two mechanisms -graded adhesion and axial adhesion- that caused convergent extension of the tissue. We demonstrated that without segment-specific adhesion, these mechanisms disturbed the segmented tissue pattern. Adding segment-specific adhesion in our model did not only preserve the segments, but also enhanced the extension of the long tissue axis. Furthermore, segment-specific adhesion by itself was sufficient for convergent extension both in unconstrained and constrained tissue, and can be combined with persistence to enhance extension in larger and stiffer tissue. Finally, we have shown that this differential-adhesion based mechanism also extends to non-segmented tissues with opposite gradients of adhesion proteins, although the amount of extension is more modest.

Source of directionality of convergent extension

An important question concerning convergent extension is where the directional signal for the orientation of tissue extension comes from. A number of earlier models was constructed to elucidate the various mechanisms behind convergent extension through cell intercalation in different organisms (Backes *et al.*, 2009, Brodland and Veldhuis, 2012, Honda *et al.*, 2008, Rauzi *et al.*, 2008, Weliky *et al.*, 1991, Zajac *et al.*, 2003). Most of these predefined a direction of extension either by biasing protrusions or constrictions of the cell membrane ((Brodland and Veldhuis, 2012, Brodland, 2006, Honda *et al.*, 2008,

Rauzi *et al.*, 2008)), or including a boundary which restricts cell motion in certain directions (Weliky *et al.*, 1991). Only two models did not impose such a direction. In the model by Backes *et al.* (Backes *et al.*, 2009), a positive tension between two cell types instructed the intercalation direction of forcibly elongated cells, and led to a direction of extension and narrowing which was perpendicular to that in our model. This mechanism only worked for tissues which were already quite narrow, and generated very little actual tissue extension. In the original version of the axial adhesion mechanism, constructed by Zajac *et al.* (anisotropic differential adhesion, (Zajac *et al.*, 2003)), the adhesion polarity of cells was not fixed, but rather depended on the orientation of the cell long axis (cells were forced to be elongated). In this case, the direction of tissue extension was not pre-defined, but emerged through alignment of the elongated cells. As a consequence, the direction of axis extension was random and differed between simulations. Finally, Shinbrot *et al.* (Shinbrot *et al.*, 2009) demonstrated that cell-cell adhesion and repulsion can generate segmented and elongated tissue patterns from random initial cell configurations. Rather than through convergent extension, the elongated and segmented patterns in their simulations form from cells condensing from a dispersed state while sorting into disks, with the tissue assuming a random orientation with respect to the field axes.

In this paper, we started out with two superimposed mechanisms for convergent extension, in which the direction of extension was also superimposed. One had an explicit gradient defining the position of the extending axis (graded adhesion), while the other imposed an internal, fixed polarity on the cells (axial adhesion), thus implicitly assuming the presence of some kind of signalling gradient. Interestingly, when segment-specific adhesion drove convergent extension alone, directionality emerged without such an imposed signalling gradient or polarity. Instead, the interface between segments provided enough information to allow the tissue to stretch in the direction perpendicular to it. The ability of segment-specific adhesion to provide the extension direction was further emphasised when we combined it with the persistence mechanism, which by itself could not produce convergent extension, but could speed up tissue extension considerably when combined with segment-specific adhesion. Therefore, to our knowledge, segment-specific adhesion is the first convergent extension mechanism which yields a predictable direction of convergent extension without imposing polarity on the cellular or tissue level.

Mechanism of convergent extension by differential adhesion

In our model, the degree of tissue extension by segment-specific adhesion was determined by the balance between surface tension between red and green segments, and the tension of the tissue with the surrounding medium. The red-green surface tension provided the elongating force by reducing the contact surface between the segments (pulling the segment interface inward), whereas the surface tension with the medium opposed this force by making the tissue as a whole as round as possible (pulling the segment interface

outward). This agrees well with findings in *Xenopus*, where the axial mesoderm needs to be enveloped in epithelium in order to extend (Ninomiya and Winklbauer, 2008). Without the epithelial layer, the surface tension of the mesoderm with the environment is too high, resulting in a spherical tissue.

Because differential adhesion minimises the contact area between tissues of different types, the initial ratio between segment width and length is another factor influencing the extent of convergent extension in our simulations. The smaller the initial ratio between segment width and length, the larger the contraction of the contact between segments, and the more extreme the resulting tissue elongation will be (compare Fig. 5.3 with ratio 3/10, to S2 Figure with ratio 2/15). As segments are typically organised perpendicular to the length axis of the tissue, the initial segment width corresponds to the tissue width before convergent extension, while the initial segment length corresponds to the tissue length before convergent extension divided by the number of segments. The segment width length ratios used in our simulations are well within the naturally occurring ranges when considering the segment numbers and tissue widths and lengths observed in for example *Tribolium*, *Drosophila* and other arthropods.

Limitations

We observed an apparent limit to the extent in which segment-specific adhesion can drive convergent extension. When differential adhesion tensions or persistence levels exceeded a certain threshold, the segments started to rotate and merged with other segments of the same type, thus further minimising intersegment boundary surface. However, we suggest that this may largely be an artefact of our simplified 2D model; the risk of tissue bending may be much lower for a 3D tissue, and/or if the tissue is also embedded in other tissues (as in *Xenopus*) that restrict its movements and aid convergent extension at the same time. Furthermore, the phenomenon did not occur for most of the parameter region we tested, and we obtained strong tissue elongation within the biologically relevant region. In addition, for persistence it is reasonable to expect that once convergent extension has completed cell motility is downregulated again as part of the further progression of the development program (note that persistence is not required for maintenance of tissue extension). This termination of persistence provides an additional safeguard against segment fusion.

In the current study we have shown that in a coherent, fully presegmented tissue, segment-specific differential adhesion is a suitable candidate mechanism both for maintaining segment integrity and driving convergent extension. We did not take into account other processes that may take place at the same time as convergent extension. For instance in *Tribolium* and other short-germ insects, the segments are laid down sequentially instead of simultaneously, from a growth zone where cell division provides a steady

source of undifferentiated tissue. It appears that in this case, convergent extension occurs shortly after a new segment is laid down (Choe *et al.*, 2006, Sarrazin *et al.*, 2012). Based on preliminary results, we expect that segment-specific adhesion will also suffice to drive convergent extension during sequential segmentation, but given the complexity of the growth zone and segment-definition dynamics, it is beyond the scope of this article to investigate this.

Furthermore, we assumed for the sake of simplicity that a cell's adhesion is a fixed property. However, we recognise that this may not always be the case, for instance when cells change the concentration of adhesion molecules on their membrane in response to interactions with other cells that possess different (concentrations of) adhesion molecules (see (Ninomiya *et al.*, 2012)). This may influence the ability of differential adhesion to drive convergent extension.

Comparison to experiments

We found that for loosely connected, mesenchyme-like tissues differential adhesion alone or combined with a limited persistence of motion can drive convergent extension. As such, we expect differential adhesion to contribute to axial extension in organisms such as *Xenopus* in which an antero-posterior pattern is present, and which indeed served as one of the inspirational starting points for this study. Possibly this mechanism also plays a role in short-germ insects such as *Tribolium*, which undergoes convergent extension simultaneously with segmentation, if the tissue emanating from the growth zone is indeed flexible enough.

For stiffer tissues we found that in order to obtain substantial tissue extension, segment-specific adhesion needs to be combined with a significant level of persistent cell motion. Notably, the persistence alone would not produce any convergent extension, but requires differential adhesion to instruct and coordinate cell movement. Furthermore, the strength of persistence required for proper extension was so low that inspection by eye would most likely not reveal the presence of this mechanism in *in vivo* tissues, as the cell displacement is similar to that of tissues where cells are not persistent. Persistence strong enough to be visible led to turbulence and segment merging, and would require other, more global directional cues than segment-specific adhesion to yield convergent extension.

Although clearly not a one-to-one match, persistence bears intriguing similarities to the case of the *Drosophila* germband. Here, parasegmental actomyosin barriers prevent intersegmental cell mixing (Monier *et al.*, 2010, 2011), while the segments also serve as a directional signal for planar cell polarity, which subsequently instructs the anisotropically directed actomyosin contractions that drive the T1 transitions underlying convergent

extension (Bertet *et al.*, 2004, Rauzi *et al.*, 2008, Zallen and Wieschaus, 2004). The similarities reside in the fact that the segmental pattern instructs the direction of cell movement, and that cell movement requires active cytoskeletal remodelling. It remains to be established whether segment-specific adhesion can act in combination with and thereby enhance the mechanisms observed in *Drosophila*, or whether it may act as an alternative strategy deployed in other organisms.

Unfortunately, the similarities between the differential adhesion mechanism and the *Drosophila* type mechanism make the design of an experimental setup to discriminate against these two possibilities highly non-trivial. As an example, if we experimentally disrupt genetic factors regulating segmentation these will not only hamper segment-specific adhesion, but also the aforementioned planar polarity such as occurs in *Drosophila*. As a consequence results would be inconclusive. Likewise, active cytoskeletal dynamics are involved both in convergent extension driven by the combination of differential adhesion and persistence and in planar-polarised junctional tension driven convergent extension. Thus, failure of convergent extension upon actomyosin disturbance will again be inconclusive. Similarly, although one could try to experimentally increase the adhesiveness of the whole tissue by ubiquitously expressing e.g. N-cadherin; this would certainly hinder convergent extension via segment-specific adhesion, but unfortunately is likely to also hinder other convergent extension mechanisms by increasing the energy required to break the bonds between cells. This problem of distinguishing between the two mechanisms is further aggravated by the fact that the mechanisms may be likely to work in combination. One experiment that may allow for a distinction between the two convergent extension mechanisms is to apply pulling forces on the tissue in the direction parallel and perpendicular to the segmentation pattern. If less force is required to tear the tissue along segment boundaries than to tear it in the perpendicular direction this is a strong indicator that differential adhesion is involved. Still, this does not allow one to establish the importance of this differential adhesion for convergent extension.

Conclusion

In summary, we have shown that differential adhesion is sufficient to drive convergent extension in presegmented tissues, and represents a convergent extension mechanism not requiring any directional signal. While the investigated convergent extension mechanism may not be universal, in segmented tissues the presence of segmental boundaries is likely to contribute to convergent extension, either via differential adhesion or via alternative mechanisms such as actomyosin bands or planar cell polarity. Likewise, while not all tissue is segmented, anterior posterior patterning may also allow for differential adhesion-based convergent extension. In the current study we focused on the role of a fully presegmented tissue pattern in driving convergent extension. However, in many cases segmentation and convergent extension occur simultaneously.

Therefore, in future work we aim to investigate the dynamic interplay between sequential segmentation and convergent extension. Considering such bidirectional feedback between patterning and morphogenesis may bring to light important principles of coordinating growth and patterning.

5.4 Methods

5.4.1 Model setup

CPM model formalism

We consider convergent extension of monolayer tissues, and therefore we use a simple 2D Cellular Potts Model for all our simulations. In the CPM model formalism, cells consist of multiple lattice sites with 2D coordinates i and j and have a cell type τ and identification number σ . The lattice (“field”) is updated using the Monte Carlo algorithm. For each Monte Carlo step (MCS), lattice sites are drawn randomly, as many times as there are lattice cells. For each site belonging to the boundary of a cell, a random neighbour is selected which may copy its identity into this lattice site. The probability of this event is calculated with the Hamiltonian, which depends on the change in cell surface energy and cell volume that would be caused by the potential copy event. The surface energy of lattice sites at the cell boundary, $J_{\tau(\sigma_{ij}),\tau(\sigma_{i'j'})}$, depends on the type of the cell ($\tau(\sigma_{ij})$) and that of the neighbour that the lattice site contacts ($\tau(\sigma_{i'j'})$). Cells are assumed to minimise their surface energy while at the same time trying to maintain their volume (or in 2D, area) at a target value A_σ . The Hamiltonian is then as follows:

$$H = \sum_{ij} \sum_{i'j'} J_{\tau(\sigma_{ij}),\tau(\sigma_{i'j'})} (1 - \delta_{\sigma_{ij},\sigma_{i'j'}}) + \sum_{\sigma} \lambda_a (a_\sigma - A_\sigma)^2$$

The first term represents the sum of all surface energies J , where δ is the Kronecker delta (0 if σ_{ij} and $\sigma_{i'j'}$ are different, and 1 if they are equal). $\sigma_{i'j'}$ sums over all 8 neighbouring sites in the 3×3 neighbourhood. The second term serves to keep the actual area a of a cell close to the target area A , where λ_a is the resistance of cells against volume changes. The probability that a neighbouring site extends into the lattice site under consideration is 1 if $\Delta H < 0$, and $e^{-(\Delta H)/T}$ otherwise, where ΔH is the change in the Hamiltonian due to the considered modification, and T is the simulation temperature, determining the membrane fluctuation amplitude of cells. The model was implemented using the C++ programming language.

J values and adhesion

As mentioned before, the J values represent the surface energies at the interface of cells and their surroundings. The higher the J value, the less favourable the interaction, and

therefore a pixel copy which reduces such an interface will be rather likely. Therefore, mechanisms which increase adhesion reduce the J value of that cell-cell interaction. J values act in a relative manner. If the J value of a cell-medium interaction is lower than that of a cell-cell interaction, cells will disperse through the medium rather than form a coherent tissue. Likewise, if the J value between red and green cells is higher than both the J value of red-red and green-green interactions, red and green cells will form a checkerboard pattern (see (Glazier and Graner, 1993) for more details). The behaviour of cells under a certain set of J values can be more easily seen from the surface tensions, which are calculated from the J values as follows:

$$\gamma_{i,j} = J_{i,j} - \frac{J_{i,i} + J_{j,j}}{2}$$

where i and j represent different cell types. Now, a positive γ value for the interaction between red and green cells keeps them separated (while a negative one causes mixing), and the same is true for the γ value of cell and medium interactions.

Simulation initialisation

Cells are initiated as blocks of 5×5 lattice sites, put closely together in a grid-like manner in the centre of the field. At this point, we already divide the *in silico* tissue into regular segments of alternating cell types. In the simulations of Fig. 5.2 and Fig. 5.3, each segment is 3 cells thick and 10 cells wide upon initialisation. In the simulations for Fig. 5.4, the segments were each 2×10 cells for the small tissues, and 4×20 for the large tissues.

5.4.2 Explicit convergent extension mechanisms

Two alternative mechanisms for convergent extension were used: graded adhesion and axial adhesion. To implement such a mechanism, ΔH is modified for certain lattice sites to establish a bias in copy probability.

Graded adhesion

For the graded adhesion mechanism, the J value of two adjacent cells is modified depending on their location within the field. We defined a Gaussian function with a maximum at the centre of the field in the x -direction (Fig. 5.1A), but homogeneous in the y -direction.

$$J' = J - w * e^{-\frac{(x-b)^2}{2*c^2}}$$

where w is the maximum amplitude of the modification, b is the centre of the x -axis of the field, and c is the standard deviation of the Gaussian. As a consequence, J values

become gradually lower towards the centre of the field.

Axial adhesion

For the axial adhesion mechanism, cells have an increased adhesion on the upper and lower faces of the cell (Fig. 5.1B). This was incorporated using the following modification of the J values:

$$J' = J - \beta^2 * |\sin(\alpha)| * |\sin(\alpha')|$$

where β represents the maximum reduction of surface energy because of the mechanism, and α and α' are angles with the horizontal (x-) axis. α is the angle of the vector pointing from the centre of cell σ_{ij} to the membrane segment where the copy takes place, and α' is the angle of the vector in the neighbouring cell ($\sigma_{i'j'}$). By taking the sine, we ensure that the modification of J is highest for membrane segments both at the top and bottom of the cell (so with vectors along the y-axis), correcting with the absolute value for the fact that angles larger than π yield negative sines. See Zajac *et al.* (Zajac *et al.*, 2003) for the original version.

Persistent motion

Persistent motion is implemented by enhancing the probability of extension in the direction of previous movement. In practice, this means that cells have a target direction, and extensions that have a small angle with this direction occur with a higher probability than extensions with a large angle to the target. The target direction is updated every s MCS to the direction of the actual previous displacement of the cell. Persistence was incorporated by extending ΔH as follows:

$$\Delta H' = \Delta H - \mu \cos(\zeta)$$

where μ is the strength of persistence, and ζ is the angle between the target direction and the displacement vector under consideration (i.e., the vector given by the mean position of the cell and the coordinates of the position to be modified). Both larger μ and larger update times contribute to higher persistence (time during which the cell moves in a more or less straight line) and larger cell speed. See also Fig. 5.1C and (Beltman *et al.*, 2007).

Opposing adhesion protein gradients

Cells have two adhesion proteins A and B, whose concentrations together add up to 1.0 (arbitrary units). So $A = 1.0 - B$. For the sorting simulations, cells are assigned a random concentration of B between 0 and 1. Subsequently, their concentration for A follows from $A = 1.0 - B$. In the convergent extension simulations, where the cells are already sorted, the concentration of B is increased stepwise from 0 to 1 in every

subsequent row. The protein concentrations are translated to a J value between two cells i and j as follows:

$$J_{i,j} = J_{i,j} - mm * (\min(A_i, A_j) + \min(B_i, B_j))$$

where A_i is the concentration in cell i, min compares the concentrations of cell i and j and takes the minimum, and mm is the maximum reduction in J value (when cells have identical protein concentrations). MM therefore determines the ‘steepness’ of the adhesion gradients. The cell with the lowest concentration of a protein dictates the amount of adhesion between two cells via that protein, so a cell with high A still adheres poorly to a cell with low A. The two adhesion proteins do work additively: having a bit of both leads to more adhesion than just having a bit of one; therefore, cells with intermediate concentration adhere more strongly to each other.

5.4.3 Model analysis

Cell displacement vectors

The mean position of each cell is registered every 1000 MCS. The position at time 5000 (to allow for an initialisation period) and the final time point are used to determine the overall displacement vector (using Python scripts). The vector plots are corrected for whole-tissue rotation with respect to the y-axis of the field. For this, the rotation of the long axis of the tissue with respect to the y-axis is registered throughout the simulation.

Correlation of neighbouring cell displacements

The displacement vectors of figure 2 and supplementary figure 1 have been coloured according to their average angle with the neighbouring vectors. A neighbour of a vector has a starting point which lies within a neighbourhood of size 30 (field pixels) around the starting point of the vector. The colour bar is scaled relative to the largest average angle in one of the two simulations to be compared (see figure 2). The colour of the vector is relative to the largest value

Emergent cell displacement speed

The mean positions are also used to calculate cell speeds, which is both an average over the variable speed of a cell and an average between cells. We used customised Perl scripts for the calculations.

5.4.4 Model parameters

The current study serves as a proof of principle, illustrating how convergent extension may disrupt pre-existing tissue patterns, while these pre-existing tissue patterns may also drive convergent extension through differential adhesion. Because of the conceptual nature of our model, we do not aim to quantitatively fit convergent extension dynamics in a particular model organism. However, if this were to be the case, model parameters could be adjusted to obtain cell movement speeds and trajectories matching experimental data. In contrast, in the current study we aim to illustrate that differential adhesion either alone or combined with persistence of motion, represents a feasible new mechanism for convergent extension. As such, we aimed to ensure that differential adhesion driven convergent extension occurs for a wide range of parameters, making it a plausible mechanism in broad range of contexts. For persistence, parameter scaling was done internally: we matched persistence tendencies to membrane fluctuations and overall tissue deformation so that we remained in the domain of biologically realistic behaviour, avoiding the merging of segments or of turbulent tissue dynamics. As a consequence, we applied considerably lower persistence tendencies than in the study by Beltman et al (Beltman *et al.*, 2007), where it was used to simulate migrating lymphocyte dynamics. Note the limited persistence tendencies applied in our studies significantly altered quantitative model behaviour, as shown in Fig. 5.4. Default parameter values are shown in Table 1.

5.5 Acknowledgements

We thank Ramiro Magno for his help with CPM implementation and parametrisation, Joost Beltman for cell speed measurement scripts and Ioana Niculescu for discussions and suggestions for the manuscript.

Table 5.1. Default parameter values

| parameter | value | comments |
|--------------------------------------|--------------------------------------|--|
| <i>common parameters</i> | | |
| field size | 400x400 - 800x800 (lattice sites) | size varied to accommodate tissue even when extended |
| duration | $5 * 10^5$ (MCS) | |
| initialisation period | 500 (MCS) | |
| simulation temperature (T) | 15 | |
| neighbourhood order | 2 | |
| target cell area | 500 or 200 (lattice sites) | smaller cell size used for all simulations with persistence (for computational efficiency) |
| λ_a | 1. | |
| $J_{c,m}$ | 10-22 | ranges used in all simulations except in those with stiff tissue (Fig. 5.4D-F) |
| $J_{r,g}$ | 12-18 | |
| $J_{r,r}$ | 4-18 | |
| | | |
| <i>mechanism-specific parameters</i> | | |
| w | 11 | |
| c | 7 | |
| β | 2.0 | |
| μ | 0.5-2.0 | |
| s | 5-30 (MCS) | |
| mm | 12-24 | |

Where applicable, parameter dimensions are indicated behind the value between parentheses.

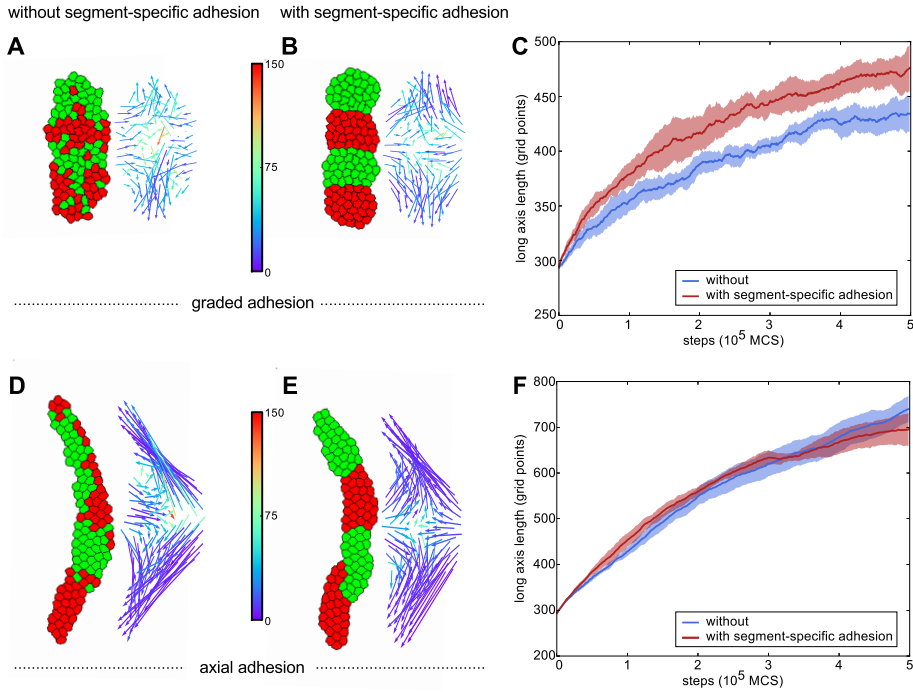


Figure S5.1. Strong graded adhesion and axial adhesion lead to more elongation and more mixing at the segment boundaries. Left images display tissue at the final step of the simulation (at 500,000 MCS). Right images contain the displacement vectors of each cell in the simulation. The tail is located at the start position of the cell, the head at the end. (A,B,C: row 1) Simulations with graded adhesion, strength $w = 12$. (D,E,F: row 2) Simulations with axial adhesion, strength $\beta = 2.66$. (A,D) Simulations without segment-specific adhesion. (B,E) Simulations with segment-specific adhesion ($\gamma_{r,g} = 4$). (C,F) Length of the long axis of the tissue as a function of simulation steps. Blue is without and red is with segment-specific adhesion. The curves are averaged over 5 runs of the model, shading indicates standard deviation. Note that the added effect of segment-specific adhesion on axis extension is smaller here than when the convergent extension mechanisms are weaker (compare to figure 2 in main article).

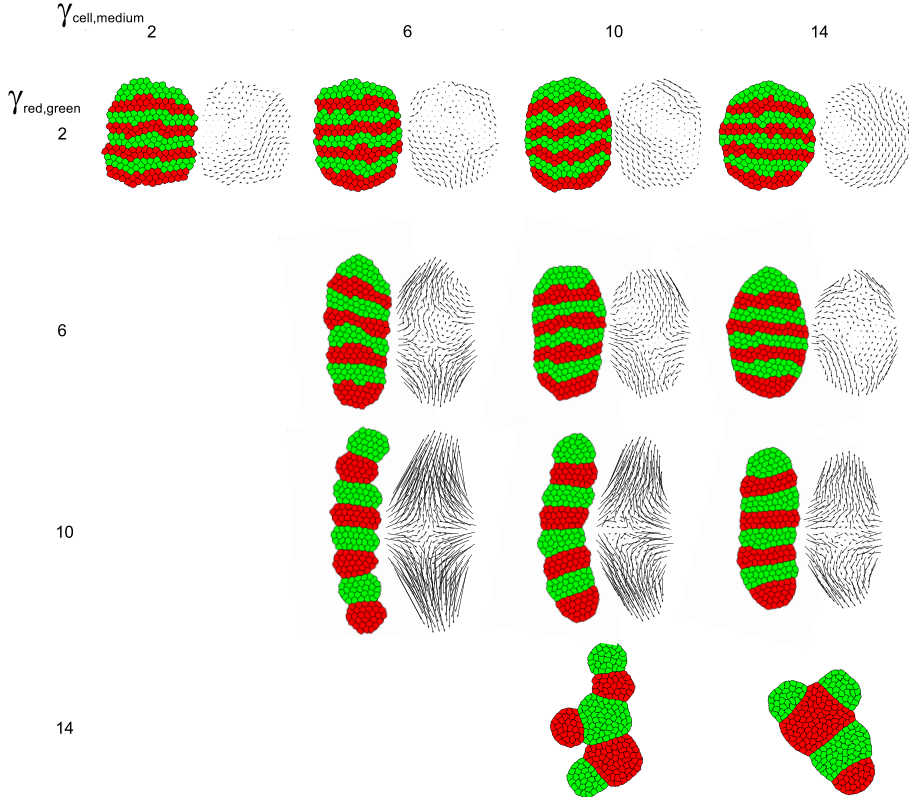


Figure S5.2. Segment-specific adhesion leads to greater extension with more and narrower segments. Parameter space of a tissue of eight segments with varying values for $\gamma_{c,m}$ and $\gamma_{r,g}$, same as figure 2 (See Supplementary Table 1 for J values). Initial segment width:2, length:15 cells. For each set of parameters, 10 simulations were run over 100,000 MCS, representative final states are displayed. In the following simulations we observed merging of segments [(row, col), # out of 10 sims]: (10,10) 1; (10,14) 2; (14,10) 10; (14,14) 10. Vector plots were corrected for whole-tissue rotation.

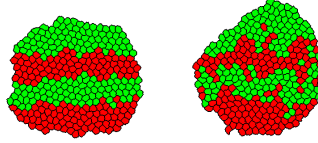


Figure S5.3. Without segment-specific adhesion, persistence does not lead to convergent extension. On the left: a tissue without segment-specific adhesion, and no persistence mechanism. On the right: a tissue without segment-specific adhesion, having a persistence mechanism with $\mu = 2.0$, $s = 10$, leading to an average cell speed of 0.181 (lattice sites/MCS). J values are $J_{c,m} = 12$, $J_{c,c} = 18$

Table S5.1. J values of the parameter space of figure 3 and supplementary figure 2.
The order of the values is $J_{c,m}$, $J_{r,g}$, $J_{r,r}$.

| $\gamma_{c,m} \rightarrow$ $\gamma_{r,g}$ \downarrow | 2 | 6 | 10 | 14 |
|--|----------|----------|----------|----------|
| 2 | 10,18,16 | 14,18,16 | 18,18,16 | 22,18,16 |
| 6 | | 12,18,12 | 16,18,12 | 20,18,12 |
| 10 | | 10,18,8 | 14,18,8 | 18,18,8 |
| 14 | | | 12,18,4 | 16,18,4 |

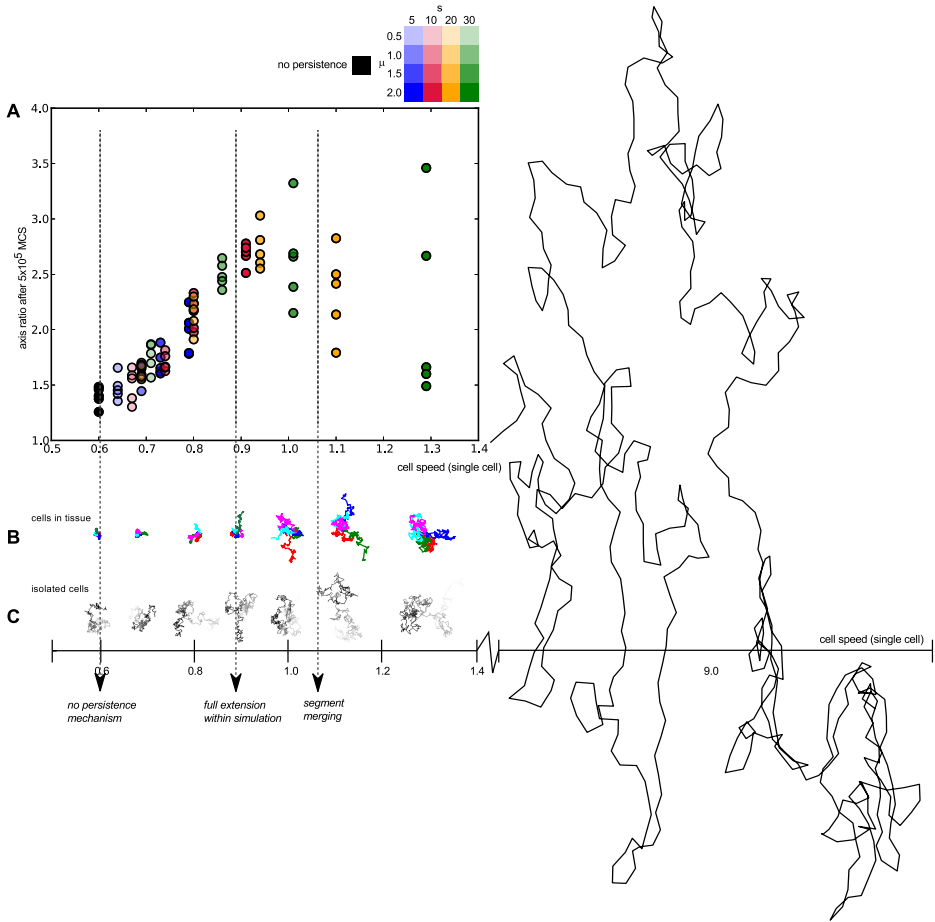


Figure S5.4. Influence of persistence on cell and tissue dynamics. (A) Influence of persistence on tissue elongation. Results are shown for simulations with varying persistence parameters ($\mu = 0 - 2.0$, $s = 5 - 30$) with the long/short axis ratio at the end of the simulation (duration 5×10^5 MCS) plotted against the measured average cell speed of a single isolated cell with those parameters (lattice sites/MCS). J values are $J_{c,m} = 12$, $J_{r,g} = 22$, $J_{r,r} = 16$, $\gamma_{r,g} = 6$. (B) For a subset of the persistence levels in A, cell tracks from 5 random cells part of the same extending tissue are shown (1 of the 5 simulated tissues shown in A; parameters correspond to the following cell speeds (single, tissue): (0.60,0.117), (0.69,0.137), (0.80,0.169), (0.91,0.211), (1.01,0.309), (1.10,0.343), (1.29, 0.501)). The tracks are measured over 100 000 MCS, with the start of each track shifted to the centre. Different tracks are depicted with different colors. (C) For the same subset of persistence levels as shown in B, cell tracks of single-cell simulations (100 000 MCS) are shown. The tracks become lighter with age to indicate directionality. The right-most cell track is of a single cell with strong, lymphocyte-like persistence ($\mu = 16$, $s = 8$), parameters are as in Vroomans *et al.*, PLoS Comp. Biol. 2011. Note the qualitative difference: the cell turns less often, and has more straight stretches (field size 2000x2000). N.B. Track does not become lighter with age.

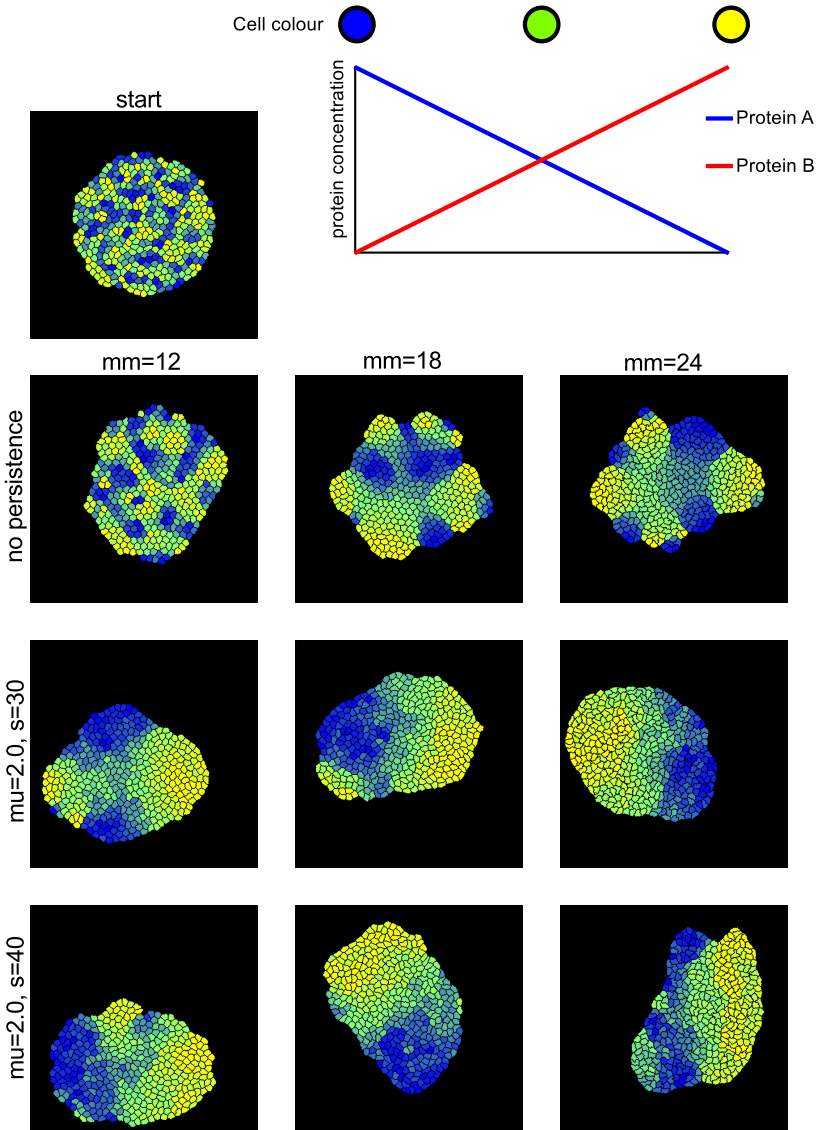


Figure S5.5. Opposing adhesion gradients lead to (partial) sorting out of tissue. The graph shows how the adhesion proteins are distributed in the tissue, and the corresponding cell colour. The images show the tissue at the end of the simulation (2,000,000 MCS) for varying strengths of the maximum adhesion difference mm , without or with persistence (parameters μ : 2.0 and s : 30-40). $J_{c,m} = 15$, maximum $J_{i,j} = 28$.

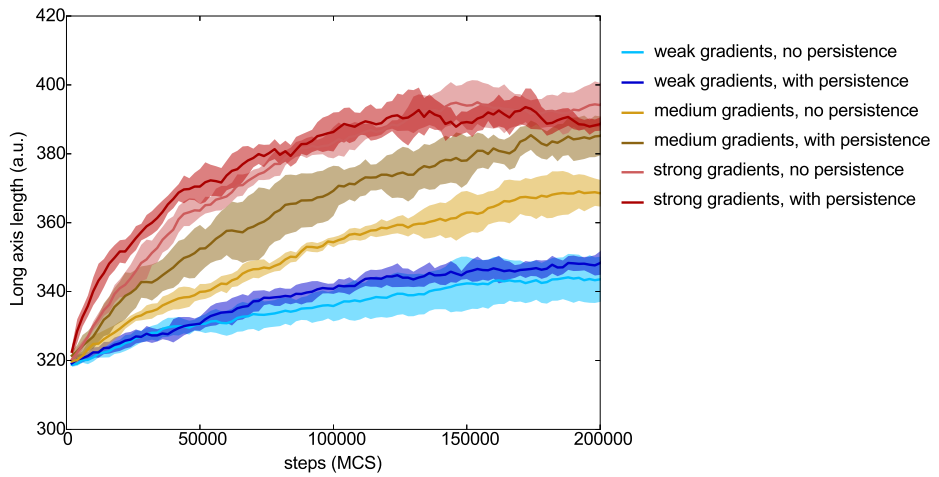


Figure S5.6. Oposing adhesion gradients lead to modest tissue extension. The graph plots the length of the long axis of the tissue over simulation steps for varying values of the maximum adhesion strength (mm: 12, 18, 24), and without or with persistence mechanism (parameters μ : 1.0 and s : 10). $J_{c,m} = 15$, maximum $J_{i,j} = 28$.

5.6 Supplementary information

5.6.1 On surface tensions and segment interfaces

In the differential adhesion hypothesis, tissues are assumed to behave like immiscible fluids. When two drops of liquid have a positive interfacial tension, they reduce the contact area with each other; there is a force pulling inwards on the interface and the drops will round up. For a drop of liquid l on a surface s and in contact with the environment/air/medium e , Young's equation tells us that when the equilibrium state (minimum free energy) is reached, the contact angle θ of the drop with said surface is given by the surface tensions involved (see figure A below, and e.g. Wikipedia - contact angle):

$$\cos \theta = \frac{\gamma_{s,e} - \gamma_{s,l}}{\gamma_{l,e}}$$

Without gravity, a 2D droplet will form a circle cut off by the surface, which means that from this angle and the total area of the droplet, we can calculate the amount of equilibrium contact between the droplet and the surface. For this, we use the fact that the droplet is a segment of the circle for $\theta < \frac{1}{2}\pi$, and that it is a circle with a segment missing for $\theta > \frac{1}{2}$. The area of a segment is given by $\frac{1}{2}(\alpha - \sin \alpha)r^2$, where α is the angle of the sector enclosing the segment. See e.g. www.mathsisfun.com/geometry/circle-sector-segment.html.

Please note that in our case, the area and the surface tensions are in arbitrary (model), not physical units, and we mainly use this to deduce droplet shapes and ratios. Below, r is the radius of the circle, A the area, L the amount of contact:

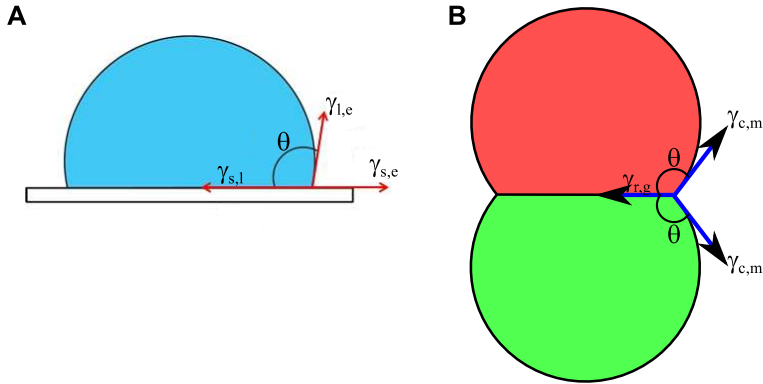
$$r = \sqrt{\frac{2A}{2\theta - \sin 2\theta}}$$

$$L = 2r \sin \theta$$

This can be extended to the situation where there are two drops of liquid adjacent to each other, as is the case in our model (figure b). In our calculations we assume that these two liquids have the same surface tension with the medium $\gamma_{c,m}$ and a (positive) tension with each other $\gamma_{r,g}$. The angle then is given by:

$$\cos \theta = \frac{-\gamma_{r,g}}{2\gamma_{c,m}}$$

The calculation of the length of the interface then remains the same. Note that θ is $\frac{1}{2}\pi$ for $\gamma_{r,g} = 0$. If $\gamma_{r,g} < 0$, the two liquids will mix instead, forming one homogeneous spherical blob (Glazier & Graner, Phys. Rev. E 1993).



The preservation of aspect ratio for larger tissues

In the section concerning persistence, we assume that tissues of different sizes with the same number of segments and the same surface tensions, will end up with the same aspect ratio ($\frac{\text{length}}{\text{width}}$). This follows from the equations above. With the same surface tensions, the angles will be the same for the larger tissue, only the area of each segment is four times as large (we doubled both the length and the width of each segment). The radius of the circle then doubles, and so does the equilibrium contact length. In other words, if we start with a tissue that is twice the size in every dimension (so same aspect ratio), we end up with an elongated tissue with twice the equilibrium contact length, so still the same aspect ratio.

5.7 supplementary videos

- **S1 Video** Simulation with graded adhesion, strength $w = 11$, and no segment-specific adhesion. Duration is 500 000 MCS.
- **S2 Video** Simulation with axial adhesion, strength $\beta = 2$, and no segment-specific adhesion. Duration is 500 000 MCS.
- **S3 Video** Simulation with graded adhesion, strength $w = 11$, and segment-specific adhesion ($\gamma_{r,g} = 4$). Duration is 500 000 MCS.
- **S4 Video** Simulation with axial adhesion, strength $\beta = 2$, and segment-specific adhesion ($\gamma_{r,g} = 4$). Duration is 500 000 MCS.
- **S5 Video** Simulation with segment-specific adhesion, ($\gamma_{c,m} = 6$, $\gamma_{r,g} = 10$). Duration is 1 000 000 MCS.
- **S6 Video** Simulation with segment-specific adhesion (segment collapse), ($\gamma_{c,m} = 10$, $\gamma_{r,g} = 14$). Duration is 1 000 000 MCS.

"It's still magic even if you know how it's done."

Terry Pratchett, *A Hat Full of Sky*

6

Summarising Discussion

Animal segmentation is a well-studied developmental process with a rich history of theoretical modelling. In this thesis, we built on this earlier work to address several open questions, focusing on specific aspects of the developmental mechanism and its evolution. In chapter 2 we investigated the evolution and potential function of the frequency gradient (responsible for travelling waves) in vertebrate segmentation. In chapter 3 we studied how differences in the somite boundary determination mechanism of three vertebrate species, result in different abnormal phenotypes. In chapter 4 we asked under what conditions posterior growth and sequential segmentation can evolve. And finally in chapter 5 we investigated the interplay between convergent extension and an existing segment pattern. In this final chapter I will provide a synthesis of these topics, and demonstrate how together, they contribute to a more comprehensive understanding of the mechanism and evolution of sequential segmentation.

6.1 Making waves – frequency gradients

One of the most characteristic features of vertebrate somitogenesis is the wave of gene expression that travels from the posterior to the anterior PSM, a phenomenon caused by the slowing of gene expression oscillations towards the anterior PSM. With regards to its function, previous studies have argued that such a frequency gradient may be a necessary feature of the somite determination mechanism, by slowing down and ultimately freezing the oscillation phase (Jaeger and Goodwin, 2001), or because it creates a phase difference in the anterior PSM between the travelling wave oscillator and a second oscillator, which then triggers somite formation (Beaupeux and François, 2016, Harima and Kageyama, 2013). In other models, the frequency gradient arises spontaneously as a result of the underlying somite determination mechanism, such as synchronisation with frozen anterior cells (Murray *et al.*, 2011) or reaction-diffusion dynamics (Cotterell *et al.*, 2015). Alternatively, there is some experimental evidence that the posterior Wnt gradient influences the speed of oscillations (Gibb *et al.*, 2009). While investigating asymmetric somitogenesis in chapter 3, we found that the frequency gradient may have a different role in different vertebrate species. In zebrafish, the timing of somite formation is dictated by the speed of the travelling wave, while somite size is dictated by how far the FGF gradient retracts during one cycle. We speculated that the travelling wave may instigate somite polarity by switching on the determination program with a posterior-to-anterior timing. In mouse instead, the travelling wave oscillator seems to interact with an oscillator with fixed frequency (Harima and Kageyama, 2013, Niwa *et al.*, 2011). Due to the frequency gradient, these two are out of phase in the anterior: when the travelling wave reaches the anterior, the fixed frequency oscillator is in a low expression phase and a somite forms (similar to the mechanism in (Beaupeux and François, 2016)). Thus, the frequency gradient here determines the size of a somite, but not the timing of its formation.

In chapter 2, we showed that frequency gradients can evolve spontaneously. They arise a bit more often in tissues with a shallow, far-reaching morphogen gradient, which have a larger undifferentiated region anterior to the tailbud than tissues with a steep gradient: there is simply more room for waves with a shallow gradient. Strikingly, the percentage of simulations evolving a frequency gradient significantly increased in the presence of gene expression noise, indicating that frequency gradients may have a function in noise mitigation in shallow gradients. Tissues with a steep gradient instead only evolved more persistent oscillations in the presence of noise in our simulations. We therefore concluded that a combination of noise and a shallow gradient is required for the reliable evolution of a frequency gradient. This is consistent with an earlier suggestion by El-Sherif *et al.*, (El-Sherif *et al.*, 2014), that frequency gradients may lead to sharper segment boundaries in the presence of noisy morphogen gradients. There is a caveat however: segmentation took longer to evolve in simulations with a shallow gradient than in simulations with a steep gradient. We therefore speculate that persistent oscillations may have first evolved in the presence of noise and a steep gradient, and that a shallow gradient and consequent frequency gradient evolved secondarily.

The segmentation process of invertebrate animals supports the notion that segmentation evolved in tissues with a steep gradient. Take for instance the cephalochordate amphioxus, a model species from the sister group to vertebrates; this animal does not have an extensive PSM, instead its somites bud off directly from the tailbud (Beaster-Jones *et al.*, 2008). The short-germ insect *Tribolium* also has a short unsegmented region, with stripes that travel only a small distance before halting (El-Sherif *et al.*, 2012); this is likely also true for other arthropods (Brena and Akam, 2013). So why do vertebrates have such a long unsegmented region with long-range, shallow morphogen gradients and a frequency gradient? To answer this, we may need to consider that amphioxus has asymmetric somites: somitogenesis is not buffered against the left-right signalling pathway. Indeed, the antagonistic FGF and RA gradients – which are essential for (symmetric) vertebrate somitogenesis – do not play an important role in somite formation in amphioxus (Bertrand *et al.*, 2011, 2015). This suggests that the activity of FGF and RA gradients in vertebrate somitogenesis may have evolved to generate symmetric somites, which is required for proper development of the vertebrae (Brent, 2005). We speculate that the interactions between these gradients increased the size of the undifferentiated PSM, either to act explicitly as a buffer for small asymmetries, or just as a side-effect of their antagonism. Our simulations with a shallow gradient and gene expression noise (which is always present in biological systems) often yield frequency gradients across the unsegmented tissue, which suggests that the recruitment of FGF and RA would likely have been followed by the evolution of a frequency gradient. Once such a frequency gradient evolves, the resulting travelling wave could be adapted for roles in somite polarity determination or even become an essential part of the mechanism, like those we saw in chapter 3.

Arthropods also have left-right asymmetries, yet only a short unsegmented region probably associated with a steep morphogen gradient and absent or limited a frequency gradient. This suggests that there is no compensatory mechanism protecting segmentation from left-right asymmetries such as in vertebrates, but still their segments are symmetric. It turns out that *Drosophila* and perhaps also other insects establish left-right asymmetry in particular organs at a later stage than segmentation, precluding any interaction between these two processes (Namigai *et al.*, 2014). Furthermore, in arthropods, segmentation applies to all tissues, which are connected during segment generation; thus there is no need to coordinate the left and right parts as extensively as in vertebrates, where the left and right PSM are fully separated.

Future directions An intriguing direction for future research is to unravel the molecular interactions between the left-right pathway and somitogenesis in different species, and the compensatory mechanism that allows for symmetric somitogenesis. Once more elaborate experimental data are available, we can extend the phenomenological descriptions used in chapter 3. We can also approach this from a different angle, using evolutionary simulations to investigate what kind of compensatory mechanisms can evolve, and how this influences the evolution of the somite boundary formation mechanism. Furthermore, these simulations could potentially give more insight into the transition from amphioxus-like, steep-gradient segmentation to vertebrate, shallow-gradient segmentation.

6.2 Making shapes – axis extension

While it is an interesting question whether the common ancestor of bilaterians was segmented, it has not received a lot of attention in this thesis. Rather, we focused on the question why almost all animals generate their segments sequentially, which is interesting even if segmentation arose through common ancestry. In previous computational studies, it was shown that sequential segmentation easily evolves, is evolvable, and robust to gene expression noise (François *et al.*, 2007, ten Tusscher and Hogeweg, 2011). In chapter 4 however, we showed that despite these favourable properties, sequential segmentation does not evolve in the absence of a posterior signalling centre. Instead, a less robust mechanism emerged which relied on noise in division timing or cell-cell signalling, and in which all segments were generated at roughly the same time after a tissue-wide burst of divisions. Once we imposed a persistently present posterior morphogen, sequential segmentation indeed became the most frequently evolving segmentation strategy. Furthermore, we showed how a selection pressure for determinate growth should only emerge secondarily, in order for sequential segmentation to evolve. Our results demonstrate the importance of prior evolutionary innovations (e.g., posterior morphogen signal), and concurrent selection pressures (e.g., for determinate growth) for the type of mechanism that evolves.

In chapter 4, we assumed for simplicity that the main mechanism of axis extension is cell division. For vertebrate somitogenesis, this may be a good approximation: although the shape of the main axis is mostly generated via convergent extension, the PSM itself is laid down via divisions and/or cell ingression in the retracting tailbud (Tada and Heisenberg, 2012, Wilson *et al.*, 2009). Cells are thus "deposited" by the tailbud, which can be reasonably approximated by posteriorly localised divisions. In annelids, while there is no evidence for a segmentation clock, posterior growth is definitely responsible for segment addition (Balavoine, 2014). In arthropods instead, divisions are not restricted to the posterior undifferentiated zone and would anyway be insufficient for axis extension (Mayer *et al.*, 2010, Nakamoto *et al.*, 2015). There is also no active retraction mechanism like the vertebrate node, so the randomly oriented cell divisions in this posterior zone would more likely contribute to a blob than an extended axis. Instead, convergent extension is the major driver of axis extension; it occurs in the region where segments form in sequentially segmenting arthropods (like *Tribolium*), and after segment specification in simultaneously segmenting animals (like *Drosophila*).

We showed in chapter 5 that convergent extension can disturb an already-formed segmented pattern, but that segment-specific adhesion can prevent this. Moreover, segment-specific adhesion can actually drive convergent extension by itself, providing a direct way in which the insect segmental pattern could generate convergent extension. Segment-specific adhesion can also inform the directionality of an initially randomly oriented active cell migration mechanism, which may provide the energy required for tissue shape change. It has been shown that the Toll receptors that drive convergent extension in many arthropod species, are expressed in stripes (a pattern likely informed by the pair-rule genes (Benton *et al.*, 2016)). In *Drosophila*, these receptors inform actomyosin contractile cell polarity rather than driving convergent extension via segment-specific adhesion (Benton *et al.*, 2016, Irvine and Wieschaus, 1994, Mao and Lecuit, 2016, Paré *et al.*, 2014, Zallen and Wieschaus, 2004). We speculate that in other arthropods, segment-specific adhesion might play a role instead, or act in combination with cell-level actomyosin polarity.

We speculate that even if sequential segmentation evolved in parallel, arthropods and vertebrates started with very similar-looking mechanisms. Axis extension may have been very simple, with posterior divisions generating most of the main body axis. This allows for the rapid evolution of a segmentation mechanism with a simple clock operating in the posterior zone. (We will ignore for the moment the issue whether axis extension evolved before segmentation (Jacobs *et al.*, 2005), as well as the confounding possibility of a complex life cycle.) Very early in the lineage leading to velvet worms and arthropods, the mechanism in which the stripes drive convergent extension may have evolved, which would have made posteriorly localised divisions less important for axis elongation. As a consequence, more widespread divisions could have evolved (nearly) neutrally, leading

to the patterns we see now in *Tribolium* and velvet worms (Mayer *et al.*, 2010, Nakamoto *et al.*, 2015). Ultimately, this could have paved the way for the *Drosophila* style, simultaneous segmentation, where axis extension is necessarily driven by (segment-mediated) intercalation (Irvine and Wieschaus, 1994). In vertebrates, the restriction of segmentation to part of the mesoderm may have made it more difficult to let the segmentation pattern dictate overall axis extension. Instead, the overall mechanism of axis extension (divisions or convergent extension) varies between vertebrates depending on the availability of resources for the embryo (Steventon *et al.*, 2016). Differential adhesion could rather drive the separation of somites (Hester *et al.*, 2011), and may guide their epithelialisation process (Dias *et al.*, 2014).

Future directions In this thesis, we left the evolutionary order of posterior growth and sequential segmentation mostly unexplored, since our current results do not allow us to determine whether posterior growth evolved simultaneously with, or prior to, sequential segmentation. In the case of prior evolution of posterior growth, determinate growth also could have evolved prior to segmentation, while our current work suggests that this is unlikely if posterior growth and segmentation evolve concurrently. Furthermore, we did not take into account the existence of other patterning processes that occur at the same time as axis extension and segmentation, and which may influence their evolution. For example, the HOX genes divide the main body axis into several regions with different fates, and are connected to the segmentation clock in vertebrates (Cordes *et al.*, 2004, Zákány *et al.*, 2001), and required for the termination of posterior growth (Denans *et al.*, 2015). Thus, the prior or simultaneous evolution of HOX patterning with growth and segmentation, may provide a signal for terminating growth at the right time during development. This could potentially resolve the bias towards simultaneous segmentation when selecting for determinate growth, allowing for the concurrent evolution of determinate posterior growth and sequential segmentation.

Another interesting new direction is the evolution of segment-driven convergent extension as observed in arthropods. In order to investigate this, the current evolutionary models should be extended to 2D tissues, with the possibility to regulate tissue remodelling mechanisms such as differential adhesion. This would also provide additional degrees of freedom for evolution to explore; for instance, such an extended model might yield new insights into the evolution of segment splitting (period doubling) observed in *Tribolium* (Clark and Akam, 2016, Sarrazin *et al.*, 2012).

6.3 Using developmental and evolutionary models

The models we used in this thesis are relatively simple, trying to explain particular phenomena or assess evolutionary consequences with a fairly minimal set of underlying assumptions. We used the developmental models in chapters 3 and 5 as tools for testing the developmental consequences of known mechanisms and exposing what information is still missing: for example, the effect of convergent extension on a segmented pattern in chapter 5. In the evo-devo models of chapters 2 and 4 instead, we had in mind what type of developmental phenomenon we want to obtain – a frequency gradient (chapter 2), or sequential segmentation (chapter 4) – and investigated what prior conditions are required for this mechanism to be the predominant outcome in our simulations (similar to (Troein *et al.*, 2009)). Note that we do not directly select for the mechanism we are looking for, but for a more general outcome – a segmented pattern at the end of development. The evolutionary process is therefore free to generate any mechanism to accomplish this final pattern. We thereby found that despite the favourable properties of sequential segmentation, it does not always evolve, for instance in the presence of a selection pressure for determinate growth. Thus, our work is a nice example of the importance of prior conditions for the type of mechanism that evolves (François, 2014).

A disadvantage of our evo-devo models is their larger complexity compared to the developmental models, as they incorporate scales of organisation ranging from the genome, through the individual to the population. This also means that the outcome is not always easy to interpret and may require extensive analysis. Chapter 2 is a prime example of this. We set out to find differences between networks evolved in tissues with steep or shallow gradients, which proved to be difficult due to the size and sheer entanglement of the evolved networks. To circumvent this issue and to analyse many simulations efficiently, we developed an automated pipeline to assess the general network properties (size, connections) and perform Fourier analysis on the oscillatory dynamics of the network. The Fourier analysis revealed the existence of frequency gradients, the presence of which we suspected from the space-time plots. Moreover, the pipeline allowed us to assess the likelihood of evolving frequency gradients in several different conditions, such as the presence of expression noise or the availability of cell-cell signalling. We applied a similar approach in chapter 4: while we did look into the details of particularly interesting evolved mechanisms, we again mainly focused on the larger pattern of evolutionary likelihood, this time of sequential segmentation in general.

Our way of assessing evolutionary likelihood in chapters 2 and 4 is different from the approach in previous studies on *in silico* evolution of segmentation (François *et al.*, 2007, Salazar-Ciudad *et al.*, 2001a, ten Tusscher and Hogeweg, 2011). These studies did not explicitly assess the likelihood of a particular mechanism but instead demonstrated that, compared to the other mechanisms that evolved *in silico*, the mechanism that actually is

present in animals has more favourable properties – such as robustness and evolvability. They also showed how sequential segmentation can easily evolve gradually from scratch, from simple small networks producing few segments, to more complex networks generating many (François *et al.*, 2007, ten Tusscher and Hogeweg, 2011). As we focus more on the likelihood of evolving a particular mechanism, we also try to connect our results more strongly to the actual evolutionary history of segmentation. We should note that these are model-guided suggestions, not absolute statements: our models are not able to prove that a specific evolutionary trajectory was actually taken, only to suggest that one is more likely than another.

6.4 Moving models beyond standard clock-and-wavefront

Somite polarity determination In this thesis as well as the work of others, the evolutionary simulations worked with a fairly simple fitness criterion: segments are defined by the alternating high and low expression of a single gene, with the high and low expression domains possibly representing segment polarity, or a higher-level double segment periodicity. The result of selection for this phenotype is typically a mechanism that combines oscillation with a bistability motif, translating the oscillation phase directly to one of two alternative attractors upon passing of the wavefront. The superimposed wavefront or morphogen gradient retraction results in the cell-by-cell transition from oscillations to determination, i.e., the point of segment determination also retracts smoothly ((François *et al.*, 2007, ten Tusscher and Hogeweg, 2011), chapters 2 and 4); furthermore, high- and low expression stripes are formed consecutively. While these dynamics seem to correspond well to arthropod segmentation, this is not the case for vertebrates, where somites are first defined as a block with roughly homogeneous expression before resolving into a clear anterior and posterior half (Harima and Kageyama, 2013, Oginuma *et al.*, 2010). During somitogenesis, oscillations terminate well after somite boundary definition has occurred (Shih *et al.*, 2015), while in our simulations the two events are simultaneous (chapter 2,4). If we want to study the evolution of vertebrate segmentation more in-depth, we may need to consider different patterning criteria and functional constraints in the model.

An alternative could be to first select for block-like formation of segments, like the blocks in the original clock-and-wavefront model (Cooke and Zeeman, 1976). A recent study selected for the detection of a phase difference between two oscillators, which also yielded roughly block-like somite formation and which is reminiscent of the mouse mechanism we implemented in chapter 3 (Beaupeux and François, 2016). Perhaps somite polarity can then be selected for secondarily, a trick we also used for determinate growth in chapter 4. Conversely, it could be that the block-like mechanism evolved secondarily from a cell-by-cell determination mechanism, perhaps in conjunction with a frequency

gradient (see section 6.1). To resolve this, we need more experimental data on somite polarity determination in amphioxus and other early-branching chordates.

An alternative patterning mechanism Recently, a different somite patterning mechanism was proposed based on an ensemble approach, investigating all three-gene networks capable of performing sequential segmentation (Cotterell *et al.*, 2015). Strikingly, the dominant mechanism emerging from this study was not a clock-and-wavefront model, but a reaction-diffusion mechanism (PORD), where an inhibitor diffusing from the formed somites causes the formation of the next somite. The biggest difference between this mechanism and the more widely-accepted clock and wavefront mechanism is that the PORD mechanism does not rely on the morphogen gradients to instruct the position of the determination front (Cotterell *et al.*, 2015). Regardless of the question whether the PORD model is the actual mechanism behind (all) vertebrate species (or even applicable to insects?), it is an interesting new mechanism, especially since the ensemble approach suggests that it is the more robust and likely evolutionary outcome (Cotterell *et al.*, 2015). In general, it is an interesting question whether evolution can select a solution with as simple a core as a three gene network. Furthermore, as we discussed in the previous section, a mechanism with selective advantages like robustness may still not be the one that evolves, and to our knowledge, none of the evolutionary simulations has yielded a purely PORD-like mechanism; a similar mechanism only evolved in combination with feed-forward mechanisms, yielding roughly simultaneous segmentation (Kohsokabe and Kaneko, 2016). This could be due to a lack of diffusive gene products in the models with a retracting morphogen wavefront (in this thesis we only included direct cell-cell signalling for example). It would be an interesting exercise to see if and when a PORD-like mechanism can evolve, which would also provide more arguments for or against the model itself.

From sequential to simultaneous segmentation In this thesis we mostly ignored simultaneous, *Drosophila*-like segmentation. This mode is thought to have evolved secondarily from a sequential mode, but how this happened is only recently becoming more clear (Clark and Akam, 2016), El-Sherif and Francois, personal communication). In chapter 4, individuals with sequential segmentation transitioned to a simultaneous mode once the selection for determinate growth became too high. This was not a smooth transition – individuals first lost the ability to grow and make segments, then invented the simultaneous mode from scratch. Moreover, this evolved simultaneous mechanism did not resemble the hierarchical manner in which *Drosophila* generates its segments. It would be very exciting to investigate what priorly evolved developmental processes, and what genetic and cellular tools are required to make a smooth transition possible from sequential segmentation to a truly *Drosophila*-like mode. Consider for instance the tissue dynamics: there is a whole range of short, intermediate and long-germ insects, which generate very few to all of their segments in the early syncytial stage before cells are

formed and the body axis takes shape. To enable a smooth transition, it may therefore be necessary to add the evolution of a syncytial stage to the model. Besides the smoothness of the transition, the model should be able to evolve the genetic hierarchy of *Drosophila*, which specifies each segment with a unique combination of upstream genes. Previous evolutionary models have shown that selection for a specific tissue pattern may yield a hierarchical mode of segmentation (François *et al.*, 2007, Salazar-Ciudad *et al.*, 2001a), so requiring the preservation of the specific priorly evolved pattern may aid in obtaining a *Drosophila*-like simultaneous mode. Evolving hierarchic gene regulation may be further facilitated by selection for axis regionalisation, like in (ten Tusscher and Hogeweg, 2011).

6.5 Concluding remarks

Modelling has a solid place in research on segmentation; from the very first model by Cooke and Zeeman to the most recent developments on the clock-to-stripe transition. We contributed with evolutionary and developmental models that give insights into the process of vertebrate somitogenesis and on the interplay between axis extension and segmentation. As we discussed here, there are still plenty of open questions about segmentation, which will require several different approaches, both theoretical and experimental. We think that ultimately, segmentation should be studied within the greater context of early embryonic development and its evolution, not as an isolated mechanism.

Bibliography

- Abley, K., De Reuille, P. B., Strutt, D., Bangham, A., Prusinkiewicz, P., Marée, A. F. M., Grieneisen, V. A., and Coen, E. 2013. An intracellular partitioning-based framework for tissue cell polarity in plants and animals. *Development*, 140(10): 2061–2074. (Cited on page 12.)
- Akiyama, R., Masuda, M., Tsuge, S., Bessho, Y., and Matsui, T. 2014. An anterior limit of FGF/Erk signal activity marks the earliest future somite boundary in zebrafish. *Development*, 141(5): 1104–1109. (Cited on pages 9, 58, 69, 83, and 86.)
- Ares, S., Morelli, L. G., Jörg, D. J., Oates, A. C., and Jülicher, F. 2012. Collective modes of coupled phase oscillators with delayed coupling. *Phys. Rev. Lett.*, 108: 204101. (Cited on pages 58 and 85.)
- Aulehla, A. and Pourquié, O. 2008. Oscillating signaling pathways during embryonic development. *Curr. Opin. Cell Biol.*, 20(6): 632 – 637. (Cited on page 22.)
- Aulehla, A. and Pourquié, O. 2010. Signaling gradients during paraxial mesoderm development. *Cold Spring Harb. Perspect. Biol.*, 2(2): 1–17. (Cited on pages 2, 7, 57, 84, 86, and 106.)
- Backes, T. M., Latterman, R., Small, S. A., Mattis, S., Pauley, G., Reilly, E., and Lubkin, S. R. 2009. Convergent extension by intercalation without mediolaterally fixed cell motion. *J. Theor. Biol.*, 256(2): 180 – 186. (Cited on pages 133, 134, 147, and 148.)
- Bajard, L., Morelli, L. G., Ares, S., Pécéréaux, J., Jülicher, F., and Oates, A. C. 2014. Wnt-regulated dynamics of positional information in zebrafish somitogenesis. *Development*, 141(6): 1381–1391. (Cited on page 8.)
- Balavoine, G. 2014. Segment formation in annelids: patterns, processes and evolution. *Int. J. Dev. Biol.*, 58: 469–483. (Cited on pages 7, 13, 14, 99, 116, 121, and 171.)
- Balavoine, G. and Adoutte, A. 2003. The segmented urbilateria: A testable scenario. *Integr. Comp. Biol.*, 43(1): 137–147. (Cited on pages 2, 13, 21, and 99.)
- Beaster-Jones, L., Kaltenbach, S., Koop, D., Yuan, S., Chastain, R., and Holland, L. 2008. Expression of somite segmentation genes in amphioxus: a clock without a wavefront? *Dev. Genes Evol.*, 218(11-12): 599–611. (Cited on pages 49 and 169.)
- Beaupeux, M. and François, P. 2016. Positional information from oscillatory phase shifts : insights from *in silico* evolution. *Phys. Biol.*, 13(3): 036009. (Cited on pages 3, 8, 10, 28, 49, 84, 168, and 174.)
- Belmonte, J. M., Swat, M. H., and Glazier, J. A. 2016. Filopodial-tension model of convergent-extension of tissues. *PLoS Comput. Biol.*, 12(6): 1–20. (Cited on page 13.)
- Beltman, J. B., Marée, A. F. M., Lynch, J. N., Miller, M. J., and De Boer, R. J. 2007. Lymph node topology dictates T cell migration behaviour. *J. Exp. Med.*, 204(4): 771–780. (Cited on pages 135, 143, 154, and 156.)
- Bénazéraf, B. and Pourquié, O. 2013. Formation and segmentation of the vertebrate body axis. *Annu. Rev. Cell Dev. Biol.*, 29(1): 1–26. PMID: 23808844. (Cited on pages 7, 11, and 99.)
- Benton, M., Pechmann, M., Frey, N., Stappert, D., Conrads, K., Chen, Y.-T., Stamatakis, E., Pavlopoulos, A., and Roth, S. 2016. Toll genes have an ancestral role in axis

- elongation. *Curr. Biol.*, 26(12): 1609 – 1615. (Cited on pages 12 and 171.)
- Benton, M. A., Akam, M., and Pavlopoulos, A. 2013. Cell and tissue dynamics during *Tribolium* embryogenesis revealed by versatile fluorescence labeling approaches. *Development*, 140(15): 3210–3220. (Cited on page 134.)
- Bertet, C., Sulak, L., and Lecuit, T. 2004. Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. *Nature*, 429: 667–671. (Cited on pages 13, 133, and 151.)
- Bertrand, S., Camasses, A., Somorjai, I., Belgacem, M. R., Chabrol, O., Escande, M.-L., Pontarotti, P., and Escriva, H. 2011. Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits. *PNAS*, 108(22): 9160–9165. (Cited on pages 84 and 169.)
- Bertrand, S., Aldea, D., Oulion, S., Subirana, L., de Lera, A. R., Somorjai, I., and Escriva, H. 2015. Evolution of the role of RA and FGF signals in the control of somitogenesis in chordates. *PLoS ONE*, 10(9): e0136587. (Cited on pages 49, 84, and 169.)
- Blair, S. S. 2008. Segmentation in animals. *Curr. Biol.*, 18(21): R991 – R995. (Cited on pages 2 and 99.)
- Boettger, T., Wittler, L., and Kessel, M. 1999. FGF8 functions in the specification of the right body side of the chick. *Curr. Biol.*, 9(5): 277 – 280. (Cited on page 58.)
- Brena, C. and Akam, M. 2013. An analysis of segmentation dynamics throughout embryogenesis in the centipede *Strigamia maritima*. *BMC Biol.*, 11(1): 1–18. (Cited on pages 7 and 169.)
- Brend, T. and Holley, S. A. 2009. Balancing segmentation and laterality during vertebrate development. *Semin. Cell Dev. Biol.*, 20(4): 472 – 478. (Cited on page 57.)
- Brent, A. E. 2005. Somite formation: Where left meets right. *Curr. Biol.*, 15(12): R468 – R470. (Cited on pages 57, 58, 65, 84, and 169.)
- Brodland, G. and Veldhuis, J. 2012. Assessing the mechanical energy costs of various tissue reshaping mechanisms. *Biomech. Model. Mechanobiol.*, 11(8): 1137–1147. (Cited on pages 133 and 147.)
- Brodland, G. W. 2006. Do lamellipodia have the mechanical capacity to drive convergent extension? *Int. J. Dev. Biol.*, 50: 151–155. (Cited on pages 13, 133, and 147.)
- Budd, G. E. 2001. Why are arthropods segmented? *Evol. Dev.*, 3(5): 332–342. (Cited on pages 13, 14, 21, and 99.)
- Chipman, A. D. 2010. Parallel evolution of segmentation by co-option of ancestral gene regulatory networks. *BioEssays*, 32(1): 60–70. (Cited on pages 14, 21, and 99.)
- Choe, C. P., Miller, S. C., and Brown, S. J. 2006. A pair-rule gene circuit defines segments sequentially in the short-germ insect *Tribolium castaneum*. *Proc. Natl. Acad. Sci. U.S.A.*, 103(17): 6560–6564. (Cited on pages 2, 48, 134, and 150.)
- Clark, E. and Akam, M. 2016. Odd-paired controls frequency doubling in *Drosophila* segmentation by altering the pair-rule gene regulatory network. *eLife*, 5: e18215. (Cited on pages 3, 172, and 175.)
- Cooke, J. and Zeeman, E. 1976. A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *J. Theor. Biol.*, 58: 455–476. (Cited on pages 2, 4, 8, 9, 57, and 174.)
- Cordes, R., Schuster-Gossler, K., Serth, K., and Gossler, A. 2004. Specification of vertebral identity is coupled to Notch signalling and the segmentation clock. *Development*, 131(6): 1221–1233. (Cited on page 172.)

- Cotterell, J. and Sharpe, J. 2010. An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. *Mol. Syst. Biol.*, 6(1). (Cited on page 5.)
- Cotterell, J., Robert-Moreno, A., and Sharpe, J. 2015. A local, self-organizing reaction-diffusion model can explain somite patterning in embryos. *Cell Syst.*, 1(4): 257–269. (Cited on pages 3, 8, 10, 83, 168, and 175.)
- Couso, J. P. 2009. Segmentation, metamerism and the cambrian explosion. *Int. J. Dev. Biol.*, 53: 1305–1316. (Cited on pages 21 and 99.)
- Crampin, E. J., Hackborn, W. W., and Maini, P. K. 2002. Pattern formation in reaction-diffusion models with nonuniform domain growth. *Bull. Math. Biol.*, 64(4): 747–769. (Cited on pages 110 and 112.)
- Crombach, A. and Hogeweg, P. 2008. Evolution of evolvability in gene regulatory networks. *PLoS Comput. Biol.*, 4(7): 1–13. (Cited on pages 23 and 101.)
- Crombach, A., Wotton, K. R., Cicin-Sain, D., Ashyraliyev, M., and Jaeger, J. 2012. Efficient reverse-engineering of a developmental gene regulatory network. *PLoS Comput. Biol.*, 8(7): 1–21. (Cited on page 8.)
- Davis, G. K. and Patel, N. H. 1999. The origin and evolution of segmentation. *Trends Genet.*, 15(12): M68 – M72. (Cited on pages 2, 21, and 99.)
- Denans, N., Imura, T., and Pourquié, O. 2015. Hox genes control vertebrate body elongation by collinear Wnt repression. *eLife*, 4: e04379. (Cited on page 172.)
- Dequéant, M.-L. and Pourquié, O. 2008. Segmental patterning of the vertebrate embryonic axis. *Nat. Rev. Genet.*, 9: 370–382. (Cited on pages 22 and 47.)
- Dequéant, M.-L., Glynn, E., Gaudenz, K., Wahl, M., Chen, J., Mushegian, A., and Pourquié, O. 2006. A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science*, 314(5805): 1595–1598. (Cited on page 22.)
- Dias, A. S., de Almeida, I., Belmonte, J. M., Glazier, J. A., and Stern, C. D. 2014. Somites without a clock. *Science*, 343(6172): 791–795. (Cited on pages 8 and 172.)
- Diez del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M., and Storey, K. 2003. Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron*, 40(1): 65 – 79. (Cited on pages 60, 65, and 84.)
- Dobzhansky, T. 1973. Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.*, 35(3): 125–129. (Cited on page 5.)
- Dubrulle, J., McGrew, M. J., and Pourquié, O. 2001. FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation. *Cell*, 106(2): 219 – 232. (Cited on pages 8 and 65.)
- El-Sherif, E., Averof, M., and Brown, S. J. 2012. A segmentation clock operating in blastoderm and germband stages of *Tribolium* development. *Development*, 139(23): 4341–4346. (Cited on pages 2, 3, 7, 48, 134, and 169.)
- El-Sherif, E., Zhu, X., Fu, J., and Brown, S. J. 2014. Caudal regulates the spatiotemporal dynamics of pair-rule waves in *Tribolium*. *PLoS Genet.*, 10(10): e1004677. (Cited on pages 22, 45, 48, and 169.)
- François, P. 2014. Evolving phenotypic networks in silico. *Semin. Cell Dev. Biol.*, 35(0): 90 – 97. (Cited on pages 22 and 173.)
- François, P. and Siggia, E. D. 2012. Phenotypic models of evolution and development: geometry as destiny. *Curr. Opin. Genet. Dev.*, 22(6): 627 – 633. Genetics of system

- biology. (Cited on page 80.)
- François, P., Hakim, V., and Siggia, E. D. 2007. Deriving structure from evolution: metazoan segmentation. *Mol. Syst. Biol.*, 3(1). (Cited on pages 10, 15, 22, 26, 28, 82, 100, 103, 108, 120, 170, 173, 174, and 176.)
- Fujimoto, K., Ishihara, S., and Kaneko, K. 2008. Network evolution of body plans. *PLoS ONE*, 3(7): e2772. (Cited on pages 6, 15, 100, 103, and 120.)
- Gibb, S., Zagorska, A., Melton, K., Tenin, G., Vacca, I., Trainor, P., Maroto, M., and Dale, J. K. 2009. Interfering with Wnt signalling alters the periodicity of the segmentation clock. *Dev. Biol.*, 330(1): 21 – 31. (Cited on pages 9 and 168.)
- Gissis, S. B., Jablonka, E., and Zeligowski, A. 2011. *Transformations of Lamarckism: From Subtle Fluids to Molecular Biology*. MIT Press. (Cited on page 5.)
- Glazier, J. A. and Graner, F. m. c. 1993. Simulation of the differential adhesion driven rearrangement of biological cells. *Phys. Rev. E*, 47: 2128–2154. (Cited on pages 134, 135, and 153.)
- Gold, D. A., Runnegar, B., Gehling, J. G., and Jacobs, D. K. 2015. Ancestral state reconstruction of ontogeny supports a bilaterian affinity for Dickinsonia. *Evol. Dev.*, 17(6): 315–324. (Cited on pages 14, 99, 100, and 122.)
- Goldbeter, A. and Pourquié, O. 2008. Modeling the segmentation clock as a network of coupled oscillations in the Notch, Wnt and FGF signaling pathways. *J. Theor. Biol.*, 252(3): 574–85. (Cited on pages 8, 22, and 48.)
- Graham, A., Butts, T., Lumsden, A., and Kiecker, C. 2014. What can vertebrates tell us about segmentation? *EvoDevo*, 5: 1–8. (Cited on page 99.)
- Graner, F. m. c. and Glazier, J. A. 1992. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys. Rev. Lett.*, 69: 2013–2016. (Cited on page 134.)
- Grima, R. and Schnell, S. 2007. Can tissue surface tension drive somite formation? *Dev. Biol.*, 307(2): 248 – 257. (Cited on pages 8 and 141.)
- Hannibal, R. L. and Patel, N. H. 2013. What is a segment? *EvoDevo*, 4(1): 1–10. (Cited on page 14.)
- Harima, Y. and Kageyama, R. 2013. Oscillatory links of Fgf signaling and Hes7 in the segmentation clock. *Curr. Opin. Genet. Dev.*, 23(4): 484 – 490. Developmental mechanisms, patterning and evolution. (Cited on pages 8, 10, 73, 84, 87, 168, and 174.)
- Harper, J., Rosen, B., and White, J. 1986. Growth and form of modular organisms. *Philos. Trans. R. Soc. Lond.*, 313(1159): p. 1–250. (Cited on page 116.)
- Hayashi, S., Shimoda, T., Nakajima, M., Tsukada, Y., Sakumura, Y., Dale, J. K., Maroto, M., Kohno, K., Matsui, T., and Bessho, Y. 2009. Sprouty4, an FGF inhibitor, displays cyclic gene expression under the control of the Notch segmentation clock in the mouse PSM. *PLoS ONE*, 4(5): 1–8. (Cited on pages 74 and 78.)
- Herrgen, L., Ares, S., Morelli, L. G., Schröter, C., Jülicher, F., and Oates, A. C. 2010. Intercellular coupling regulates the period of the segmentation clock. *Curr. Biol.*, 20(14): 1244 – 1253. (Cited on pages 8, 57, and 85.)
- Hester, S. D., Belmonte, J. M., Gens, J. S., Clendenon, S. G., and Glazier, J. A. 2011. A multi-cell, multi-scale model of vertebrate segmentation and somite formation. *PLoS Comput. Biol.*, 7(10): e1002155. (Cited on pages 4, 8, 9, 134, and 172.)
- Hogeweg, P. 2012. *Evolutionary Systems Biology*, chapter Toward a Theory of Multilevel Evolution: Long-Term Information Integration Shapes the Mutational Landscape and Enhances Evolvability, pages 195–224. *Adv. Exp. Med. Biol.* (Cited on page 5.)

- Honda, H., Nagai, T., and Tanemura, M. 2008. Two different mechanisms of planar cell intercalation leading to tissue elongation. *Dev. Dynam.*, 237(7): 1826–1836. (Cited on pages 12, 13, 133, and 147.)
- Huang, S., Ma, J., Liu, X., Zhang, Y., and Luo, L. 2011. Retinoic acid signaling sequentially controls visceral and heart laterality in zebrafish. *J. Biol. Chem.*, 286(32): 28533–28543. (Cited on page 58.)
- Hubaud, A. and Pourquié, O. 2014. Signalling dynamics in vertebrate segmentation. *Nat. Rev. Mol. Cell Biol.*, 15: 709–721. (Cited on pages 7, 21, 57, 68, and 71.)
- Ingham, P. W. 1988. The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature*, 335: 25–34. (Cited on page 8.)
- Irvine, K. and Wieschaus, E. 1994. Cell intercalation during *Drosophila* germband extension and its regulation by pair-rule segmentation genes. *Development*, 120(4): 827–841. (Cited on pages 12, 13, 134, 147, 171, and 172.)
- Jacobs, D. K., Hughes, N. C., Fitz-Gibbon, S. T., and Winchell, C. J. 2005. Terminal addition, the Cambrian radiation and the Phanerozoic evolution of bilaterian form. *Evol. Dev.*, 7(6): 498–514. (Cited on pages 99, 122, and 171.)
- Jacobs-McDaniels, N. L. and Albertson, R. C. 2011. Chd7 plays a critical role in controlling left-right symmetry during zebrafish somitogenesis. *Dev. Dyn.*, 240(10): 2272–2280. (Cited on page 58.)
- Jaeger, J. and Goodwin, B. C. 2001. A cellular oscillator model for periodic pattern formation. *J. Theor. Biol.*, 213(2): 171 – 181. (Cited on pages 8, 9, 22, 61, 82, 85, and 168.)
- Jaeger, J. and Sharpe, J. 2014. *Towards a Theory of Development*, chapter On the concept of mechanism in development, pages 56–78. Oxford University Press. (Cited on page 4.)
- Jaeger, J., Blagov, M., Kosman, D., Kozlov, K. N., Manu, Myasnikova, E., Surkova, S., Vanario-Alonso, C. E., Samsonova, M., Sharp, D. H., and Reinitz, J. 2004. Dynamical analysis of regulatory interactions in the Gap gene system of *Drosophila melanogaster*. *Genetics*, 167(4): 1721–1737. (Cited on page 8.)
- Jiang, Y.-J., Aerne, B. L., Smithers, L., Haddon, C., Ish-Horowitz, D., and Lewis, J. 2000. Notch signalling and the synchronization of the somite segmentation clock. *Nature*, 408: 475–479. (Cited on page 57.)
- Jiménez, A., Munteanu, A., and Sharpe, J. 2015. Dynamics of gene circuits shapes evolvability. *PNAS*, 112(7): 2103–2108. (Cited on page 5.)
- Kaern, M., Menzinger, M., and Hunding, A. 2000. Segmentation and somitogenesis derived from phase dynamics in growing oscillatory media. *J. Theor. Biol.*, 207(4): 473 – 493. (Cited on page 22.)
- Kappen, C., Schughart, K., and Ruddle, F. H. 1989. Two steps in the evolution of antenpedia-class vertebrate homeobox genes. *Proc. Natl. Acad. Sci. USA*, 86: 5459–5463. (Cited on page 122.)
- Kato, Y. 2011. The multiple roles of Notch signaling during left-right patterning. *Cell. Mol. Life Sci.*, 68(15): 2555–2567. (Cited on page 58.)
- Kawakami, Y., Ángel Raya, Raya, R. M., Rodríguez-Esteban, C., and Belmonte, J. C. I. 2005. Retinoic acid signalling links left-right asymmetric patterning and bilaterally symmetric somitogenesis in the zebrafish embryo. *Nature*, 435: 165–171. (Cited on pages 57, 58, 60, 64, 72, 80, and 81.)
- Kawamura, A., Koshida, S., Hijikata, H., Sakaguchi, T., Kondoh, H., and Takada, S.

2005. Zebrafish Hairy/Enhancer of Split protein links FGF signaling to cyclic gene expression in the periodic segmentation of somites. *Genes Dev.*, 19(10): 1156–1161. (Cited on page 69.)
- Keller, R., Davidson, L., Edlund, A., Elul, T., Ezin, M., Shook, D., and Skoglund, P. 2000. Mechanisms of convergence and extension by cell intercalation. *Phil. Trans. R. Soc. Lond. B*, 355(1399): 897–922. (Cited on page 133.)
- Kohsokabe, T. and Kaneko, K. 2016. Evolution-development congruence in pattern formation dynamics: Bifurcations in gene expression and regulation of networks structures. *J. Exp. Zool. B Mol. Dev. Evol.*, 326(1): 61–84. (Cited on pages 6, 15, and 175.)
- Komatsu, Y. and Mishina, Y. 2013. Establishment of left–right asymmetry in vertebrate development: the node in mouse embryos. *Cell. Mol. Life Sci.*, 70(24): 4659–4666. (Cited on page 64.)
- Krebs, L. T., Iwai, N., Nonaka, S., Welsh, I. C., Lan, Y., Jiang, R., Saijoh, Y., O’Brien, T. P., Hamada, H., and Gridley, T. 2003. Notch signaling regulates left-right asymmetry determination by inducing Nodal expression. *Genes Dev.*, 17(10): 1207–1212. (Cited on page 58.)
- Krol, A. J., Roellig, D., Dequéant, M.-L., Tassy, O., Glynn, E., Hattem, G., Mushegian, A., Oates, A. C., and Pourquié, O. 2011. Evolutionary plasticity of segmentation clock networks. *Development*, 138(13): 2783–2792. (Cited on pages 11, 58, 68, 69, and 73.)
- Käfer, J., Hayashi, T., Marée, A. F. M., Carthew, R. W., and Graner, F. 2007. Cell adhesion and cortex contractility determine cell patterning in the *Drosophila* retina. *Proc. Natl. Acad. Sci. U.S.A.*, 104(47): 18549–18554. (Cited on page 134.)
- Lack, J. B., Monette, M. J., Johanning, E. J., Sprengelmeyer, Q. D., and Pool, J. E. 2016. Decanalization of wing development accompanied the evolution of large wings in high-altitude *Drosophila*. *PNAS*, 113(4): 1014–1019. (Cited on page 4.)
- Lauschke, V. M., Tsiarris, C. D., François, P., and Aulehla, A. 2013. Scaling of embryonic patterning based on phase-gradient encoding. *Nature*, 493: 101–105. (Cited on pages 8, 10, 48, and 49.)
- Lewis, J. 2003. Autoinhibition with transcriptional delay: A simple mechanism for the zebrafish somitogenesis oscillator. *Curr. Biol.*, 13(16): 1398 – 1408. (Cited on page 8.)
- Mallo, M. 2016. Revisiting the involvement of signaling gradients in somitogenesis. *FEBS J.*, 283(8): 1430–1437. (Cited on page 8.)
- Mallo, M. and Alonso, C. R. 2013. The regulation of Hox gene expression during animal development. *Development*, 140(19): 3951–3963. (Cited on page 106.)
- Mao, Q. and Lecuit, T. 2016. Evo-devo: Universal toll pass for the extension highway? *Curr. Biol.*, 26(14): R680 – R683. (Cited on pages 12 and 171.)
- Marée, A. F. M. and Hogeweg, P. 2002. Modelling *Dictyostelium discoideum* morphogenesis: The culmination. *Bull. Math. Biol.*, 64(2): 327–353. (Cited on page 4.)
- Marée, A. F. M., Grieneisen, V. A., and Hogeweg, P. 2007. The Cellular Potts Model and biophysical properties of cells, tissues and morphogenesis. In A. R. A. Anderson, M. A. J. Chaplain, and K. A. Rejniak, editors, *Single-Cell-Based Models in Biology and Medicine*, pages 107–136. Birkhäuser Verlag, Basel. (Cited on page 134.)
- Martin, B. L. and Kimelman, D. 2009. Wnt signaling and the evolution of embryonic posterior development. *Curr. Biol.*, 19(5): R215 – R219. (Cited on page 121.)
- Marée, A. F. M. and Hogeweg, P. 2001. How amoeboids self-organize into a fruiting body: Multicellular coordination in *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. U.S.A.*, 98(7): 3879–3883. (Cited on pages 4 and 134.)

- Mayer, G., Kato, C., Quast, B., Chisholm, R. H., Landman, K. A., and Quinn, L. M. 2010. Growth patterns in Onychophora (velvet worms): lack of a localised posterior proliferation zone. *BMC Evol. Biol.*, 10:339(1): 1–12. (Cited on pages 99, 171, and 172.)
- Mazzitello, K. I., Arizmendi, C. M., and Hentschel, H. G. E. 2008. Converting genetic network oscillations into somite spatial patterns. *Phys. Rev. E*, 78: 021906 1–8. (Cited on page 9.)
- McMillan, W., Monteiro, A., and Kapan, D. D. 2002. Development and evolution on the wing. *Trends Ecol. Evol.*, 17(3): 125–133. (Cited on page 4.)
- Meinhardt, H. 1982. *Models of Biological Pattern Formation*. Academic Press, London. (Cited on pages 83, 106, and 122.)
- Meinhardt, H. 1986. *Somites in developing embryos*, chapter Models of segmentation, pages 179–189. NATO ASI Series. (Cited on page 9.)
- Meinhardt, H. 2015. Models for patterning primary embryonic body axes: The role of space and time. *Semin. Cell Dev. Biol.*, 42: 103 – 117. Claudins and Time, Space and the Vertebrate Body Axis. (Cited on page 122.)
- Meyers, E. N. and Martin, G. R. 1999. Differences in left-right axis pathways in mouse and chick: Functions of FGF8 and SHH. *Science*, 285(5426): 403–406. (Cited on pages 80 and 82.)
- Minelli, A. and Fusco, G. 2004. Evo-devo perspectives on segmentation: model organisms, and beyond. *Trends Ecol. Evol.*, 19(8): 423 – 429. (Cited on pages 21 and 99.)
- Monier, B., Pélissier-Monier, A., Brand, A. H., and Sanson, B. 2010. An actomyosin-based barrier inhibits cell mixing at compartmental boundaries in *Drosophila* embryos. *Nat. Cell Biol.*, 12: 60–65. (Cited on page 150.)
- Monier, B., Pélissier-Monier, A., and Sanson, B. 2011. Establishment and maintenance of compartmental boundaries: role of contractile actomyosin barriers. *Cell. Mol. Life Sci.*, 68(11): 1897–1910. (Cited on pages 13 and 150.)
- Morelli, L. G., Ares, S., Herrgen, L., Schröter, C., Jülicher, F., and Oates, A. C. 2009. Delayed coupling theory of vertebrate segmentation. *Hfsp J.*, 3(1): 55–66. PMID: 19492022. (Cited on pages 8, 16, 47, 58, 61, 62, 69, 82, 83, and 85.)
- Murray, P. J., Maini, P. K., and Baker, R. E. 2011. The clock and wavefront model revisited. *J. Theor. Biol.*, 283(1): 227 – 238. (Cited on pages 8, 9, 69, 83, 85, and 168.)
- Nakamoto, A., Hester, S. D., Constantinou, S. J., Blaine, W. G., Tewksbury, A. B., Matei, M. T., Nagy, L. M., and Williams, T. A. 2015. Changing cell behaviours during beetle embryogenesis correlates with slowing of segmentation. *Nat. Commun.*, 6(6635): 1–9. (Cited on pages 7, 11, 99, 112, 171, and 172.)
- Nakaya, M.-a., Biris, K., Tsukiyama, T., Jaime, S., Rawls, J. A., and Yamaguchi, T. P. 2005. Wnt3a links left-right determination with segmentation and anteroposterior axis elongation. *Development*, 132(24): 5425–5436. (Cited on pages 58 and 80.)
- Namigai, E. K., Kenny, N. J., and Shimeld, S. M. 2014. Right across the tree of life: The evolution of left-right asymmetry in the Bilateria. *Genesis*, 52(6): 458–470. (Cited on page 170.)
- Niehrs, C. 2010. On growth and form: a cartesian coordinate system of Wnt and BMP signaling specifies bilaterian body axes. *Development*, 137(6): 845–857. (Cited on pages 100, 106, and 121.)
- Nikolaidou, K. K. and Barrett, K. 2005. Getting to know your neighbours; a new mechanism for cell intercalation. *Trends Genet.*, 21(2): 70 – 73. (Cited on page 11.)

- Ninomiya, H. and Winklbauer, R. 2008. Epithelial coating controls mesenchymal shape change through tissue-positioning effects and reduction of surface-minimizing tension. *Nat. Cell Biol.*, 10: 61–69. (Cited on page 149.)
- Ninomiya, H., Elinson, R. P., and Winklbauer, R. 2004. Antero-posterior tissue polarity links mesoderm convergent extension to axial patterning. *Nature*, 430: 364–367. (Cited on pages 134, 145, and 147.)
- Ninomiya, H., David, R., Damm, E. W., Fagotto, F., Niessen, C. M., and Winklbauer, R. 2012. Cadherin-dependent differential cell adhesion in *xenopus* causes cell sorting in vitro but not in the embryo. *J. Cell Sci.*, 125(8): 1877–1883. (Cited on pages 145 and 150.)
- Niwa, Y., Shimojo, H., Isomura, A., González, A., Miyachi, H., and Kageyama, R. 2011. Different types of oscillations in Notch and Fgf signaling regulate the spatiotemporal periodicity of somitogenesis. *Genes Dev.*, 25(11): 1115–1120. (Cited on pages 9, 10, 58, 59, 73, 74, 78, 83, 87, and 168.)
- Oginuma, M., Niwa, Y., Chapman, D. L., and Saga, Y. 2008. *Mesp2* and *Tbx6* cooperatively create periodic patterns coupled with the clock machinery during mouse somitogenesis. *Development*, 135(15): 2555–2562. (Cited on pages 10 and 74.)
- Oginuma, M., Takahashi, Y., Kitajima, S., Kiso, M., Kanno, J., Kimura, A., and Saga, Y. 2010. The oscillation of Notch activation, but not its boundary, is required for somite border formation and rostral-caudal patterning within a somite. *Development*, 137(9): 1515–1522. (Cited on pages 10 and 174.)
- Özbudak, E. M. and Lewis, J. 2008. Notch signalling synchronizes the zebrafish segmentation clock but is not needed to create somite boundaries. *PLoS Genet.*, 4(2): 1–11. (Cited on page 57.)
- Palmeirim, I., Henrique, D., Ish-Horowicz, D., and Pourquié, O. 1997. Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell*, 91(5): 639 – 648. (Cited on pages 2, 22, 48, and 57.)
- Pankratz, M. J. and Jäckle, H. 1990. Making stripes in the *Drosophila* embryo. *Trends Genet.*, 6(0): 287 – 292. (Cited on page 106.)
- Paré, A. C., Vichas, A., Fincher, C. T., Mirman, Z., Farrell, D. L., Mainieri, A., and Zallen, J. A. 2014. A positional toll receptor code directs convergent extension in *Drosophila*. *Nature*, 515: 523–527. (Cited on pages 12 and 171.)
- Peel, A. 2004. The evolution of arthropod segmentation mechanisms. *BioEssays*, 26(10): 1108–1116. (Cited on pages 3, 7, 8, 14, and 99.)
- Peel, A. and Akam, M. 2003. Evolution of segmentation: Rolling back the clock. *Curr. Biol.*, 13(18): R708 – R710. (Cited on pages 2, 21, and 99.)
- Petersen, C. P. and Reddien, P. W. 2009. Wnt signaling and the polarity of the primary body axis. *Cell*, 139(6): 1056 – 1068. (Cited on page 121.)
- Pourquié, O. 2011. Vertebrate segmentation: From cyclic gene networks to scoliosis. *Cell*, 145(5): 650 – 663. (Cited on page 57.)
- Rauzi, M., Verant, P., Lecuit, T., and Lenne, P-F. 2008. Nature and anisotropy of cortical forces orienting *Drosophila* tissue morphogenesis. *Nat. Cell Biol.*, 10: 1401–1410. (Cited on pages 13, 133, 147, 148, and 151.)
- Rauzi, M., Lenne, P-F., and Lecuit, T. 2010. Planar polarized actomyosin contractile flows control epithelial junction remodelling. *Nature*, 468: 1110–1114. (Cited on pages 13 and 133.)

- Raya, Á., Kawakami, Y., Rodríguez-Esteban, C., Büscher, D., Koth, C. M., Itoh, T., Morita, M., Raya, R. M., Dubova, I., Bessa, J. G., de la Pompa, J. L., and Belmonte, J. C. I. 2003. Notch activity induces Nodal expression and mediates the establishment of left-right asymmetry in vertebrate embryos. *Genes Dev.*, 17(10): 1213–1218. (Cited on pages 58 and 80.)
- Resende, T. P., Ferreira, M., Teillet, M.-A., Tavares, A. T., Andrade, R. P., and Palmeirim, I. 2010. Sonic hedgehog in temporal control of somite formation. *PNAS*, 107(29): 12907–12912. (Cited on pages 82 and 86.)
- Resende, T. P., Andrade, R. P., and Palmeirim, I. 2014. Timing embryo segmentation: Dynamics and regulatory mechanisms of the vertebrate segmentation clock. *Biomed Res. Int.*, 2014: 12. (Cited on page 57.)
- Richmond, D. L. and Oates, A. C. 2012. The segmentation clock: inherited trait or universal design principle? *Curr. Opin. Genet. Dev.*, 22(6): 600 – 606. Genetics of system biology. (Cited on page 99.)
- Ridley, A. J., Schwartz, M. A., Burridge, K., Firtel, R. A., Ginsberg, M. H., Borisy, G., Parsons, J. T., and Horwitz, A. R. 2003. Cell migration: Integrating signals from front to back. *Science*, 302(5651): 1704–1709. (Cited on pages 135 and 141.)
- Rivera, A. S. and Weisblat, D. A. 2009. And lophotrochozoa makes three: Notch/Hes signaling in annelid segmentation. *Dev. Genes Evol.*, 219(1): 37–43. (Cited on page 99.)
- Saga, Y. 2012. The mechanism of somite formation in mice. *Curr. Opin. Genet. Dev.*, 22(4): 331 – 338. Developmental mechanisms, patterning and evolution. (Cited on pages 70 and 75.)
- Salazar-Ciudad, I. and Jernvall, J. 2010. A computational model of teeth and the developmental origins of morphological variation. *Nature*, 464: 583–586. (Cited on page 4.)
- Salazar-Ciudad, I., Newman, S. A., and Solé, R. V. 2001a. Phenotypic and dynamical transitions in model genetic networks i. Emergence of patterns and genotype-phenotype relationships. *Evol. Dev.*, 3(2): 84–94. (Cited on pages 6, 15, 28, 105, 120, 173, and 176.)
- Salazar-Ciudad, I., Solé, R. V., and Newman, S. A. 2001b. Phenotypic and dynamical transitions in model genetic networks ii. Application to the evolution of segmentation mechanisms. *Evol. Dev.*, 3(2): 95–103. (Cited on page 28.)
- Sarrazin, A. F., Peel, A. D., and Averof, M. 2012. A segmentation clock with two-segment periodicity in insects. *Science*, 336(6079): 338–341. (Cited on pages 2, 7, 112, 134, 147, 150, and 172.)
- Saúde, L., Lourenço, R., Gonçalves, A., and Palmeirim, I. 2005. Terra is a left-right asymmetry gene required for left-right synchronization of the segmentation clock. *Nat. Cell Biol.*, 7: 918–920. (Cited on page 64.)
- Schaerli, Y., Munteanu, A., Gili, M., Cotterell, J., Sharpe, J., and Isalan, M. 2014. A unified design space of synthetic stripe-forming networks. *Nat. Commun.*, 5(4905): 1–10. (Cited on page 28.)
- Schröter, C., Ares, S., Morelli, L. G., Isakova, A., Hens, K., Soroldoni, D., Gajewski, M., Jülicher, F., Maerkl, S. J., Deplancke, B., and Oates, A. C. 2012. Topology and dynamics of the zebrafish segmentation clock core circuit. *PLoS Biol.*, 10(7): e1001364. (Cited on page 8.)
- Schubert, M., Holland, L. Z., Stokes, M., and Holland, N. D. 2001. Three amphioxus Wnt genes (AmphiWnt3, AmphiWnt5, and AmphiWnt6) associated with the tail bud: the evolution of somitogenesis in chordates. *Dev. Biol.*, 240(1): 262 – 273. (Cited on

- page 84.)
- Seaver, E. C. 2003. Segmentation: mono- or polyphyletic? *Int. J. Dev. Biol.*, 47: 583 – 595. (Cited on pages 13, 21, and 99.)
- Sepich, D. S., Calmelet, C., Kiskowski, M., and Solnica-Krezel, L. 2005. Initiation of convergence and extension movements of lateral mesoderm during zebrafish gastrulation. *Dev. Dynam.*, 234(2): 279–292. (Cited on page 133.)
- Shankland, M. and Seaver, E. C. 2000. Evolution of the bilaterian body plan: What have we learned from annelids? *PNAS*, 97(9): 4434–4437. (Cited on pages 21 and 99.)
- Shih, N. P., François, P., Delaune, E. A., and Amacher, S. L. 2015. Dynamics of the slowing segmentation clock reveal alternating two-segment periodicity. *Development*, 142(10): 1785–1793. (Cited on pages 3, 9, 47, 48, 59, 69, 71, 83, 87, and 174.)
- Shimizu, T. and Nakamoto, A. 2001. Segmentation in annelids: Cellular and molecular basis for metameric body plan. *Zool. Sci.*, 18: 285–298. (Cited on page 7.)
- Shinbrot, T., Chun, Y., Caicedo-Carvajal, C., and Foty, R. 2009. Cellular morphogenesis in silico. *Biophys. J.*, 97(4): 958 – 967. (Cited on page 148.)
- Sirbu, I. O. and Duester, G. 2006. Retinoic-acid signalling in node ectoderm and posterior neural plate directs left-right patterning of somitic mesoderm. *Nat. Cell Biol.*, 8: 271–277. (Cited on pages 58, 65, and 66.)
- Skoglund, P. and Keller, R. 2010. Integration of planar cell polarity and ECM signaling in elongation of the vertebrate body plan. *Curr. Opin. Cell Biol.*, 22(5): 589 – 596. (Cited on page 133.)
- Solé, R. V., Salazar-Ciudad, I., and Garcia-Fernández, J. 2002. Common pattern formation, modularity and phase transitions in a gene network model of morphogenesis. *Physica A*, 305(3?4): 640 – 654. (Cited on page 5.)
- Soroldoni, D., Jörg, D. J., Morelli, L. G., Richmond, D. L., Schindelin, J., Jülicher, F., and Oates, A. C. 2014. A Doppler effect in embryonic pattern formation. *Science*, 345(6193): 222–225. (Cited on page 48.)
- Soza-Ried, C., Öztürk, E., Ish-Horowicz, D., and Lewis, J. 2014. Pulses of Notch activation synchronise oscillating somite cells and entrain the zebrafish segmentation clock. *Development*, 141(8): 1780–1788. (Cited on page 57.)
- Steventon, B., Duarte, F., Lagadec, R., Mazan, S., Nicolas, J.-F., and Hirsinger, E. 2016. Species-specific contribution of volumetric growth and tissue convergence to posterior body elongation in vertebrates. *Development*, 143(10): 1732–1741. (Cited on pages 11 and 172.)
- Tada, M. and Heisenberg, C.-P. 2012. Convergent extension: using collective cell migration and cell intercalation to shape embryos. *Development*, 139(21): 3897–3904. (Cited on pages 11, 12, 13, 133, 136, and 171.)
- Tanaka, Y., Okada, Y., and Hirokawa, N. 2005. FGF-induced vesicular release of Sonic hedgehog and retinoic acid in leftward nodal flow is critical for left-right determination. *Nature*, 435: 172–177. (Cited on page 58.)
- Tautz, D. 2004. Segmentation. *Dev. Cell*, 7(3): 301 – 312. (Cited on pages 21 and 99.)
- ten Tusscher, K. 2013. Mechanisms and constraints shaping the evolution of body plan segmentation. *Eur Phys J E*, 36(5): 1–12. (Cited on pages 3, 15, 22, 24, and 120.)
- ten Tusscher, K. H. and Hogeweg, P. 2011. Evolution of networks for body plan patterning; interplay of modularity, robustness and evolvability. *PLoS Comput. Biol.*, 7(10): e1002208. (Cited on pages 6, 10, 15, 16, 22, 23, 26, 27, 28, 33, 83, 100, 101, 103,

- 105, 108, 170, 173, 174, and 176.)
- Terry, A. J., Sturrock, M., Dale, J. K., Maroto, M., and Chaplain, M. A. J. 2011. A spatio-temporal model of Notch signalling in the zebrafish segmentation clock: Conditions for synchronised oscillatory dynamics. *PLoS ONE*, 6(2): e16980. (Cited on page 8.)
- Tiedemann, H. B., Schneltzer, E., Zeiser, S., Rubio-Aliaga, I., Wurst, W., Beckers, J., Przemeck, G. K., and de Angelis, M. H. 2007. Cell-based simulation of dynamic expression patterns in the presomitic mesoderm. *J. Theor. Biol.*, 248(1): 120 – 129. (Cited on page 8.)
- Tiedemann, H. B., Schneltzer, E., Zeiser, S., Hoesel, B., Beckers, J., Przemeck, G. K. H., and de Angelis, M. H. 2012. From dynamic expression patterns to boundary formation in the presomitic mesoderm. *PLoS Comput. Biol.*, 8(6): e1002586. (Cited on pages 10 and 69.)
- Tiedemann, H. B., Schneltzer, E., Zeiser, S., Wurst, W., Beckers, J., Przemeck, G. K. H., and de Angelis, M. H. 2014. Fast synchronization of ultradian oscillators controlled by Delta-Notch signaling with cis-inhibition. *PLoS Comput. Biol.*, 10(10): 1–12. (Cited on page 8.)
- Troein, C., Locke, J. C., Turner, M. S., and Millar, A. J. 2009. Weather and seasons together demand complex biological clocks. *Curr. Biol.*, 19(22): 1961 – 1964. (Cited on pages 6 and 173.)
- Tsiairis, C. and Aulehla, A. 2016. Self-organization of embryonic genetic oscillators into spatiotemporal wave patterns. *Cell*, 164(4): 656 – 667. (Cited on page 3.)
- Verd, B., Crombach, A., and Jaeger, J. 2014. Classification of transient behaviours in a time-dependent toggle switch model. *BMC Sys. Biol.*, 8(1): 1–19. (Cited on page 80.)
- Vermot, J. and Pourquié, O. 2005. Retinoic acid coordinates somitogenesis and left-right patterning in vertebrate embryos. *Nature*, 435: 215–220. (Cited on pages 57, 58, 60, 64, 65, and 82.)
- Vermot, J., Llamas, J. G., Fraulob, V., Niederreither, K., Chambon, P., and Dollé, P. 2005. Retinoic acid controls the bilateral symmetry of somite formation in the mouse embryo. *Science*, 308(5721): 563–566. (Cited on pages 57, 58, 60, 64, 65, 76, and 83.)
- Vilhais-Neto, G. C., Maruhashi, M., Smith, K. T., Vasseur-Cognet, M., Peterson, A. S., Workman, J. L., and Pourquié, O. 2010. Rere controls retinoic acid signalling and somite bilateral symmetry. *Nature*, 463: 953–957. (Cited on page 80.)
- von der Hardt, S., Bakkers, J., Inbal, A., Carvalho, L., Solnica-Krezel, L., Heisenberg, C.-P., and Hammerschmidt, M. 2007. The Bmp gradient of the zebrafish gastrula guides migrating lateral cells by regulating cell-cell adhesion. *Curr. Biol.*, 17(6): 475 – 487. (Cited on page 136.)
- Vroomans, R. M., Hogeweg, P., and ten Tusscher, K. H. 2016. In silico evo-devo: reconstructing stages in the evolution of animal segmentation. *EvoDevo*, accepted: 1–10. (Cited on pages 22, 28, 29, and 47.)
- Wallingford, J. B., Fraser, S. E., and Harland, R. M. 2002. Convergent extension: The molecular control of polarized cell movement during embryonic development. *Dev. Cell*, 2(6): 695 – 706. (Cited on page 133.)
- Wanglar, C., Takahashi, J., Yabe, T., and Takada, S. 2014. Tbx protein level critical for clock-mediated somite positioning is regulated through interaction between Tbx and Ripply. *PLoS ONE*, 9(9): 1–12. (Cited on pages 69, 70, 71, 83, and 86.)
- Wanninger, A., Kristof, A., and Brinkmann, N. 2009. Sipunculans and segmentation. *Commun. Integr. Biol.*, 2(1): 56–59. (Cited on page 99.)

- Webb, A. B., Lengyel, I. M., Jörg, D. J., Valentin, G., Jülicher, F., Morelli, L. G., and Oates, A. C. 2016. Persistence, period and precision of autonomous cellular oscillators from the zebrafish segmentation clock. *eLife*, 5: e08438. (Cited on pages 10 and 69.)
- Weliky, M., Minsuk, S., Keller, R., and Oster, G. 1991. Notochord morphogenesis in *Xenopus laevis*: simulation of cell behavior underlying tissue convergence and extension. *Development*, 113(4): 1231–1244. (Cited on pages 13, 133, 134, 147, and 148.)
- Williams, T., Blachuta, B., Hegna, T. A., and Nagy, L. M. 2012. Decoupling elongation and segmentation: Notch involvement in anostracan crustacean segmentation. *Evol. Dev.*, 14(4): 372–382. (Cited on page 99.)
- Wilson, V., Olivera-Martinez, I., and Storey, K. G. 2009. Stem cells, signals and vertebrate body axis extension. *Development*, 136(12): 2133–2133. (Cited on pages 11 and 171.)
- Wolpert, L. 2007. *Principles of Development*. Oxford University Press. (Cited on page 2.)
- Yabe, T. and Takada, S. 2016. Molecular mechanism for cyclic generation of somites: Lessons from mice and zebrafish. *Dev. Growth Differ.*, 58(1): 31–42. (Cited on pages 68, 70, 74, and 75.)
- Yin, C., Ciruna, B., and Solnica-Krezel, L. 2009. Chapter 7: Convergence and extension movements during vertebrate gastrulation. In T. Lecuit, editor, *Current Topics in Developmental Biology*, volume 89, pages 163 – 192. Academic Press. (Cited on page 133.)
- Young, T., Rowland, J. E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J.-N., Beck, F., Mallo, M., and Deschamps, J. 2009. Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev. Cell*, 17(4): 516 – 526. (Cited on page 122.)
- Zajac, M., Jones, G. L., and Glazier, J. A. 2003. Simulating convergent extension by way of anisotropic differential adhesion. *J. Theor. Biol.*, 222(2): 247 – 259. (Cited on pages 13, 133, 134, 136, 147, 148, and 154.)
- Zákány, J., Kmita, M., Alarcon, P., de la Pompa, J.-L., and Duboule, D. 2001. Localized and transient transcription of Hox genes suggests a link between patterning and the segmentation clock. *Cell*, 106(2): 207 – 217. (Cited on page 172.)
- Zallen, J. A. and Wieschaus, E. 2004. Patterned gene expression directs bipolar planar polarity in *Drosophila*. *Dev. Cell*, 6(3): 343 – 355. (Cited on pages 12, 13, 134, 147, 151, and 171.)
- Zhang, Y., Thomas, G. L., Swat, M., Shirinifard, A., and Glazier, J. A. 2011. Computer simulations of cell sorting due to differential adhesion. *PLoS One*, 6(10): e24999. (Cited on pages 145 and 146.)
- Ángel Raya, Kawakami, Y., Rodríguez-Esteban, C., Ibañes, M., Rasskin-Gutman, D., Rodríguez-León, J., Büscher, D., Feijó, J. A., and Belmonte, J. C. I. 2004. Notch activity acts as a sensor for extracellular calcium during vertebrate left-right determination. *Nature*, 427: 121–128. (Cited on page 58.)

Samenvatting

Segmenten zijn herhaalde structuren langs de hoofd-staartas, die bij veel dieren voorkomen: denk bijvoorbeeld aan de ringen van regenwormen, de ringen met pootjes van duizendpoten of de ruggenwervels van gewervelden. Deze structuren worden al zeer vroeg tijdens de embryonale ontwikkeling aangelegd. Meestal worden ze één voor één geproduceerd vanuit een groeizone, een weefsel dat aan het begin van de ontwikkeling net achter het hoofd ligt, en hier steeds verder vanaf komt te liggen naarmate er meer segmenten zijn gevormd – dit wordt sequentiële segmentatie genoemd. Voordat segmenten hun uiteindelijke vorm krijgen, worden de grenzen tussen segmenten en tussen de verschillende delen van elk segment vastgesteld door een gestreept patroon van verschillende segmentatiegenen in het embryonale weefsel. In insecten en gewervelden¹ gebeurt dit door de interactie tussen een moleculaire klok en een morfogengradiënt in de cellen in het weefsel dat wordt gesegmenteerd. De klok bestaat uit meerdere genen, wiens expressie afwisselend hoog en laag is (oscilleert): de periode van deze oscillaties bepaalt het tempo waarmee segmenten gevormd worden. Het morfogen is een speciaal eiwit dat cellen informeert over hun plek in het te segmenteren weefsel. Dit eiwit heeft een hoge concentratie in de groeizone, maar neemt af richting het hoofd waar de reeds gevormde segmenten liggen. Waar de concentratie van het morfogeneiwit laag genoeg is, worden de oscillaties van de klok omgezet in een gedifferentieerd segment. Het mechanisme waarmee dit gebeurt staat echter nog steeds ter discussie. Ook over de evolutionaire oorsprong van segmenten is nog veel onduidelijk: was de voorouder van de segmenteerde dieren ook gesegmenteerd? Welke evolutionaire selectiedrukken waren verantwoordelijk voor het ontstaan van de huidige segmentatiemechanismen?

In het onderzoek naar segmentatie worden regelmatig computermodellen gebruikt om hypothesen te testen. Computermodellen zijn uitermate geschikt om complexe mechanismen mee te ontrafelen en om evolutionaire hypothesen te testen; in een computer kan een evolutionair proces van duizenden jaren worden gesimuleerd in enkele dagen, en vele malen herhaald onder verschillende condities. In dit proefschrift gebruiken we dan ook computermodellen om de evolutie en het mechanisme van sequentiële segmentatie te onderzoeken.

¹Segmenten worden op een heel andere manier aangelegd in annelide wormen. Hier is minder over bekend, en we behandelen het niet in dit proefschrift.

In het eerste deel richten we ons op specifieke kenmerken van segmentatie in gewervelden, en in het tweede deel op de algemene relatie tussen het segmentatiepatroon en de groei van het weefsel.

In zowel insecten als gewervelden zijn een klok en morfogengradiënt betrokken bij segmentatie, maar er zijn verschillen: in gewervelden is de morfogengradiënt lang en gradueel, terwijl deze in insecten een stuk steiler en korter lijkt te zijn. Ook suggereren de beschikbare data dat de moleculaire klok in gewervelden complexer is (uit meer interacterende genen bestaat) dan in insecten. In hoofdstuk 2 onderzoeken we of en hoe de lengte en steilheid van de morfogengradiënt de evolutie van de klok beïnvloedt wat betreft het aantal betrokken genen en de dynamica van de oscillaties. We zien dat een langere, minder steile (flauwe) gradiënt de evolutie van segmenten ietwat bemoeilijkt, en leidt tot de evolutie van grotere genregulatiernetwerken. Bovendien leidt het net iets vaker tot de spontane evolutie van een zogeheten frequentiegradiënt, waarbij oscillaties langzamer worden bij een lagere morfogen-concentratie voordat de cellen differentiëren - dit is een kenmerkende eigenschap van segmentatie in gewervelden. Evolutie leidt vaker tot een frequentiegradiënt als er noise (ruis) zit in de genexpressie en de morfogengradiënt lang en flauw is, wat suggereert dat gewervelden mogelijk een frequentiegradiënt evolueerden om het effect van ruis te verminderen. Onze resultaten suggereren dat insecten, met hun steile morfogengradiënt, minder kans hadden om een frequentiegradiënt en complexe klok te evolueren.

Een ander essentieel kenmerk van segmentatie in gewervelden is de gelijktijdige, symmetrische aanleg van segmenten aan beide zijden van de notochord². Als echter een van de betrokken morfogenen, retinoic acid (retinezuur), afwezig is, dan worden segmenten asymmetrisch gevormd onder invloed van het links-rechts-differentiatiesysteem, met ernstige gevolgen voor de verdere ontwikkeling van het dier. Maar de ernst van het asymmetrische fenotype verschilt tussen diersoorten: in de zebravis is er sprake van een tijdelijke vertraging in de segmentvorming aan de rechterkant, die later netjes wordt opgelost, terwijl in muis de segmenten ook scheef worden aangelegd, of soms zelfs afwezig blijven. In hoofdstuk 3 laten wij zien hoe deze verschillen kunnen worden verklaard aan de hand van verschillen in het segmentatie-mechanisme van deze diersoorten. Wij concluderen dat de verschillen in segmentatiemechanisme tussen de soorten niet volledig functioneel neutraal zijn. We wijzen in dit hoofdstuk op onbekende factoren in de mechanismen van verschillende diersoorten, en doen suggesties voor de interacties tussen het links-rechts differentiatiesysteem en segmentatie.

²een centrale buis die later de kern van de ruggenwervel vormt

Zoals we eerder vermeldden, is sequentiële segmentatie de meest voorkomende vorm van segmentontwikkeling. In hoofdstuk 4 onderzoeken we wat voor condities nodig zijn om sequentiële segmentatie inderdaad de meest waarschijnlijke evolutionaire uitkomst te laten zijn in onze simulaties, waarin zowel de locatie van celdelingen als het genexpressiepatroon kunnen evolueren. Met andere woorden: we willen achterhalen onder welke omstandigheden sequentiële segmentatie met een groeizone is geëvolueerd, en in welke volgorde bepaalde eigenschappen zijn ontstaan. Onze simulaties laten bijvoorbeeld zien dat een morfogengradiënt vanuit de staart noodzakelijk is voor de evolutie van sequentiële segmentatie met een groeizone in de staart: zonder zo'n gradiënt evolueert een minder robuuste vorm van segmentatie waarbij alle segmenten tegelijkertijd gevormd worden. Als de morfogengradiënt wel aanwezig is, wordt sequentiële segmentatie ruim de dominante evolutionaire uitkomst. We verliezen deze dominantie echter weer als er ook een selectiedruk is voor het op tijd stoppen met celdelingen aan het einde van *in silico* segmentontwikkeling (gedetermineerde groei). We concluderen dat voor de evolutie van sequentiële segmentatie, een staartzone met een duidelijke identiteit en de potentie voor het produceren van een morfogengradiënt aanwezig moet zijn geweest, en dat gedetermineerde groei waarschijnlijk niet tegelijk met sequentiële groei is geëvolueerd.

In hoofdstuk 5 bekijken we de interactie tussen het gestreepte genpatroon en de vorming van de lichaams-as. We laten zien dat eerder voorgestelde mechanismen voor convergente extensie (het langer en smaller worden van weefsel) een bestaand segmentatiepatroon kunnen verstoren. Als cellen met dezelfde identiteit in een segment sterker aan elkaar hechten dan aan andere cellen, blijft het patroon wel behouden; sterker nog, deze "segment-specifieke adhesie" kan zelf zorgen voor extensie van de lichaamsas. Als we bovendien toestaan dat cellen bewegen, kan deze adhesie deze beweging omzetten in convergente extensie. Dit resultaat suggereert dat vergaande interacties tussen de vorming en segmentatie van de lichaamsas mogelijk zijn en meer aandacht verdienen in experimenten.

De resultaten in dit proefschrift geven nieuwe inzichten in het mechanisme en de evolutie van segmentatie, en wijzen op nieuwe richtingen voor verder onderzoek.



Curriculum Vitæ

Renske Vroomans was born in Den Haag, The Netherlands on June 21st 1988. She attended the St Maartenscollege in Voorburg, where she obtained her Gymnasium diploma *cum laude* in June 2006. In the same year she started her bachelor's degree in Biology at Utrecht University, graduating *cum laude* in February 2010. She then pursued a master's degree in the Theoretical Biology and Bioinformatics group, conducting a research project on T cell migration in lymph nodes under the supervision of prof.dr. Rob de Boer and dr. Joost Beltman. At the Computational Biology group of the John Innes Centre in Norwich (UK), she performed a minor research project on auxin transport in *Arabidopsis* fruit with dr. Stan Marée and dr. Verônica Grieneisen. In October 2011, she obtained her master's degree *cum laude* and started her PhD research at the Theoretical Biology & Bioinformatics group at Utrecht University, under the supervision of dr. Kirsten ten Tusscher and Prof.dr. Paulien Hogeweg. The results of her PhD research are described in this thesis.



List of Publications

Renske M.A. Vroomans, Athanasius F. M. Marée, Rob J. de Boer, Joost B. Beltman. Chemotactic Migration of T Cells towards Dendritic Cells Promotes the Detection of Rare Antigens. *PLoS Computational Biology*; 11(8)

Renske M.A. Vroomans, Paulien Hogeweg, Kirsten H.W.J. ten Tusscher. Segment-specific adhesion as a driver of convergent extension. 2015. *PLoS Computational Biology*; 11(2)

Renske M.A. Vroomans, Paulien Hogeweg, Kirsten H.W.J. ten Tusscher. In silico evo-devo: reconstructing stages in the evolution of animal segmentation. 2016. *EvoDevo*, 7:14

Renske M.A. Vroomans, Kirsten H.W.J. ten Tusscher. Modelling asymmetric somitogenesis: a pointer to potentially non-neutral species differences 2016. *Submitted*

Renske M.A. Vroomans, Paulien Hogeweg, Kirsten H.W.J. ten Tusscher. Analysing the influence of wavefront dynamics on clock evolution. 2016. *In preparation*



Acknowledgements

First of all, I would like to thank my co-promotor Kirsten ten Tusscher; Kirsten, I believe I was your first ‘real’ PhD student, to be supervised by you from the start. Thank you for the boundless patience you had with me and my bugs (not just those in my simulations). Your thorough, honest and ever-critical attitude to science has been a great example. I wish you all the best with your group working on root modelling – and who knows, oscillations may again become a theme one day!

Paulien Hogeweg, in the first-year Systems Biology course you introduced a whole new range of amazing research with models like CPM. After this, and the (in)famous third-year Systems Biology course, I was sold. I am honoured to have had you as my promotor; thank you for your many inspiring ideas and for your unique scientific vision.

Writing a thesis is pointless unless people read it, and I am therefore very grateful to my reading committee – dr. Paul François, prof. Roeland Merks, prof. Kunihiko Kaneko, prof. Andrew Oates and dr. James Sharpe. They did not just read my thesis, but also performed much of the research on which my thesis is built. Special thanks to dr. Merks for providing extensive commentary on the first two chapters. Hilje and Jaap helped out with my Dutch summary, and my fourth chapter was scrutinised by the evolutionary journal club of the TBB: Rutger, Thomas, Bram, Hilje en Sandro. Ioana, thank you for the discussions on CPM, and for reading what is now the fifth chapter. Sophia was my master student for over a year, continuing the work in my second chapter in a “2D setting”. Her efforts also breathed new life into my 1D simulations.

It was a great privilege to be a part of the TBB group, a rare collection of bright and interested people who care deeply about science and the world in general. These people and their research are supported by Jan Kees, whose efforts allowed me to run large numbers of simulations (sorry for the thousands of .pngs). Not to mention all those past and present group members who contributed to the lovely atmosphere: thank you all!

Outside of academia, my family and friends were the best support I could possibly wish for. Mam, Pap, jullie hebben mij begeleid op mijn pad naar de wetenschap vanaf het moment dat ik kon lezen – dank jullie voor dat grote geschenk. En Lex, voor de broodnodige realiteitszin als ik weer eens te lang met mijn hoofd in de boeken had gezeten.

Finally, my dear Sandro... thank you for sharing this road with me. I look forward to our new scientific and personal adventures together.

