Evolving developmental trajectories to generate phenotypic diversity in Lake Malawi Cichlids

Shannon Taylor

St Hilda's College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

April 2025

Abstract

A significant question in evolutionary developmental biology is how morphogenesis and gene regulation interact to achieve phenotypic diversity. In this thesis, I tackle this question using two complementary approaches.

First, I investigate how morphological diversity is generated in the near-absence of genetic diversity. Lake Malawi cichlid fishes are an excellent system with which to study this problem, varying in phenotype but having extremely limited genetic diversity. In particular, cichlid species vary in somite and vertebral number, with *Astatotilapia calliptera* and *Rhamphochromis chillingali* forming 32 and 40 somites respectively. I characterized in detail the process of axial development during somitogenesis in both species, and found that both species differ only at the onset of segmentation: at the early stages of segmentation, the *R. chillingali* PSM is longer, greater in volume, and contains more cells, than that of *A. calliptera*. Other aspects of development, such as the anatomy and gene expression patterning of the PSM, are identical between these species. This suggests that *R. chillingali* have evolved an increased somite count by evolving the initial conditions of somitogenesis, and not the process of somitogenesis itself. This suggests that minor modifications of initial conditions can result in dramatic phenotypic innovations.

Secondly, I investigate how morphogenesis in the form of cell movements influences pattern formation during zebrafish somitogenesis. I develop and validate a method to reverse-engineer gene regulatory networks in a tissue undergoing extensive cell movements and rearrangements: the zebrafish presomitic mesoderm (PSM) during T-box pattern formation. Combining static and dynamic microscopy data, I reverse-engineer GRNs which are able to recapitulate key aspects of the T-box GRN, including pattern formation, the sign of known genetic interactions, and the response to perturbations in signalling. Characterizing these networks reveals that the tbox GRN could act as a cell-autonomous timer of differentiation as cells adopt a somitic fate. This finding explains a surprising ability of cells in the zebrafish PSM, to autonomously differentiate and upregulate *tbx6/mesp-ba* expression when isolated from the tailbud.

Altogether, this investigates how morphogenesis and genetic processes co-operate to produce phenotypic diversity during embryonic development.

Evolving developmental trajectories to generate phenotypic diversity in Lake Malawi Cichlids



Shannon Taylor

St Hilda's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

April 2025

Acknowledgements

This thesis would not have been possible without the assistance of many people.

I must first thank my supervisor, Dr Berta Verd, for being such an inspiring and supportive mentor for the duration of my studies. I have learnt a huge amount from you.

I also want to thank James Hammond, Callum Bucklow, Georgina Stooke-Vaughan, and other members of the Verd lab: it has been a lot of fun working with you and I have learnt a lot from you all.

I am also incredibly grateful to my family: my parents, my sister, and my extended Whanau. You have supported me both emotionally and practically during my PhD, and you have made me the person I am today.

My PhD would not have been possible without generous funding: the Clarendon scholarship made it possible for me to study at Oxford, and the William Georgetti fund made it possible for me to move to the UK. I am immensely grateful for this opportunity.

This work would also have been impossible without the tireless work of our fish technicians, Helen Saunders, Christine Sloper, and Allex Turner, who came in every day, even Christmas, to look after our cichlids. Thank you so much for looking after our fish so well.

I was very lucky to find wonderful communities at Oxford. I am particularly grateful to the Broken Spoke, especially Abha, Charlie, Loïse, Jonathan, Inês, Alice, Frank, and the rest of Team BikeReCycling for being such a lovely and welcoming community — and for training me how to fix bikes!

I had companions during the writing process - Stellar-boo the cat, as well as pīwakawaka, korimako, and tūī. Thank you for your complete indifference to my PhD!

I also want to thank my former supervisor Peter Dearden, and other members of the Laboratory for Evolution and Development in Dunedin.

Abstract

A significant question in evolutionary developmental biology is how morphogenesis and gene regulation interact to achieve phenotypic diversity. In this thesis, I tackle this question using two complementary approaches.

First, I investigate how morphological diversity is generated in the near-absence of genetic diversity. Lake Malawi cichlid fishes are an excellent system with which to study this problem, varying in phenotype but having extremely limited genetic diversity. In particular, cichlid species vary in somite and vertebral number, with *Astatotilapia calliptera* and *Rhamphochromis chillingali* forming 32 and 40 somites respectively. I characterized in detail the process of axial development during somitogenesis in both species, and found that both species differ only at the onset of segmentation: at the early stages of segmentation, the *R. chillingali* PSM is longer, greater in volume, and contains more cells, than that of *A. calliptera*. Other aspects of development, such as the anatomy and gene expression patterning of the PSM, are identical between these species. This suggests that *R. chillingali* have evolved an increased somite count by evolving the initial conditions of somitogenesis, and not the process of somitogenesis itself. This suggests that minor modifications of initial conditions can result in dramatic phenotypic innovations.

Secondly, I investigate how morphogenesis in the form of cell movements influences pattern formation during zebrafish somitogenesis. I develop and validate a method to reverse-engineer gene regulatory networks in a tissue undergoing extensive cell movements and rearrangements: the zebrafish presomitic mesoderm (PSM) during T-box pattern formation. Combining static and dynamic microscopy data, I reverse-engineer GRNs which are able to recapitulate key aspects of the T-box GRN, including pattern formation, the sign of known genetic interactions, and the response to perturbations in signalling. Characterizing these networks reveals that the tbox GRN could act as a cell-autonomous timer of differentiation as cells adopt a somitic fate. This finding explains a surprising ability of cells in the zebrafish PSM, to autonomously differentiate and upregulate *tbx6/mesp-ba* expression when isolated from the tailbud.

Altogether, this investigates how morphogenesis and genetic processes co-operate to produce phenotypic diversity during embryonic development.

Contents

List of Figures viii

List of Tables x

List of Abbreviations xi

1 Introduction 1

1.1 Evolvability of somite counts . . . 3
1.1.1 Conceptualising the evolution of somite counts using the Clock and Wavefront model . . . 4
1.1.2 Mechanisms of elongation differ between species . . . 6
1.1.3 Patterning of the PSM . . . 7
1.1.4 Co-operation between morphogenesis and patterning . . . 9
1.2 Studying pattern formation alongside morphogenesis . . . 10
1.3 Lake Malawi cichlids: a system to study the evolution of phenotypic diversity . . . 12
1.3.1 Adaptive radiation in Lake Malawi | Lake Nyasa | Lago Niassa . 13
1.3.2 Development of *A. calliptera* and other cichlid species . . . 15
1.3.3 Genetic and embryological tools in *A. calliptera* and *R. chillingali* . 16
1.4 Aims of this thesis . . . 17

2 Materials and Methods 19

2.1 Wet-lab methods . . . 19
2.1.1 Animal husbandry . . . 19
2.1.2 Extraction and fixation of cichlid embryos . . . 19
2.1.3 Hybridization chain reaction . . . 19
2.1.4 Phalloidin staining . . . 21
2.1.5 Imaging . . . 21
2.2 Modelling methods . . . 22
2.2.1 AGET construction . . . 22
2.2.2 Model formulation . . . 23
2.2.3 Model fitting: MCMC . . . 24
2.2.4 Calculating a likelihood score . . . 25
2.2.5 Filtering and clustering parameter sets for analysis . . . 26
2.2.6 Identification and classification of fixed points in the tbox GRN . 26
2.3 Image analysis methods . . . 27

2.3.1 Morphometrics . . . 27
2.3.2 Manual tissue segmentation . . . 28
2.3.3 Measuring patterning across the AP axis . . . 28
2.3.4 Cell counting . . . 29

3 Characterization of axial elongation during somitogenesis in the Lake Malawi cichlid *A. calliptera* **30**

3.1 Introduction . . . 31
3.1.1 Cichlids are an important system for understanding evolution and development . . . 31
3.1.2 Prior knowledge of *A. calliptera* development . . . 31
3.1.3 Evo-devo of teleost axial elongation . . . 32
3.1.4 Mechanisms of vertebrate axial elongation . . . 33
3.1.5 Patterning of the PSM . . . 38
3.2 Results . . . 40
3.2.1 Formation of the body axis: 0ss . . . 41
3.2.2 Structural integrity of the PSM arises at 2-4ss . . . 47
3.2.3 The early segmentation stage: 5-15ss . . . 49
3.2.4 Separation of the PSM from the lateral plate mesoderm: 14-16ss . . . 52
3.2.5 Distinct cell types in the embryo: 20ss . . . 55
3.2.6 Late segmentation: 25ss onwards . . . 56
3.2.7 Cell density in the PSM . . . 57
3.2.8 Patterning of the *A. calliptera* midline . . . 59
3.2.9 Oct4 patterning of the *A. calliptera* neural tube . . . 61
3.2.10 Gene expression patterning of the PSM . . . 63
3.3 Discussion . . . 67
3.3.1 The morphology of the *A. calliptera* PSM changes over time . . . 68
3.3.2 *A. calliptera* differs to zebrafish in patterning . . . 69
3.3.3 *A. calliptera* elongate via compaction-extension . . . 70
3.4 Conclusions . . . 71

4 Evolving somite counts without evolving the segmentation clock **72**

4.1 Introduction . . . 73
4.2 Results . . . 77
4.2.1 PSM size differs at the onset of somitogenesis between *R. chillingali* and *A. calliptera* . . . 77
4.2.2 Structural anatomy of the tailbud is similar between species . . . 81
4.2.3 Molecular anatomy of the tailbud is similar between species . . . 84
4.2.4 Anatomy of the body axis and the PSM is indistinguishable between species at the onset of segmentation . . . 86
4.3 Discussion . . . 87
4.3.1 Cichlids may have evolved somite counts by modifying the initial conditions of somitogenesis . . . 87
4.3.2 A unique mechanism of epiboly in African cichlids may contribute to morphological evolution . . . 89
4.3.3 Evolution via boundary conditions of embryogenesis . . . 91

4.4 Conclusions . . . 92

5 Approximated gene expression trajectories for gene regulatory network inference on cell tracks **93**

5.1 Introduction . . . 94
5.1.1 Studying pattern formation in the absence of morphogenesis . . . 95
5.1.2 Zebrafish tbox patterning is an ideal system to study the interaction of patterning and morphogenesis . . . 96
5.1.3 Live modelling overview and prior work . . . 100
5.2 Results . . . 102
5.2.1 Different AGET construction methods yield similar results. . . . 102
5.2.2 Optimal fits can be obtained using as few as 10% of AGETs for reverse engineering. . . . 102
5.2.3 Network Validation . . . 107
5.2.4 Parameter sets reproduce gene expression at the single cell level . 108
5.2.5 Parameter sets recapitulate tissue-level gene expression patterning 110
5.2.6 Parameter sets predict known genetic interactions . . . 112
5.2.7 Parameter sets recapitulate some genetic perturbations . . . 115
5.2.8 Parameter sets can reproduce upregulation of *tbx6 in silico* . . . 117
5.3 Discussion . . . 119
5.3.1 Parameter sets recapitulate key biological behaviours of the tbox GRN . . . 119
5.3.2 "Live modelling" allows one to investigate patterning in a tissue undergoing morphogenesis . . . 120
5.3.3 Future work to model the tbox GRN in zebrafish . . . 122
5.4 Conclusions . . . 124

6 The role of signalling and GRN dynamics in pattern formation in the developing zebrafish PSM **125**

6.1 Introduction . . . 126
6.1.1 The zebrafish tbox GRN achieves pattern formation alongside cell movement . . . 126
6.1.2 Using dynamical systems theory to study pattern formation and differentiation . . . 128
6.2 Results . . . 130
6.2.1 Signalling modifies speed of differentiation . . . 130
6.2.2 Change in signalling as cells transit PSM do not impact pattern formation. . . . 136
6.2.3 The MAP network is bistable, allowing different long-term behaviour of cells . . . 142
6.2.4 Multiple potential dynamical mechanisms in the tbox GRN . . . 146
6.3 Discussion . . . 148
6.3.1 Two dynamical mechanisms achieving Tbox patterning . . . 149
6.3.2 The Tbox GRN: a cell-autonomous timer of differentiation? . . . 151
6.3.3 Degeneracy in the genotype-phenotype map . . . 152
6.3.4 Is positional information required for pattern maintenance in the tbox GRN? . . . 153

6.4 Conclusions 154

7 General discussion **155**

7.1 Live modelling allows one to reverse-engineer parameter sets reproducing gene expression patterning alongside morphogenesis 156
7.2 The Tbox GRN is a cell-autonomous timer of differentiation 157
7.3 A detailed description of *A. calliptera* development during axial elongation 159
7.4 Key differences between cichlid species established at the onset of segmentation 159
7.5 Overall conclusions 160

References **162**

Appendices

A Experimental appendices **177**

List of Figures

1.1 Vertebrate species develop with different relationships to the yolk and extra-embryonic structures. . . . 6
1.2 Transcriptional states in the zebrafish tailbud . . . 8

3.1 Anatomy of the embryo at prior to somite formation. . . . 42
3.2 Anatomy of the PSM at 0ss. . . . 43
3.3 Tbox gene expression in the embryo at 1ss. . . . 45
3.4 Heterogenous *Oct4/Pouf5a* expression in 1ss *A. calliptera* . . . 46
3.5 Travelling waves of *delta-c* in the early embryo. . . . 47
3.6 Anatomy of the PSM at 2-4ss. . . . 48
3.7 Anatomy of the PSM at 10ss. . . . 50
3.8 Patterning of *A. calliptera* at 5ss. . . . 51
3.9 Detail of the boundary between the PSM and the lateral plate mesoderm. 53
3.10 Separation of the PSM from the lateral plate mesoderm . . . 54
3.11 Anatomy of the PSM at mid segmentation. . . . 55
3.12 MyoD expression in adaxial cells . . . 56
3.13 Anatomy of the PSM at late segmentation . . . 57
3.14 Cell density increases over time, and differs over space at early segmentation. . . . 58
3.15 Expression of *noto* in *A. calliptera*. . . . 60
3.16 Gene expression within Kuppfer's vesicle . . . 61
3.17 Patterning of the neural tube of *A. calliptera*. . . . 62
3.18 Expression of the tbox genes retracts over time. . . . 64
3.19 *A. calliptera mespaa* is consistently anterior to *tbx6*, in contrast to zebrafish. 65
3.20 mespaa marks somite specification . . . 67

4.1 Overview of cichlid somitogenesis . . . 76
4.2 PSM size differs only at the onset of segmentation . . . 78
4.3 PSM anatomy differs only at the onset of segmentation. . . . 82
4.4 Patterning of the PSM is identical between species . . . 84
4.5 Anatomy of the early embryo and unsegmented region at the onset of segmentation. . . . 86
4.6 Delayed completion of epiboly is unique to African cichlids. . . . 90

5.1 Visualization of *tbox* gene expression in the PSM . . . 97
5.2 Known genetic interactions in the tbox GRN . . . 97
5.3 AGET construction: pointcloud alignment . . . 101
5.4 Different strategies to construct AGETs result in near-identical outcomes at the cellular level. . . . 103

5.5 Different strategies to construct AGETs result in near-identical outcomes at the cellular level. . . . 104
5.6 Using 200 AGETs for reverse-engineering is the best balance between computation efficiency and network quality . . . 105
5.7 Using 200 AGETs for reverse engineering leads to the most robust patterns. 106
5.8 GRN Clustering . . . 108
5.9 Simulated gene expression from MAP network. . . . 109
5.10 For the MAP network, simulated gene expressions across the embryo closely recapitulate the known gene expression patterns. . . . 110
5.11 Other clustered parameter values recapitulate pattern well . . . 111
5.12 Heatmap of parameter values. . . . 113
5.13 GRN topology for different clusters. . . . 114
5.14 Over/Underexpression of FGF and Wnt is able to recapitulate patterns seen in the literature. . . . 116
5.15 Some parameter sets can upregulate *tbx6 in silico*. . . . 118

6.1 Types of fixed point in dynamical systems . . . 128
6.2 Gene expression pattern and network topology for best-fitting networks. 131
6.3 Cells moving anteriorly and posteriorly occupy different regions of signalling and gene expression space. . . . 132
6.4 Cells moving anteriorly are move likely to differentiate than those moving posterior. . . . 132
6.5 Initial gene expression used to control for gene expression heterogeneity. 133
6.6 Cells moving anteriorly more likely to begin to differentiate, after correcting for initial conditions. . . . 134
6.7 Cells experience different signalling values as they transit the PSM. . . . 136
6.8 Gene expression patterning is robust to signalling fluctuations. . . . 137
6.9 Gene expression heterogeneity is due to cell movement . . . 138
6.10 Gene expression pattern in perturbed simulations for all clusters. . . . 140
6.11 Heterogeneity in gene expression is mostly due to cell movement. . . . 141
6.12 Signalling profiles chosen for phase portrait analysis. . . . 142
6.13 Phase portrait for the MAP network. . . . 143
6.14 Condensed phase portrait visualization. . . . 145
6.15 Different networks produce different phase portraits. . . . 147
6.16 Percentage of cells in the PSM present in a given fixed point's basin of attraction. . . . 148

A.1 Somite numbers in *A. calliptera* and *R. chillingali* . . . 178
A.2 *A. calliptera* somite numbers at the onset of segmentation. . . . 181
A.3 *R. chillingali* somite numbers at the onset of segmentation . . . 182
A.4 PSM measurements used for PCA. . . . 183

List of Tables

2.1 Recipes for HCR 21
2.2 Microscope configuration 22
2.3 Software versions used for modelling 27
2.4 Software versions used for image analysis 29

3.1 Relationship between the onset of segmenation and epiboly in cichilds . . 34
3.2 Relationship between the onset of segmentation and epiboly in other teleosts 35
3.3 Comparison of staging in this chapter with Marconi et al., (2023) 40

5.1 Evidence for known genetic interactions in the tbox GRN 98

6.1 Cells moving anteriorly are more likely to differentiate, even after controlling for initial conditions. 135

A.1 Statistics on PSM metrics 179
A.2 Probability of interaction between species and somite stage 180

List of Abbreviations

AGET . . . . . Approximated gene expression trajectory
EMT . . . . . Epithelial to mesenchymal transition
GRN . . . . . Gene regulatory network
hpf . . . . . Hours post fertilization
HCR . . . . . Hybridization chain reaction
ICP . . . . . iterative closed point algorithm
MAP network . Maximum a posteriori (best fitting) network
MET . . . . . Mesenchymal to epithelial transition
MCMC . . . . . Marcov chain monte carlo
ODE . . . . . Ordinary differential equation
PSM . . . . . Presomitic mesoderm

Chapter 1: Introduction

During embryonic development, a single cell must become a whole, functioning, organism. How this occurs and is regulated is one of the great mysteries of biology. During embryonic development, two things must occur: cells must specialize, choosing adult cell fates, in a process called differentiation. Alongside cell differentiation, the embryo must change in shape, eventually adopting the adult form, a process called morphogenesis.

The processes of cell differentiation and morphogenesis have been extensively studied in isolation from one another, and are relatively well understood individually. Cell differentiation is controlled by gene regulatory networks (GRNs), that is, complex networks of genetic interactions that control the overall differentiation trajectories of individual cells. The ultimate fate of a particular cell is determined by these GRNs, and the external regulation (usually cell signalling) it is exposed to during development (Green & Sharpe, 2015; Sáez et al., 2022a; Fulton et al., 2022a). Tissue morphogenesis, on the other hand, involves a complex interplay of cell movements, tissue mechanics, and extra-embryonic constraints (Stooke-Vaughan & Campàs, 2018).

GRNs and morphogenesis influence each other. The transcriptional state of cells influences their physical properties and levels of motility: for example, whether a given cell is an epithelial (non-motile) or mesenchymal state is mediated by the expression of genes like *vimentin* and integrins (Baum et al., 2008). In turn, cell and tissue morphogenesis influence gene expression: changing the position of tissues relative to each other within the embryo during morphogenesis can influence the signalling environments a cell is exposed to, and therefore how

it will differentiate (Busby & Steventon, 2021). The position of a cell within a tissue will also determine its signalling environment and therefore the positional information that it is exposed to, which in turn affects its transcriptional state (Kicheva & Briscoe, 2015; Fulton et al., 2022a).

While GRNs and morphogenesis are well understood in isolation, and experimental work is beginning to unpack how these processes interact, we lack data-driven modelling frameworks to simulate and interrogate the interaction of these processes (Fulton et al., 2022a). To fully understand how morphogenesis and cell movements regulate gene expression and pattern formation within the embryo, a mathematical modelling framework is required which can incorporate cell movement in a model of pattern formation and cell differentiation (Fulton et al., 2022a). In this thesis, I take a first step in addressing this methodological gap, by establishing and validating a methodology to reverse-engineer gene regulatory networks achieving pattern formation alongside cell movement (Chapter 5). In Chapter 6 I go on to systematically analyse these networks, investigating how cell movements influence patterning of the zebrafish posterior body.

Most of what we know about the evolution of phenotypic diversity is focused on the role of GRN evolution. In the first half of my thesis I developed a methodology to investigate this more fully. In the second half, I turn to a related question: how phenotypic diversity can evolve with minimal genomic, and by extension GRN, evolution. During development, Lake Malawi cichlid species are able to produce a multitude of different phenotypes, including diversity in somite counts, craniofacial structure, and colouration of the body (Turner, 2007; Marconi et al., 2023; Bucklow et al., 2024). Despite this phenotypic diversity, they exhibit extremely limited genetic diversity, differing only 0.1-0.2% in nucleotide identity (Malinsky et al., 2018). Therefore, cichlids may achieve morphological diversity without necessarily evolving the GRNs involved in differentiation and patterning. Here, I investigate how two species of Lake Malawi cichlid achieve divergent somite counts, characterizing the process of somitogenesis in both

species and aiming to identify how any differences in this process may have evolved between species.

1.1 Evolvability of somite counts

The number of vertebrae in an adult organism is an adaptively important phenotypic trait, allowing different swimming and ecological strategies (McDowall, 2008). The number of vertebrae differs between vertebrate species: fish can have between 16 (sunfish) and 750 (eels), while snakes can form hundreds of vertebrae, and amphibians form around 10 (McDowall, 2008; Gomez et al., 2008). Vertebrae are formed from precursors called somites during embryonic development (Holley, 2007; Oates et al., 2012). While vertebral number closely correlates with somite count, not every somite forms exactly one vertebra, and the relationship between the two differs across vertebrates. In chick, the first five somites form the occipital bone and the following two somites form the first vertebra, resulting in five fewer vertebrae than somites (Ward et al., 2017; Piatkowska et al., 2021). Similarly, in zebrafish, the first two somites do not contribute to vertebrae, leading to zebrafish having two more somites than vertebrae (Morin-Kensicki et al., 2002). While in some species, somites resegment before forming vertebrae (the anterior portion of one somite contributes to the posterior of the next vertebrae), it has been found that this does not impact the total number of vertebrae formed (Morin-Kensicki et al., 2002; Ward et al., 2017; Piatkowska et al., 2021; Criswell & Gillis, 2020). As the number of somites formed during embryogenesis tightly correlates with the number of vertebrae present in the adult, development and evolvability of somite numbers within the embryo can be used as a proxy for understanding vertebral count evolution in adults.

In this thesis, I aim to identify the developmental basis of vertebral count evolution in Lake Malawi cichlid fishes. I will compare two species: *A. calliptera*, forming 30 vertebrae, and *R. chillingali*, forming 36 vertebrae. Below, I review the literature on somitogenesis and discuss how somite numbers could evolve.

1.1.1 Conceptualising the evolution of somite counts using the Clock and Wavefront model

The segmentation clock governs the rate of somite specification. To evolve differing somite counts, it is necessary to either modify the rate of the segmentation clock, or modify the total duration of this process.

During somitogenesis, somites are produced from a transient embryonic tissue known as the presomitic mesoderm (PSM). The Clock and Wavefront model proposes that somite specification is governed by two interacting entities, the segmentation clock, and the wavefront of somite determination (Cooke & Zeeman, 1976; Oates et al., 2012). The segmentation clock consists of a set of autonomously oscillating genes, the period of the which controls the rate of somite specification (Krol et al., 2011; Gomez et al., 2008; Oates et al., 2012). Oscillations in segmentation clock genes arrest as they encounter the wavefront, and cells adopt a somitic fate. The wavefront is thought to be associated with positional information provided by FGF signalling (Oates et al., 2012). At a molecular level, the segmentation clock is thought to be constituted of Delta-Notch signalling, and an autoregulatory core composed her/hes genes, such as *her7* and *hes11* in zebrafish, and *hairy1* in chick (Palmeirim et al., 1997; Jouve et al., 2002; Gomez et al., 2008; Krol et al., 2011; Schröter et al., 2012; Oates et al., 2012).

The total number of somites formed by the embryo is controlled by the rate of the segmentation clock and the total duration of somitogenesis (Oates et al., 2012). The Clock and Wavefront model predicts that increasing the rate of the segmentation clock would cause more, smaller, somites to be produced from the same volume of tissue (Oates et al., 2012). For example, the corn snake is thought to have increased the rate of somitogenesis relative to the cell cycle to evolve large numbers of somites (Gomez et al., 2008).

While the evolution of the rate of somitogenesis, and the components of the segmentation clock and wavefront, have been well-studied in vertebrates, these are not the only developmental traits impacting vertebral counts (Gomez et al., 2008; Krol et al., 2011; Oates et al., 2012; Steventon et al., 2016). Somitogenesis is

also tied to the total size of the PSM: the size of the nascent somite is controlled by the length of the PSM and as well as the wavefront of the somitogenesis clock, and, as the PSM is consumed as somitogenesis progresses, the total number of somites produced is also tied to the total length of the PSM (Gomez et al., 2008; Oates et al., 2012). Therefore, instead of evolving the segmentation clock, species could evolve increased numbers of somites by increasing the size of the PSM, and the duration of somitogenesis. This would result in more somites being produced, over a long timescale, from a larger total tissue volume. The cichlid fish *R. chillingali* may use this mechanism to evolve its increased number of somites relative to other species: James Hammond has shown that the duration of somitogenesis is extended in *R. chillingali* relative to other species (James Hammond, unpublished data).

This scenario raises the question, where could the increased tissue, potentially used to create an increased number of somites, arise from? There are a number of possibilities. Firstly, the size of the body axis could be larger at the onset of segmentation: this would mean there are more cells in the embryo that can be used to create somites. Secondly, smaller somites could be produced, resulting in more somites produced from the same tissue volume. Thirdly, the final size of the PSM could be reduced, meaning that more somites are produced from the same total tissue. Finally, there could be more addition of cells into the PSM, whether by cell division, or lateral ingression of cells into the PSM. While we do not know how these traits have evolved to alter somite counts, we do know which strategies are used during PSM elongation in different vertebrate species. For example, zebrafish do not exhibit overall growth of the body axis, so the total size of the body axis must be specified early in somitogenesis, while the corn snake, mouse, dogfish, and chick all exhibit volumetric growth of the body axis, and indeed, corn snake embryos (which make more somites) continue to grow the body axis for longer in somitogenesis than other species (Gomez et al., 2008; Steventon et al., 2016).

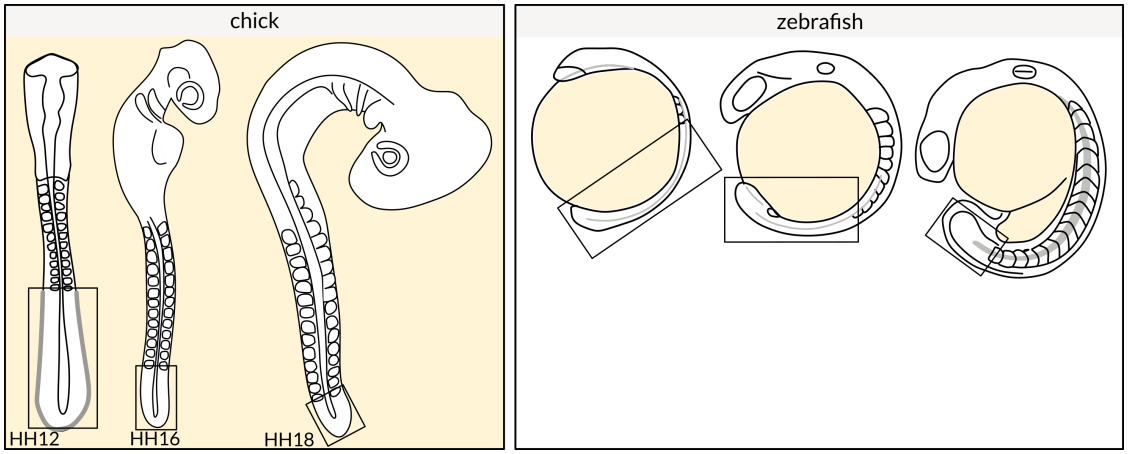


Figure 1.1: Vertebrate species develop with different relationships to the yolk and extra-embryonic structures. Chick embryos (left) develop on top of a very large yolk, with the two halves of the PSM laying beside each other. Zebrafish have a comparatively smaller yolk; prior to tail eversion (50% somitogenesis) the PSM lies flat on the yolk as in chick, but after this stage tail eversion occurs, and the tailbud folds in on itself, protruding from the yolk. Chick embryos are traced from Hamburger and Hamilton, 1951. Zebrafish embryos are traced from Kimmel et al., 1995. Black boxes indicate position of the PSM.

1.1.2 Mechanisms of elongation differ between species

To fully understand how somite counts evolve, we also need to understand the cellular mechanisms of axial elongation. These differ between species. In zebrafish, a change in the physical environment of the PSM, from a fluid-like state in the posterior to a solid-like state in the anterior, is thought to drive axial elongation (Mongera et al., 2018). There are three lines of evidence for this idea. Firstly, direct measurement of tissue pressure in the PSM proper and the mesodermal progenitor zone (MPZ) shows that pressure is higher in the PSM than in the MPZ (Mongera et al., 2018; Stooke-Vaughan et al., 2025). Second, cells in the MPZ exhibit more disordered cell motion and neighbour exchanges, consistent with a fluid-like cellular state (Lawton et al., 2013; Das et al., 2017; Mongera et al., 2018, 2023; Thomson et al., 2021). Finally, computational modelling reveals that a fluid to solid transition in the cellular environment, combined with surface tension (from extracellular membranes), is sufficient to produce tissue elongation in two dimensions (Mongera et al., 2018; Banavar et al., 2021). After tail eversion, zebrafish axial elongation is associated with compaction-extension: cell density

increases, cell volume decreases, and there is no directed cell rearrangements consistent with convergence-extension (Thomson et al., 2021). This points to the potential role of extrinsic forces in controlling elongation of the body axis at these stages (Thomson et al., 2021).

Prior to tail eversion, chick embryos achieve axial elongation using a different mechanism to zebrafish (Bénazéraf et al., 2010, 2017). Like zebrafish, cell motility is less directed in the posterior chick PSM (Bénazéraf et al., 2010). However, in chick, the posterior PSM also exhibits a decrease in cell density and increase in overall cell motility, suggesting that chick embryos elongate via a diffusion gradient, instead of compaction-extension (Bénazéraf et al., 2010, 2017). That is, cells in the posterior PSM are further away from their neighbours and are able to move more, resulting in overall extension of the body axis (Bénazéraf et al., 2010). Cell motility is controlled by FGF signalling, providing a link between gene expression patterning and morphogenesis (Bénazéraf et al., 2010). Note that this work studies chick axial elongation at early segmentation (11-15ss, or 22-29% somite production), when the PSM is still lying flat on the yolk and lateral ingression is possible. Chick could use a different mechanism of elongation at later stages. Indeed, the dynamics of growth change over time in chick: the length of the PSM increases between 3-22ss (0-40% somitogenesis), decreasing after this time (Gomez et al., 2008).

1.1.3 Patterning of the PSM

During axial elongation, cells migrate from the posterior tailbud, into the PSM, and then into the anterior PSM before forming somites. During this process, they undergo a number of transcriptional changes, ultimately differentiating into somites.

In zebrafish, cells adopting the mesodermal fate (meaning that they will form mesodermal tissue within the embryo, such as the somites, instead of neural tissue such as the neural tube) express *tbxta*, *tbx16*, *tbx6l* and *msgn*, amoung other markers, which together regulate cell motility and mesodermal differentiation in

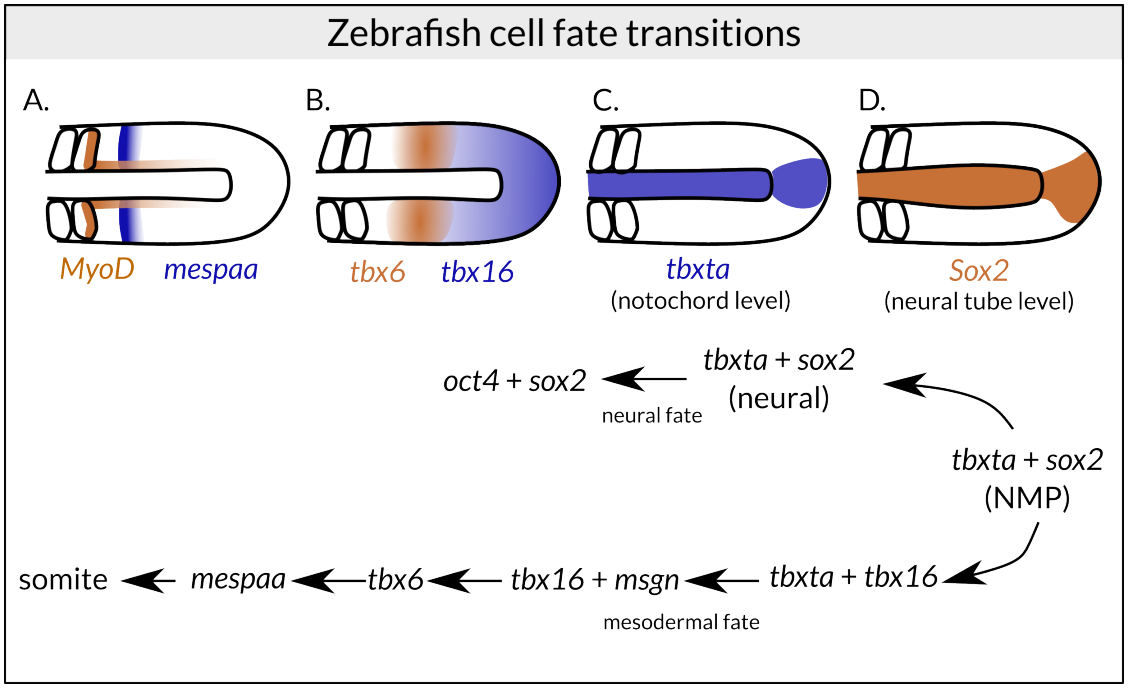


Figure 1.2: Transcriptional states in the zebrafish tailbud. Top panel depicts gene expression at mid-segmentation for important genes patterning the zebrafish PSM. Colours indicate expression of different genes. A. *MyoD* brown, *mespaa* blue. B. *tbx6* brown, *tbx16* blue. C. *tbxta* blue. Cartoon depicts the embryo as imaged through the notochord, not the neural tube. D. *Sox2* brown. Cartoon depicts the PSM as imaged through the neural tube. See Figure 1.1 for the position of the PSM relative to other tissues.

the PSM (Ruvinsky et al., 1998; Mueller et al., 2010; Manning & Kimelman, 2015; Bouldin et al., 2015; Morrow et al., 2017). In turn, expression of *tbx16* activates *tbx6*, and *tbx6* is upregulated as cells leave the posterior of the tailbud and enter the PSM proper (Bouldin et al., 2015; Fior et al., 2012). Thus, as cells move through the PSM, they express a sequence of genes: *tbxta*, followed by *tbx16*, *tbx6l*, and *msgn*, before expressing *tbx6* (Fulton, 2021) (Figure 1.2). Expression of *tbx6* is required for somite specification: embryos mutant for this gene do not form boundaries between somites, nor express somite boundary markers such as *MyoD* and *mesp*-class genes (van Eeden et al., 1996; Nikaido et al., 2002; Durbin et al., 2000; Cutty et al., 2012; Windner et al., 2015). Formation of the somite itself involves a mesenchymal to epithelial transition (MET), where an epithelial layer forms around the somite from the mesenchymal cells in the PSM (Naganathan & Oates, 2020).

While there is limited work done on patterning of the PSM in teleost fish other than zebrafish, the work that does exist suggests the function of these genes is conserved within teleost fish. The genes *tbx16* and *msgn* appear to retain their zebrafish function in medaka and the killifish (*Kryptolebias marmoratus*), being expressed in a posterior-to-anterior gradient (Saud et al., 2021). Loss of these genes also results in a circular mass of undifferentiated cells to form in the posterior PSM, suggesting that the function of *msgn* and *tbx16* is conserved in medaka and killifish (Saud et al., 2021). The medaka *mesp-aa* enhancer is also able to drive *mesb2* expression in mouse, suggesting a high degree of functional conservation of *mesp* regulation in vertebrates (Yasuhiko et al., 2008). Expression of genes involved in the segmentation clock is less well conserved. While *hes7* and *her11* are expressed in travelling waves in medaka as in zebrafish, *lunatic fringe* is not, and *fgf8* is not expressed in the PSM (Elmasri et al., 2004a). This implies, consistent with other work investigating conservation of the segmentation amongst vertebrates, that the dynamics of the segmentation clock are conserved in vertebrates but that the specific genes involved change extensively over evolutionary time (Elmasri et al., 2004a; Krol et al., 2011).

1.1.4 Co-operation between morphogenesis and patterning

There is extensive co-operation between PSM morphogenesis and patterning in zebrafish. For example, embryos mutant for *tbx16* and *mesogenin* have a large ball of cells in the tailbud, in place of a proper PSM (Kimmel et al., 1989; Griffin et al., 1998; Manning & Kimelman, 2015). This arises because *tbx16* and *msgn* are required for directed, productive cell migration in the tailbud, and loss of these genes causes undirected cell migration and therefore an accumulation of cells in the tailbud (Manning & Kimelman, 2015). In addition, FGF signalling is thought to be required for the cell motility gradient responsible for chick elongation, providing another link between morphogenesis and gene expression patterning (Bénazéraf et al., 2010).

1.2 Studying pattern formation alongside morphogenesis

As described above, during segmentation in vertebrates, extensive cell morphogenetic movements and cell rearrangements exist alongside robust and coherent pattern maintenance within the PSM. A major question in developmental biology is how gene expression patterns are established and maintained in tissues undergoing morphogenesis. Extensive work has revealed general principles underpinning pattern formation in tissues where morphogenesis and pattern formation occur over different timescales and can be separated. For example, in both the *Drosophila* blastoderm and the vertebrate neural tube, positional information as provided by signalling gradients is required for pattern formation (Jaeger, 2011; Marković et al., 2024). In the *Drosophila* blastoderm, opposing gradients of *Bicoid* and *Caudal* mRNA and protein position the initial expression of gap genes (Jaeger, 2011). However, gap gene autoregulation is also required for correct pattern formation, meaning that GRN dynamics, alongside positional information provided from signalling, are both required for pattern formation in the *Drosophila* blastoderm (Jaeger, 2011; Manu et al., 2009b).

Studying pattern formation in systems where patterning and morphogenesis occur over different timescales, such as the vertebrate neural tube and *Drosophila* blastoderm patterning, simplifies the study of GRNs, as cell movements can be ignored when inferring and analysing models (Jaeger, 2011; Fulton et al., 2022a; Marković et al., 2024). However, in the majority of developmental systems, morphogenesis and patterning co-occur and, indeed, can influence each other (Fulton et al., 2022a). For example, the zebrafish body axis forms during extensive convergence extension cell movements (Shah et al., 2019). Zebrafish tailbud morphogenesis also involves a complex interplay of cell movements, gene regulation, and extra-embryonic and physical constraints (Kanki & Ho, 1997; Lawton et al., 2013; Fulton et al., 2022a). Patterning of the PSM is achieved by genes such as *tbxta*, *tbx16*, *tbx6*, which are regulated by FGF and Wnt signalling (Martin & Kimelman, 2008; Bouldin et al., 2015; Goto et al., 2017). Cell movement,

in turn, is regulated by patterning of the PSM: *tbx16* and *msgn* are required for directed cell motility in the PSM (Manning & Kimelman, 2015; Das et al., 2017). Additionally, mechanical constraints such as the fluid-to-solid transition in the zebrafish PSM influence body axis elongation and development (Mongera et al., 2018). Altogether, then, gene expression patterning, GRNs, and morphogenesis co-operate to achieve development of the zebrafish posterior body.

Therefore, it is essential to develop modelling methodologies that can incorporate cell movement into models of pattern formation (Fulton et al., 2022a). There are two general approaches that could achieve this. Firstly, GRNs describing known genetic processes could be reverse-engineered alongside cell movements (Fulton et al., 2022a). This approach has the advantage that the GRNs describing morphogenesis can be reverse-engineered directly using gene expression and cell movement datasets, making the resultant models more biologically relevant. Alternatively, models could be constructed from the ground up and hand-tuned to resemble morphogenetic processes of interest. For example, the modelling framework EmbryoMaker can simulate morphogenesis alongside genetic interactions, modelling an embryo as a group of cells plus a GRN; the cells can have differing shapes and mechanical properties (Marin-Riera et al., 2016). This has allowed the study of how complex structures can evolve (Marin-Riera et al., 2016; Hagolani et al., 2021). The vertebrate segmentation clock has also been studied in a morphogenic context, allowing the impact of morphogenesis on somitogenesis to be studied (Uriu et al., 2017, 2021; Hammond et al., 2025).

In this thesis, I take the former approach, developing and validating Live Modelling (as outlined in Fulton et al., 2022a) as a method to study how morphogenesis impacts gene expression patterning. In Chapter 5, I develop and validate a method to reverse-engineer GRNs describing the tbox GRN during PSM morphogenesis in zebrafish. This therefore establishes Live Modelling as a viable methodology to investigate the relationship between patterning and morphogenesis. In Chapter 6, I then analyse the parameter sets generated in this chapter to explain different functions of the zebrafish tbox GRN, such as the

ability of cells in the GRN to autonomously differentiate *in vitro* and without any change in the signalling they experience (Fulton, 2021; Rohde et al., 2024).

1.3 Lake Malawi cichlids: a system to study the evolution of phenotypic diversity

As well as investigating the interaction of morphogenesis and patterning, in this thesis, I also tackle a related question: how can phenotypic diversity evolve, particularly in the absence of genetic diversity? There are two main strategies to study how phenotypic diversity is produced during development. The first is to characterize and compare the development of organisms that are very divergent in evolutionary history, phenotypic traits, and ecological behaviour. For example, Steventon et al., 2016 employ this method to compare tissue contributions to vertebrate axial elongation and propose that maternal energy/yolk contributions govern whether embryos can increase the amount of tissue in the body axis over time. Such analyses have also been employed to great effect to study the evolution of segmentation in arthropods, revealing that apparently different mechanisms of segmentation can be achieved using a conserved GRN (Peel et al., 2005; Clark, 2017; Clark & Peel, 2018; Clark et al., 2019; Taylor & Dearden, 2022). However, this method has the drawback that it is difficult to pinpoint the precise causes of phenotypic evolution in organisms that are ecologically, genetically, and developmentally diverse.

The second approach is to identify genetically closely related species with divergent phenotypes, and identify the developmental and genetic basis for this diversity. This approach has been used to study genital diversity in *Drosophila* species (Urum & Preger-Ben Noon, 2025) and to identify the developmental basis of Galapagos finch beak evolution (al-Mosleh et al., 2021). This approach makes it simpler to identify subtle, potentially causative, differences between species that contribute to species diversity.

To study the evolvability of somite counts in vertebrates, we need to identify such a system: a species group where somite counts vary, but species remain

similar genetically and developmentally, to allow detailed and specific comparison of the developmental basis of somite count evolution. Lake Malawi cichlid fishes represent an ideal system with which to approach this question, varying in somite counts while exhibiting very limited genetic diversity (Santos et al., 2023; Malinsky et al., 2018; Bucklow et al., 2024). In this thesis, I am studying the evolution of somite counts by comparing development and axial elongation of two Lake Malawi cichlid species with varying somite counts.

1.3.1 Adaptive radiation in Lake Malawi | Lake Nyasa | Lago Niassa

Lake Malawi cichlids are an amazing example of an adaptive radiation (Turner, 2007; Malinsky et al., 2018; Santos et al., 2023). There are at least 800 species of cichlid fishes in Lake Malawi (also known as Lake Nyasa and Lago Niassa), which are thought to have evolved in the last 800,000 years from a common ancestor resembling the extant generalist *A. calliptera* (Turner, 2007; Malinsky et al., 2018; Santos et al., 2023). Cichlids in Lake Malawi occupy the majority of ecological niches in the lake, and differ in phenotype: varying in head shape, tooth and jaw structure, colour and patterning, body shape, behaviour, as well as vertebral count (Santos et al., 2023). Despite this astounding phenotypic diversity, they are genetically very similar, differing on average by just 0.1-0.2% at the nucleotide level, and in some cases there is more nucleotide diversity found within a single individual than between species (Malinsky et al., 2018). While assortative mating is thought to enable speciation within the lake, individuals from different species can produce hybrid offspring in the lab and presumably also in the wild (Turner, 2007).

Cichlid species appear to be predisposed to undergo dramatic speciation events. There are two other adaptive radiations of cichlid species in the African Great Lakes (Lakes Victoria and Tanganyika), as well as in the Midas species flock in South America (Turner, 2007). Exactly what predisposes cichlids to such phenotypic diversity is unknown (Turner, 2007).

There are genomic and environmental hypotheses potentially explaining why Lake Malawi cichlids are so phenotypically diverse. There is genomic evidence for an early hybridization event between two *Astatotilapia* species, which could have facilitated diversification (Svardal et al., 2020). The genomic phylogeny within Lake Malawi cichlids suggests gene flow and therefore hybridization between species (Malinsky et al., 2018). There is no evidence that an increased mutation rate contributes to diversification in Malawi cichlids: the mutation rate of cichlids is low in comparison to other species (Malinsky et al., 2018).

In addition to genomic explanations, developmental innovations can lead to speciation. For example, in the classic example of an adaptive radiation, the shape of Galapagos finch beaks can be explained by a change in development: changing *bmp4* expression and growth dynamics during beak formation changes the shape of the adult beak (Abzhanov et al., 2004; al-Mosleh et al., 2021). Additionally, within Lake Malawi cichlids, the anatomy of the adult face is shaped by heterochrony and variation during development (Powder et al., 2015).

In this thesis, I examine the ontogeny of somite counts within Lake Malawi cichlids. Lake Malawi cichlids have between 28 and 42 vertebrae, and species where development has been examined likewise vary in somite number (Marconi et al., 2023; Bucklow et al., 2024). In Chapter 3, I describe the process of axial elongation in detail in *A. calliptera*, a generalist Lake Malawi cichlid which is thought to closely resemble the common ancestor of all cichlids within Lake Malawi (Turner, 2007; Malinsky et al., 2018; Santos et al., 2023). In Chapter 4, I perform a quantitative morphometric comparison of *A. calliptera* and *R. chillingali* development. *R. chilllingali* was chosen to study because, with 38 somites, it represents one of the most phenotypically extreme cichlids within Lake Malawi (Bucklow et al., 2024).

1.3.2 Development of *A. calliptera* and other cichlid species

Very little is known about Lake Malawi cichlid development, particularly in comparison to other teleost species like zebrafish and medaka. Marconi et al., 2023 characterize *A. calliptera*, *R. chillingali*, and *Tropheops* sp. "mauve" development from egg laying until the larval stage. This revealed that, like other cichlid species, the studied Malawi cichlids develop on top of a large, egg-shaped yolk (Marconi et al., 2023). Unusually, and as far as I can tell uniquely to African cichlids, Lake Malawi cichilds do not complete epiboly until the end of segmentation (Marconi et al., 2023). This is quite different to other teleosts fish species, which, like zebrafish, complete epiboly very early in segmentation, before more than ~ 5 somites have formed (Kimmel et al., 1995)[1]. Despite *R. chillingali* forming more somites than *A. calliptera*, formation of key anatomical landmarks such as the eye, otic vesicle, and other neural structures is tied to the number of somites produced by the embryo, not the percentage of somitogenesis completed (Marconi et al., 2023). This implies that the developmental programme is tied to overall somite count, not the percentage of somitogenesis completed, in these species (Marconi et al., 2023). One important difference between cichlids and zebrafish is that, during somitogenesis, *A. calliptera* develop as flat embryos on top of a large yolk until tail eversion occurs at 28ss (Marconi et al., 2023). In contrast, in zebrafish, 50% of somite formation occurs after tail eversion (Kimmel et al., 1995; Steventon et al., 2016). More recent work has analysed the evolution of the neural crest in more depth, showing that functional evolution of *sox10* paralogs is likely to have altered neural crest development between *A. calliptera* and *R. chillingali* (Marconi et al., 2024).

There is much that is not known about cichlid development during axial elongation. How many somites are specified at the onset of segmentation? Does the duration of somitogenesis differ between species? What is the developmental basis of somite count evolution in *R. chillingali*? What is the mechanistic basis of cichlid axial elongation? I will tackle these questions in Chapters 3 and 4.

[1]See Table 3.2 in Chapter 3 for a full description and justification of this statement

1.3.3 Genetic and embryological tools in *A. calliptera* and *R. chillingali*

A number of genetic and embryological tools exist for *A. calliptera* and *R. chillingali*. Embryos can be obtained regularly from free-breeding 'stock tanks' of fish. However, as neither *A. calliptera* nor *R. chillingali* breed readily on demand, obtaining very early (1-2 cell) embryos is difficult and in this thesis, embryos are predominantly staged via morphology, not hours post fertilization (hpf). A published genome for *A. calliptera* is available, allowing us to generate probes for *in situ* hybridization and hybridization chain reaction (HCR) (Malinsky et al., 2015; Choi et al., 2018). Though no *R. chillingali* genome is publicly available, due to the high level of genomic sequence conservation within Lake Malawi cichlids, probes designed for *A. calliptera* are also cross-reactive with *R. chillingali* gene sequences (Malinsky et al., 2015, 2018, our own observations). Recently, CRISPR-based genome editing has become possible in *A. calliptera*, enabling functional genetic studies in cichlids, and Tol2 genome editing was already possible (Juntti et al., 2013; Clark et al., 2022, 2023)

While Lake Malawi cichlids are the ideal biological system to study the evolution of somite count, they do present considerable practical challenges. While synchronized clutches of embryos are readily available, the difficulty of obtaining 1-2 cell embryos makes precise staging of embryos challenging. Live imaging of embryos is also difficult, owing to the difficulty of obtaining sufficiently early cichlid embryos, autofluorescence from the large yolk, and the long developmental timescales of these fish. In addition, the cichlid chorion appears to be impermeable to pharmacological treatments, making developmental perturbations difficult (James Hammond, unpublished data). For these reasons, in this thesis, I focus on performing a detailed, quantitative description of *A. calliptera* and *R. chillingali* development, using fixed specimens and high resolution confocal imaging.

1.4 Aims of this thesis

This thesis contains four data chapters. Each contains a dedicated Introduction, Results, and Discussion, so that they can be read independently. There is overlap in the Methods, so this is presented as a separate chapter (Chapter 2).

In Chapters 3 and 4, I investigate axial elongation and patterning in two species of Lake Malawi cichlid, *A. calliptera* and *R. chillingali*; which differ in somite count. Chapter 5 presents a detailed characterization of *A. calliptera* development and patterning during axial elongation.

In Chapter 4, I perform a quantitative morphometric comparison of axial elongation in *A. calliptera* and *R. chillingali*. I find that species differ only at the early stages of segmentation, suggesting that cichlids may have evolved divergent somite counts by modifying the initial conditions of axial elongation; rather than evolving axial elongation itself.

In Chapter 5, I present the work I have done to establish and validate the Live Modelling methodology: a technique to reverse-engineer GRNs in a tissue undergoing morphogenesis. I thoroughly validate this methodology, demonstrating that it is able to produce parameter sets that reproduce patterning at the tissue level, known genetic interactions, and predict biological functions of the GRN. The work in this chapter has also been published in Spiess et al., 2024; I am co-first author on this paper. For this chapter I re-wrote my contributions in my own words.

In Chapter 6, I analyse parameter sets obtained in Chapter 5, to investigate the interaction between morphogenesis and patterning. I find, firstly, that there are multiple distinct dynamical mechanisms that can achieve zebrafish tbox patterning. This suggests a further level of degeneracy in the genotype to phenotype map. I also find that signalling is not required for patterning of the PSM; instead, cell differentiation/upregulation of *tbx6* appears to be cell-autonomous in my models. However, the rate of differentiation is altered by signalling. This implies that, while signalling modulates the rate of cell differentiation in the PSM, over

the timescales studied, the fluctuations in patterning that cells experience as they transit the PSM are not required for correct tbox patterning.

Chapter 2: Materials and Methods

2.1 Wet-lab methods

2.1.1 Animal husbandry

A. calliptera and *R. chillingali* were reared at 26°C, under a 12:12 hour day-night cycle. *A. calliptera* were fed JBL Novo Malawi flake (for algae-eating cichlid), and *R. chillingali* were reared on JBL Novo Malawi flake for carnivores. Feed was supplemented with pellets and frozen fish for enrichment and to encourage breeding.

2.1.2 Extraction and fixation of cichlid embryos

Embryos were extracted from stock tanks at least weekly. Embryos were removed from mouthbrooding females by first catching the female, and then flushing the mouth with water using a pasteur pipette until all embryos were released.

If embryos were to be reared to a specific stage, they were then incubated in a solution of tank water with 1mg/L of methylene blue (Sigma-Aldrich)to prevent mould or bacterial contamination. Embryos were incubated on a slow rocker at 26°C. Water was changed and dead embryos removed daily.

Fixation was performed overnight at 4°C, in 4% formaldehyde (Sigma).

2.1.3 Hybridization chain reaction

Hybridization chain reaction (HCR) was performed as per (Choi et al., 2016, 2018; Fulton et al., 2022b). In short, embryos were fixed overnight at 4°C in

4% formaldehyde (though paraformaldehyde is often used for teleost fixation, we have noticed no degradation in stain quality or embryo morphology using formaldehyde). Embryos were then dehydrated in a 25%, 50%, 75% methanol:PBS series, leaving tissue to equilibrate for five minutes for each methanol concentration. Embryos were then stored at $-20°$C in 100% methanol. Note that all washes performed are for five minutes unless indicated otherwise.

To perform the HCR, embryos were first rehydrated in a 75%, 50%, 25% methanol:PBS series as before. Embryos were then dissected from the yolk in PBS, and placed in a glass staining dish. Permeabilization with Proteinase K was typically not required. If permeabilization was required, embryos were then permeabilized for one hour in 10μ g/mL proteinase K (Sigma), before being washed twice in PTw and re-fixed in 4% formaldehyde for 30 minutes.

Embryos were then pre-hybridized for 30 minutes at $37°$C in 30% hybridization buffer (Table 2.1), before 1μL of probe was added per 100μL of hybridization buffer. Note that to ensure strong and consistent staining I used 100μL of hybridization buffer per half-dozen embryos, and did not stain more than 15 embryos per staining dish. The embryos were then left overnight to hybridize at $37°$C.

The following day, I washed out the probe using 4x15 minute washes in probe wash buffer (Table 2.1) at $37°$C. At room temperature, I performed 3x5 minute washes in SSCT (Table 2.1), before pre-amplifying in amplification buffer (Table 2.1). I heat-shocked the hairpins at $95°$C for 90 seconds in separate PCR tubes, before adding these to the embryos. I then incubated embryos, overnight, in the dark.

The following morning I washed out the hairpins in SSCT and added 1μL/mL DAPI at least overnight, before mounting in 100% glycerol and imaging.

Table 2.1: Recipes for HCR

Reagent	Recipe or source
1x PBS	Produced from tablets (Sigma-Aldrich, P4417)
PTw	100mL 1X PBS, 100μL Tween 20
Probe hybridization buffer	Purchased from Molecular Instruments
Probe wash buffer	Purchased from Molecular Instruments
5X SSCT	20X SSC (Thermo-Fisher, J60561.AP) , 0.1% Tween 20, distilled water
Amplification Buffer	Purchased from Molecular Instruments
DAPI	4',6-Diamidino-2-Phenylindole, dihydrochloride (Sigma: MBD 0015-IML)

2.1.4 Phalloidin staining

Phalloidin staining must be performed on embryos that have never been dehyrated in methanol. Embryos were extracted and fixed as above, before being washed three times in PTw and dissected. Embryos were then permeabilized for a total of two days in 1% PTx, washing four times over this time period. Embryos were then washed three times in PTw before being stained overnight in half a unit (10μL) Phalloidin-CF568 (Biotin, BT00044-T) in 100μL PTw and 0.5μL DAPI (Biotin, BT40011). Finally, the stain was washed out three times in PTw, before embryos were mounted in 90% glycerol.

2.1.5 Imaging

Imaging was performed on either an Olympus FV3000, Leica LSM980, or Leica LSM880. Microscope configuration is summarized in Table 2.2

Table 2.2: Microscope configuration

Microscope	Low magnification objective	Medium magnification	High magnfication objective	Lasers (nm)
Olympus FV3000	Olympus UPLSAPO 4X (0.16NA dry)	N/A	UPlanSApo 30x silicon oil (1.05NA)	405, 488, 514, 561, 594, 631
Leica LSM980	Zeiss 10x (0.45 NA Dry)	N/A	40x (1.30 NA Oil)	405, 488, 514, 561, 639
Leica LSM880	Zeiss 10x (0.45 NA Dry)	Plan-Apochromat 20x/0.8 M27	40x (1.30 NA Oil)	405, 488, 514, 561, 639
Leica Stellaris 8	HC PL FLUOTAR 10x (0.30 NA Dry)	HC PL APO CS2 20x (0.75 NA water/multi-immersion)		White light laser

2.2 Modelling methods

All code relevant to Chapter 5 is available at `doi.org/10.5281/zenodo.12607613`. Code for other chapters is at `doi.org/10.5281/zenodo.17092281`.

2.2.1 AGET construction

To produce AGETs, imaging data was prepared by Kay Speiss, as described in Spiess et al., 2024. This produced pointclouds describing the location of nuclei within the tailbud within space. Pointclouds from different embryos were aligned using the iterative point cloud (ICP) algorithm implemented in the Open3d python library (Zhou et al., 2018; Rusinkiewicz & Levoy, 2001). To obtain gene expression trajectories, the pointclouds describing cell tracking data were aligned with pointclouds associated with gene expression levels at each timepoint. This provided multiple gene expression measurements for each cell, at each timepoint, which were then averaged.

2.2.2 Model formulation

To model the tbox GRN, a dynamical systems formulation was used. Here, the tbox GRN is modelled as a system of coupled ordinary differential equations, where expression of each gene relies on cross-regulation from other genes, signalling inputs, and the individual production and decay rates of each gene. This model formulation has been extensively used to model pattern formation in other developing systems (Jaeger et al., 2004; Crombach et al., 2012).

The state variables of the system represent expression of the tbox genes *tbxta*, *tbx16*, and *tbx6* respectively. The rate of change of each gene is represented by the following system of coupled ordinary differential equations:

$$\frac{dg_a(t)}{dt} = R_a\phi(u_a) - \lambda_a g_a(t) \tag{2.1}$$

Here, R_a represents the production rate of the gene, and λ_a represents its decay rate. The concentration of gene a at time t is given by the term $g_a(t)$. The index (a) indicates the gene being modelled. To introduce non-linearities into the model, the sigmoid function ϕ is used:

$$\phi(u_a) = \frac{1}{2}\left(\frac{u_a}{\sqrt{(u_a)^2+1}} + 1\right) \tag{2.2}$$

where

$$u_a = \sum_{b\in G} W^{ba} g_b(t) + \sum_{s\in S} E^{sa} g_s(t) + h_a \tag{2.3}$$

The tbox genes are represented by the state variables, G={tbxta, tbx16, tbx6} and signalling factors by $S = \{\text{Wnt}, \text{FGF}\}$. For this modelling framework, the level of signalling g_s is provided by the AGETs themselves, and is not simulated. Because the levels of Wnt and FGF change over time in each cell, the term $\sum_{s\in S} E^{sa} g_s(t)$ is time-dependent, and the overall model is non-autonomous (Verd et al., 2014).

The regulatory input to each gene is controlled by current levels of signalling and gene expression, weighted by the parameter sets reverse-engineered using

MCMC. These parameter sets are represented by the interconnectivity matrices W and E, for regulatory interactions amoungst the tbox genes, and from signalling, respectively. Matrix elements W^{ba} and E^{sa} represent the effect of regulator b or s on target gene a. These parameters can be positive, representing activation of b by a, negative, representing an inhibition, or close to zero, representing no interaction. The parameter h_a represents basal activity of gene a, which could be due to regulation from genes not included in the model formulation, or basal activity of the gene.

This means that every gene has eight parameters associated with it: basal activity h, degradation rate λ, production rate R, regulation by other genes (W^{tbxta}, W^{tbx16}, W^{tbx6}), and regulation by signalling (E^{FGF}, E^{Wnt}).

To simulate gene expression for each AGET, the simulation is initialized with initial gene expression from the first timepoint for each AGET. At each timepoint, the signalling profiles are updated. For simulations in Chapters 5, 6, simulations are performed either using the `lsoda` of function `numbalsoda`, using rtol and atol values of 10^{-8}, or using the `odeint` function and lsoda solver in `scipy` (Virtanen et al., 2020).

2.2.3 Model fitting: MCMC

Parameter sets were reverse-engineered using the emcee package (Goodman & Weare, 2010; Foreman-Mackey et al., 2013). As an ensemble sampler, this MCMC sampler uses an ensemble of walkers, instead of just one, and scales the proposal distribution to the current walker distribution (Goodman & Weare, 2010; Foreman-Mackey et al., 2013). This therefore makes the sampler better at sampling posterior distributions where parameter sets can differ by an order of magnitude (Goodman & Weare, 2010; Foreman-Mackey et al., 2013).

I tried using different prior distributions when reverse-engineering, which impacts results by constraining the total search space. I found that Gaussian priors did not sufficiently constrain the search space, with walkers becoming stuck in unrealistic regions of parameter space (not shown). Accordingly, I used flat priors,

capped to within ± 200. Large parameters (more extreme than ± 100) tended to have poor likelihood scores and were removed from subsequent analysis.

I optimized the number of walkers used for sampling. Initial tests revealed that changing the number of walkers did not change the parameter distributions obtained, so I used the number of walkers recommended for most computational efficiency: twice the number of compute cores used. Therefore, here, `emcee` was run using 96 walkers (twice the number of compute cores), using the University of Oxford Advanced Research Computing (ARC) facility[1]. At least five independent runs were performed using each unique set of AGETs, and MCMC was run for a minimum of 10,000 iterations. If, after this time period, MCMC had not reached convergence, MCMC was run for a further 10,000 iterations. Convergence was defined using the Kolmogorov–Smirnov test (in `scipy`, Virtanen et al., 2020): convergence was deemed to have been reached if the the final 5000 parameter sets were statistically indistinguishable from the -10000th to -8000th parameter sets, for at least 20 of the parameters, using a significance threshold of 0.001.

2.2.4 Calculating a likelihood score

The likelihood score is defined as the mean squared distance between the target (AGET) and simulated gene expression, across all timepoints, for all genes fit to. Data were normalized using a custom σ value for each gene. The likelihood function is expressed mathematically:

$$L = -0.5 \cdot \sum_{a=0}^{a_{\max}} \sum_{c=0}^{c_{\max}} \sum_{t=0}^{t_{\max}} \frac{\left(T_{at}^{c} - S_{at}^{c}\right)^2}{\sigma_a} \tag{2.4}$$

In this equation, a refers to the gene, c refers to the cell (or AGET) used for reverse-engineering. In this study c_{max} is between 10 and 200. t refers to the timepoint of the simulation; for our dataset t_{max}=60. T represents the target (or AGET) gene expression, and S refers to simulated gene expression. σ is a scaling factor, unique for each gene: here, $\sigma_{tbxta} = 0.2$, $\sigma_{tbx16} = 0.2$, $\sigma_{tbx6} = 0.1$.

[1] http://dx.doi.org/10.5281/zenodo.22558

2.2.5 Filtering and clustering parameter sets for analysis

Parameter sets for further analysis were filtered as follows. I first took only parameter set where 200 AGETs were used for reverse-engineering. I then excluded parameter sets which met any of the following conditions: the likelihood score, as computed to all 1903 AGETs, is less than -15000; or parameter sets are unrealistically large (more extreme than ± 100). I then clustered the resulting parameter sets, using k-means clustering (`scikit-learn`, Pedregosa et al., 2011). The number of clusters to produce was determined using an elbow plot, and any clusters representing less than 5% of all parameters were excluded.

2.2.6 Identification and classification of fixed points in the tbox GRN

To identify and classify fixed points in the tbox GRN, I first identified relevant seeds for the equation, identified the root of these equations numerically, and then classified these roots.

I used two strategies to identify seeds for the root finding algorithm. First, I took the first timepoint of the simulation, and used the gene expression values for each cell as seeds. Second, I computed the rate of change for all gene expression values between 0 and 5, with a step size of 0.01 (or, $\{0.01k \mid k \in [0, 5]\}$ / `numpy.arange(0, 5, 0.01)`). I kept gene expression values with a rate of change of less than 0.5 for all variables as seeds for root finding.

To identify fixed points, or roots, of the equation, I took both of the collections of seeds described above, and used the root-finding algorithm in `scipy` (Virtanen et al., 2020) to attempt to find the root. If this failed, due to the seed being too far from the fixed point, I used the seed as initial conditions to solve the ODE for 20 times the total simulation duration, and then used the solution as the seed to repeat the root-finding. I then verified that the root found was indeed a fixed point, e.g. the absolute rate of change at these gene expression values $< 10^5$. Finally, I removed duplicate fixed points.

Table 2.3: Software versions used for modelling. Full anaconda enviornment is provided on Github.

Software	version	citation
python	3.7.10	
numpy	1.20.3	Harris et al., 2020
scipy	1.6.3	Virtanen et al., 2020
sympy	1.9	Meurer et al., 2017

Fixed points were classified using the sympy library, by extracting the real and imaginary components of the eigenvalues and classifying accordingly.

Package versions are given in Table 2.3.

2.3 Image analysis methods

2.3.1 Morphometrics

Morpometric measurements were performed by manually annotating images within OMERO (Burel et al., 2015). Anatomical regions were defined as follows. The PSM length is defined as the distance between the posterior-most tip of the PSM, at the level of the notochord, and the posterior boundary of the nascent somite; measured through the notochord. The notochord length is the distance between the posterior boundary of the nascent somite, and the posterior tip of the notochord, measured through the notochord. The MPZ length is the difference between these two values. Somite length is defined as the distance between the posterior and anterior boundaries of the nascent somite, measured at the middle of the somite. PSM width was measured by placing two points at the intersection of the posterior boundary of the nascent somites, and the lateral aspect of the PSM, and by computing the Euclidean distance between these points in python. PSM depth is similarly measured, by placing points at the dorsal and ventral most points of the PSM, at the anterior-posterior position of the lateral tip of the notochord. Notochord width is measured at the middle of the notochord. Metrics are diagrammed in Figure 4.3.

The embryo length was measured as a poly-line across the AP axis of the embryo, from the tip of the nose to the tip of the tailbud, following the embryonic spinal cord.

Principal component analysis was performed in R using the 'stats' package. Data were scaled prior to PCA.

2.3.2 Manual tissue segmentation

Tissues were manually segmented using the Segmentation Editor tool in ImageJ (Schindelin et al., 2012). The whole tailbud was first outlined every 5-10 slices, then the labels were interpolated and the label added to the image. The notocord, neural tube, and somites were then segmented using the same method, and given unqiue labels for identification.

To measure the volume of different tissues, the label object was resized so that each voxel measured $1\mu m^3$. The number of voxels in each label were then counted. For somites, the number of somites segmented (one or two) was detected automatically by extracting the somite label, labelling each somite, and then counting the number of labels. The mean somite volume was then computed by dividing the total somite volume by the number of somites.

2.3.3 Measuring patterning across the AP axis

Gene expression across the anterior-posterior axis were measured in Fiji as follows. First, an average intensity projection of relevant focal planes was performed in ImageJ. The number of focal planes varied between specimens, depending on the size of the optical section. Average intensity projections were always performed on focal planes adjacent to the notochord. A polyline was then drawn across the PSM, from the posterior boundary to the nascent somite to the posterior of the notochord; following the middle of the PSM. Expression was then measured, averaging across 20-100 pixels adjacent to the line. The number of pixels was chosen such that the polyline width was as close as possible to the width of the PSM itself, without including any tissue outside the PSM.

Table 2.4: Software versions used for image analysis. Full anaconda enviornment is provided on Github.

Software	version	citation
python	3.9.20	
numpy	1.26.4	Harris et al., 2020
scipy	1.13.1	Virtanen et al., 2020
pyclesperanto	0.16.0	Rigaud et al., 2024
napari	0.5.4	Sofroniew et al., 2025
skimage	0.24.0	Walt et al., 2014
R	4.4.2	R Core Team, 2024
factoextra	1.0.9	Kassambara and Mundt, 2020
ggplot2	3.5.1	Wickham, 2016
ggpubr	0.6.0	Kassambara, 2023
dplyr	1.1.4	Wickham et al., 2023
tidyr	1.3.1	Wickham et al., 2024
data.table	1.16.2	Barrett et al., 2024
OMERO		Burel et al., 2015
Fiji	2.14.0	Schindelin et al., 2012

2.3.4 Cell counting

Cell counting was performed using a method derived from a standard spot counting protocol by Robert Haase, using the `pyclesperanto` library (Haase et al., 2020). The DAPI (nuclear) channel was first extracted from the image, and rescaled to have voxels of size $0.5\mu\text{m}^3$, for consistency. A gaussian blur was then performed on the nuclear channel, with a radius of 0.5μm (one voxel) in the X and Y dimensions, and zero voxels in the Z dimension. The local maxima was detected, using a 'box' (cubed) connectivity, and radius of 1.5μm (three voxels) in all dimensions. To merge nearby spots, a binary dilation was performed with radius 1.5μm (three voxels) in the X and Y dimensions and 0.5μm (one voxel) in the Z dimensions, before spots were labelled, the relevant tissue extracted, and the number of unique labels (e.g. number of cells) in the tissue was counted.

The quality of spot detection was verified blind after segmentation, and embryos where spot detection was not accurate were discarded.

Chapter 3: Characterization of axial elongation during somitogenesis in the Lake Malawi cichlid *A. calliptera*

Contents

3.1 Introduction **31**
3.1.1 Cichlids are an important system for understanding evolution and development 31
3.1.2 Prior knowledge of *A. calliptera* development 31
3.1.3 Evo-devo of teleost axial elongation 32
3.1.4 Mechanisms of vertebrate axial elongation 33
3.1.5 Patterning of the PSM 38
3.2 Results **40**
3.2.1 Formation of the body axis: 0ss 41
3.2.2 Structural integrity of the PSM arises at 2-4ss 47
3.2.3 The early segmentation stage: 5-15ss 49
3.2.4 Separation of the PSM from the lateral plate mesoderm: 14-16ss 52
3.2.5 Distinct cell types in the embryo: 20ss 55
3.2.6 Late segmentation: 25ss onwards 56
3.2.7 Cell density in the PSM 57
3.2.8 Patterning of the *A. calliptera* midline 59
3.2.9 Oct4 patterning of the *A. calliptera* neural tube 61
3.2.10 Gene expression patterning of the PSM 63
3.3 Discussion **67**
3.3.1 The morphology of the *A. calliptera* PSM changes over time . . . 68
3.3.2 *A. calliptera* differs to zebrafish in patterning 69
3.3.3 *A. calliptera* elongate via compaction-extension 70
3.4 Conclusions **71**

3.1 Introduction

Lake Malawi cichlids have undergone a recent adaptive radiation, diversifying in vertebral counts, craniofacial morphology, and body axis ratio, amoungst other features (Turner, 2007; Santos et al., 2023). However, the developmental basis of this diversification remains poorly understood, particularly as it relates to vertebral number evolution. In this chapter, I characterize the process of somitogenesis and axial elongation in detail in *A. calliptera*, a species thought to closely resemble the common ancestor of all cichlid species in Lake Malawi (Malinsky et al., 2018).

3.1.1 Cichlids are an important system for understanding evolution and development

Lake Malawi cichlids exhibit both evolutionary and embryological traits which make them an attractive system to study the evolution of phenotypic diversity. Lake Malawi cichlids are best known for being a dramatic example of speciation and an adaptive radiation, having evolved around 800 species in just 800,000 years (Turner, 2007; Malinsky et al., 2018; Santos et al., 2023). Despite dramatic phenotypic differences between species, different Malawi cichlid species differ by only 0.2% nucleotide divergence, and are genetically capable of hybridization (Turner, 2007). Therefore, Lake Malawi cichlid fishes have evolved dramatic phenotypic diversity, and many different species, from a very limited substrate of genetic diversity.

3.1.2 Prior knowledge of *A. calliptera* development

Marconi et al., 2023 perform a detailed analysis of *A. calliptera*, *R. chillingali*, and *Tropheops* sp. mauve development, from fertilization to hatching. They show that, like other cichlid species, the Lake Malawi cichlids form comparatively small embryos on top of a very large, nutrient-rich yolk. Unlike in zebrafish, where 50%

of development occurs after the tailbud lifts from the yolk, cichlid tail eversion occurs at ∼28ss (or approximately 3/4 of somitogenesis) (Marconi et al., 2023).

Marconi et al., 2023 also show that the rate of somite formation is very similar between *R. chillingali* and *A. calliptera*, meaning that a change in segmentation clock rate does not explain somite number evolution in this species. Specifically, *R. chillingali* develop around 10% faster than *A. calliptera*, thus explaining two of the six difference in somite count between species. James Hammond has since repeated this work and found the same result (James Hammond, unpublished data). The appearance of key morphological features, such as development of the eye and otic vesicles, is also tied to absolute somite stage, not percentage of somitogenesis (Marconi et al., 2023). This implies that the developmental trajectory of each species is tied to somite stage, rather than total progression through somitogenesis.

3.1.3 Evo-devo of teleost axial elongation

Embryonic development has been well-characterized in zebrafish, a teleost approximately 224 million years diverged from *A. calliptera* (Kimmel et al., 1995; Kumar et al., 2022). Despite the significant evolutionary difference between these species, zebrafish are the closest relative to *A. calliptera* whose development has been well-characterized, so here, I will describe key aspects of zebrafish development during axial elongation and segmentation.

In zebrafish, the body axis is established by convergence and extension cell movements during epiboly (Kimmel et al., 1995; Shah et al., 2019). During epiboly, cells migrate posterior and towards the midline of the embryo, forming what will become the embryonic body axis, and the first three somites form (Kimmel et al., 1995; Shah et al., 2019). The tailbud then begins to form, developing first as a slight prominence at the posterior body, before becoming more morphologically distinct, and lifting from the yolk at around 16ss (Kanki & Ho, 1997; Steventon et al., 2016; Shah et al., 2019). Cells can migrate from lateral tissues of the embryo into the tailbud until around 12ss: at 10ss, cells are ingressing from dorsal tissues

into the presomitic mesoderm proper, and prior to 12ss, cells are able to migrate from lateral tissues into either the PSM, or directly into somites (Kanki & Ho, 1997; Lawton et al., 2013; Steventon et al., 2016). By 16ss, the tailbud has everted from the yolk, forming a distinct tail, and it continues to elongate as development proceeds (Kanki & Ho, 1997; Steventon et al., 2016).

The vast majority of teleosts, like zebrafish, complete epiboly either prior to somitogenesis, or very early in somitogenesis. To demonstrate this, I performed a partial literature review of teleost fish embryogenesis. I took a list of species from Roux et al., 2019, and supplemented it with additional species where development has been characterized. While this list of species may not be exhaustive, it represents a broad sampling of taxa from available literature. The results are summarized in Tables 3.2 and 3.1. I was unable to identify any species which complete epiboly late in somitogenesis (eg. after more than seven somites have formed). The exception are two species of African cichlids. *A. calliptera* do not complete epiboly until late in development, 28ss (Marconi et al., 2023). In addition, an early description of *Haplochromis piceatus*, from Lake Victoria, depicts embryos with approximately 20 somites where epiboly was not yet complete (Fleig, 1993). Therefore, haplochromine cichlids are highly unusual, if not unique, amongst teleost fish in delaying completion of epiboly until segmentation is nearly completed. This represents another reason to characterize *A. calliptera* development, with the aim of understanding why this unique developmental trait occurs.

3.1.4 Mechanisms of vertebrate axial elongation

Elongation of the vertebrate body axis involves a complex interplay between development of the different tissues of the tailbud and the body axis, such as the PSM, notochord, neural tube, and the somites themselves (Bénazéraf, 2019). The contribution of the presomitic mesoderm (PSM) to axial elongation is perhaps best understood, and is thought to be driven by changes in cell motility and the tissue micro-environment (Bénazéraf et al., 2010; Lawton et al., 2013; Bénazéraf

Table 3.1: Relationship between the onset of segmentation and epiboly in cichilds.

Species & Common name	Method of development	Citation
Cichlasoma dimerus (South American cichlid)	The blastopore closes at 22-23hpf, and six somites have formed by 28hpf	Meijide and Guerrero, 2000
Blood parrot cichlid (*Cichlasoma synspilum* x *C. citrinellum* hybrid)	Epiboly completes early in segmentation/before many segments have formed. Article in Chinese; figure captions in English.	Zhijing et al., 2014
Amphilophus xiloaensis (Midas cichlid species flock)	Six somites are visible at 90% epiboly (30hpf). 12 somite are present when epiboly completes. Thus, as with Oreochromis, some somites form prior to completion of epiboly, but epiboly still completes early in segmentation.	Kratochwil et al., 2015
Thorichthys meeki (Fire mouth cichlid, Central America)	Blastopore closure completes before appreciable numbers of somites formed.	(Gültekin & Yilmaz, 2023)
Oreochromis niloticus (Nile tilapia; cichlid)	By 75% epiboly, or 26-30hpf, three somites are visible in the embryo. By 32hpf, epiboly is complete and eight somites are visible. Therefore some somites are forming prior to completion of epiboly, but not as many as in *A. calliptera*.	Fujimura and Okada, 2007
Etroplus suratensis (endemic to India)	Epiboly completes prior to somitogenesis	Bindu and Padmakumar, 2016
Amatitlania nigrofasciata (Convict cichlid, Central America)	Epiboly completes before somites are visible	Jung et al., 2018
	Haplochromine cichlids	
A. calliptera (Eastern Happy; Lake Malawi cichlid)	By 30% epiboly, six somites are visible in the embryo. By 50% epiboly, around 15 somites are visible in the embryo. Epiboly completes at approximately 28ss.	Marconi et al., 2023
Haplochomis velvet black (Lake Victoria cichlid)	Mid-segmentation embryos which have not completed epiboly are shown.	Fleig, 1993

Table 3.2: Relationship between the onset of segmentation and epiboly in other teleosts. The list of species is taken from Roux et al., 2019.

Species & Common name	Method of development	Citation
Anguilla anguilla (Eel)	Three somites are visible at the blastopore closure period (19phf). Article in Japanese but staging table and captions in English.	Yamamoto et al., 1975; Sørensen et al., 2016
Astyanax mexicanus (Cavefish)	Epiboly completes at 10.5hpf, when the first somite is visible.	Hinaux et al., 2011
Danio rerio (Zebrafish)	Epiboly completes at 10hpf; there are no morphologically visible somites at this stage	Kimmel et al., 1995; Steventon et al., 2016
Carassius auratus (Goldfish)	Epiboly completes by around 11hpf. There are 1-6 somites already present at the end of epiboly, and a small amount of yolk is excluded from the embryo proper at the end of epiboly.	Tsai et al., 2013
Twin-tailed goldfish	Interestingly, though the twin-tailed goldfish appers to exhibit similar morphogeneic movements during gastrulation as the wild-type strain, blastopore closure is delayed, and the tailbud/PSM is larger and Kuppfer's vesicle is smaller.	Li et al., 2019
Gadus morhua (Cod)	The first somites form at around 75% epiboly, and there are seven to ten somite pairs at 100% epiboly, depending on the study. Studies differ in rearing temperature.	Hall et al., 2004; Gorodilov et al., 2008; Buslov et al., 2010
Tetraton nigrovirtidis (Pufferfish)	Epiboly is complete by 20hpf, and segments do not begin to form until epiboly is complete.	Zaucker et al., 2014
Notothenia coriiceps (Antarctic bullhead northern)	Epiboly is not complete until 20 days post fertilization! By 21dpf, three somites pairs are visible.	Postlethwait et al., 2016
Gasterosteus aculeatus (Three-spined stickleback)	Blastopore closure is complete by 24hpf, and by 30hpf six somites are visible.	Vrat, 1949
Paralichthys olivaceus (Olive Flounder)	Blastopore closure completes before somites are visible in the embryonic shield. Article in Japanese but staging table and captions in English.	Cerqueira, 2005; Jung et al., 2015

Continued on next page

Table 3.2: (continued)

Species & Common name	Method of development	Citation
Chromis limbata (Azores chromis damselfish)	Epiboly appears to complete prior to somitogenesis.	Laglbauer et al., 2017
Pomacentrus amboinensis (Damselfish)	Some somites are described as being formed towards the end of epiboly, but large numbers of somites continue to form after completion of epiboly.	McCormick and Nechaev, 2002
Neopomacentrus cyanomus (Damselfish)	Images are not entirely clear, but epiboly appears to complete prior to somitogenesis	Setu and Ajith Kumar, 2010
Amphiprion ocellaris (Clownfish)	Epiboly completes over an oval yolk by 21hpf. By 23hpf, three somites are visible.	Salis et al., 2021
Oryzias latipes (Medaka)	Somites do not develop until after epiboly is complete.	Iwamatsu, 2004
Xiphophorus macelatus (Platyfish)	Five to seven pairs of somite have formed at Stage 9, after the completion of epiboly; somites may be visible at Stage 8 or 80% epiboly. Illustrations appear to depict the tailbud continuous with lateral plate mesoderm even at 7ss.	Tavolga, 1949
Fundulus heteroclitus (Killifish)	The blastopore closes at Stage 20 and there are 3-4 somite pairs by Stage 21	Armstrong and Child, 1965
Kryptolebias marmoratus (killifish)	Somitogenesis begins after the completion of epiboly	Mourabit et al., 2011
Austrofundulus limnaeus (killifish)	Epiboly completes, then the body axis disperses, before reaggregating. Somitogenesis then begins.	Podrabsky et al., 2017
Gambusia holbrooki	The blastopore definitely closes; it is very difficult to distinguish somites from yolk globules in the embryo. Authors place completion of epiboly prior to segmentation.	Mousavi and Patil, 2022
Prochilodus lineatus	Epiboly completes prior to somite formation.	Ninhaus-Silveira et al., 2006
Thunnus thynnus (Atlantic bluefin tuna)	Epiboly completes prior to or early in segmentation	Ortega et al., 2024
Sarda sarda (Atlantic bonito)	Epiboly completes prior to or early in segmentation	Ortega et al., 2024
Cyprinus carpio (carp)	Epiboly completes prior to segmentation	Park et al., 2017

Continued on next page

Table 3.2: (continued)

Species & Common name	Method of development	Citation
Clarias batrachus (catfish)	There are eight to ten somites at the yolk plug stage	Rahman et al., 2013

et al., 2017; Mongera et al., 2018; Thomson et al., 2021; Stooke-Vaughan et al., 2025). In both chick and zebrafish, cells in different regions of the PSM differ in cell movement profiles (Lawton et al., 2013; Bénazéraf et al., 2010). Cells in the zebrafish posterior PSM are known to be more motile and exhibit less directed cell movement than those in the anterior (Lawton et al., 2013; Das et al., 2017; Mongera et al., 2018). This has been proposed to be due to a fluid-to-solid transition in the posterior body: cells in the posterior PSM are in a lower-density, fluid-like state, while cells in the anterior are in a less motile, solid-like state (Mongera et al., 2018). Simulations showed that this mechanism was sufficient to achieve directed elongation (Mongera et al., 2018; Banavar et al., 2021). Chick embryos also have differences in cell motility across the anterior-posterior axis: there is a cell motility, and cell density, gradient running anterior to posterior (Bénazéraf et al., 2010). This motility gradient is thought to explain axis elongation in chick (Bénazéraf et al., 2010).

Alongside the fluid-to-solid transition, cells in the zebrafish PSM are also increasing in density over time, in a process called compaction-extension (Thomson et al., 2021). This suggests that external forces are also modifying zebrafish development (Thomson et al., 2021).

Dorsal tissues, namely the neural tube, are also involved in overall body axis elongation. In zebrafish, cells move from the neural tube into the PSM; without this cell migration, the speed of elongation is reduced (Lawton et al., 2013; Saunders et al., 2025; Stooke-Vaughan et al., 2025). Additionally, vaculation of the notochord is required for extension of the body axis after segmentation is complete (McLaren & Steventon, 2021). Finally, once somites are specified,

they will themselves elongate, developing their characteristic chevron shape and contributing to the overall elongation of the body (Narayanan et al., 2021).

In this chapter, while I focus on characterizing the development of the PSM (which directly contributes to somite count evolution), I also characterize the morphogenesis and patterning of the notochord and neural tube. I describe the process I am characterizing as "axial elongation" because, while I focus on the development of the PSM and the posterior body, I am specifically interested in how PSM morphogenesis ultimately achieves body axis growth and elongation in cichlids.

3.1.5 Patterning of the PSM

The vast majority of papers studying teleost fish development characterize the gross morphology of the embryo, and do not provide sufficient detail to characterize PSM morphogenesis and axial elongation in detail (see citations in Table 3.2). However, medaka and killifish development has been characterized in more depth (Terasaki et al., 2006; Yokoi et al., 2007; Yasuhiko et al., 2008; Saud et al., 2021). Like zebrafish, medaka embryos exhibit dynamic, likely oscillatory, expression of *hey1* and *hes7*, though not *lunatic fringe* (Elmasri et al., 2004b). However, *fgf8* is not expressed in the Medaka tailbud, implying that this gene does not control the frequency gradient of the segmentation clock as in zebrafish and chick (Dubrulle & Pourquié, 2004; Yokoi et al., 2007). Thus, while the segmentation clock does appear to operate in medaka, the genes controlling it can differ compared to other vertebrates. This suggests that while the dynamics of the segmentation clock are conserved across teleosts and other vertebrates, the GRN controlling the segmentation clock is evolvable (Krol et al., 2011).

The function and expression of genes patterning the anterior-posterior axis of the PSM also appear to be conserved between medaka and zebrafish (Terasaki et al., 2006; Yokoi et al., 2007; Yasuhiko et al., 2008; Saud et al., 2021). As in zebrafish, *tbx16* and *msgn* are expressed in a posterior-to-anterior gradient (Terasaki et al., 2006; Yokoi et al., 2007; Saud et al., 2021). Also as in zebrafish, loss of *msgn* causes

a round, undifferentiated mass of cells to form in the posterior PSM of Medaka and killifish (Saud et al., 2021); loss of *tbx16* has a similar effect (Terasaki et al., 2006; Yokoi et al., 2007; Saud et al., 2021). This suggests that *msgn* and *tbx16* are required for cell motility in killifish and medaka, as in zebrafish (Terasaki et al., 2006; Yokoi et al., 2007; Saud et al., 2021). Expression and regulation of *mespa* and *mespb* are also similar between zebrafish and medaka, with both genes expressed in the posterior of the somite about to form, as well as within the *tbx6* domain (Terasaki et al., 2006; Cutty et al., 2012). As in zebrafish, medaka *tbx6* is required for expression of the *mespba* enhancer. Even more remarkably, this enhancer is able to substitite for the *mesp2* enhancer in mouse, suggesting a high degree of functional homology during patterning of these distantly related species (Yasuhiko et al., 2008). While patterning of the killifish (*Kryptolebias marmoratus*) PSM is poorly understood, patterning of *tbx16*, *fgf8*, and *tbx6* resemble zebrafish at mid-segmentation, though *tbxta* expression is absent from the notochord (Saud et al., 2021). The genes *msgn* and *noto* are likewise required for directed cell migration, and notochord formation, respectively (Saud et al., 2021).

In this chapter, I describe axial elongation in *A. calliptera* in detail. I build on the work of Marconi et al., 2023 by investigating the morphology of the posterior unsegmented body, where somites are produced, in higher spatial and temporal detail. I first characterize the onset of segmentation in detail, characterizing both the anatomy of the embryo, and its patterning, at the earliest stages of segmentation. I will characterize the overall anatomy of the PSM, as the arrangement and shape of different cell populations in the tailbud. I then characterize patterning of the tailbud and the PSM, to understand how and whether the genetic control of axial elongation differs between zebrafish and *A. calliptera*. This chapter lays the groundwork for the detailed quantitative comparison between *A. calliptera* and *R. chillingali* that I will perform in Chapter 4.

3.2 Results

To orient the reader, I will first describe the overall development of the embryo during axial elongation. I identify a number of stages of axial elongation, which will be described further below. Note that this staging system differs from that proposed by Marconi et al., 2023 as it is based on the morphology of the PSM and tailbud, not the embryo as a whole. The differences between these staging system are summarised in Table 3.3. I propose a system specific to axial elongation because the Marconi staging does not precisely identify the timing of transitions between stages, and does not distinguish between embryos before and after separation of the PSM and lateral plate mesoderm, a key morphological transition. This new staging system should only be used for studies of axial elongation and somitogenesis, not development in general.

Table 3.3: Comparison of staging in this chapter with Marconi et al., (2023). Gastrula/segmentation onset/early segmentation stages are Stage 9 in Marconi et al., 2023.

somite count	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32
Staging (Marconi et al., 2023)	10		11				11	12			12			30	
Staging (this chapter)	early segmentation						mid seg.					late seg.			

The first stage, "formation of the body axis", occurs between 0ss and 2ss. The notochord has formed at the medial axis of the embryo. The neurectoderm covers the majority of the embryo: it consists of a single layer of cells across most of the embryo, but is thicker in the future head region, and in the medial portion of the embryo (the future neural tube). Cell morphologies in the PSM are less organized than at later stages; there is no pseudoepitheial layer surrounding the PSM as at later stages.

The second stage, "Development of the PSM", occurs from 2ss to 5ss. Somites have formed, and have begin to epithelialize anteriorly and posteriorly, though not laterally. A pseudoepithelial layer has formed dorsally and ventrally to the

PSM, separating the PSM from the other tissues in these dimensions, though not laterally. A morphologically distinct set of cells has formed in the posterior of the embryo, posterior and dorsal to the posterior tip of the notochord. The neurectoderm still covers the entire PSM at the anterior, but has formed a distinct neural tube in the midline of the embryo.

After 5ss, development of anterior structures is well-described by Marconi et al., 2023 so I will not repeat this analysis here.

The third stage, "early segmentation", occurs between 5 and 15ss. During these stages, the neurectoderm retracts and neurulation proceeds; by 10ss the neural tube formation is completed.

The fourth stage, "separation of PSM", occurs at 15-16ss. At this stage, the pseudoepithelial layer of cells begins to form laterally to the PSM, separating the PSM from the lateral plate mesoderm. This stage marks a key morphological transition for the embryo, as after this stage, lateral ingression of cells into the PSM is no longer possible.

The fifth stage, "mid segmentation", occurs between 15 and 25ss in *A. calliptera*. At this stage, separation of the PSM from lateral plate mesoderm is complete. An obvious, elongate pseudoepithelial layer has formed around the PSM; this is morphologically distinct to inner mesenchymal cells.

At 25ss, tail eversion occurs. After this stage (25-30ss), or "late segmentation", the PSM is structured similarly to at mid segmentation; but has turned on its side so that it is always imaged from the lateral aspect. Figure will show this more clearly.

In the following sections I will describe these stages in more detail.

3.2.1 Formation of the body axis: 0ss

The cichlid embryo develops as a thin, transparent layer of tissue on top of a very large, yellow, fatty yolk [1] . This makes staging very early stages of development

[1] It is clear that the yolk is fatty because, when live embryos are dissected, large globules of fat are removed from the yolk.

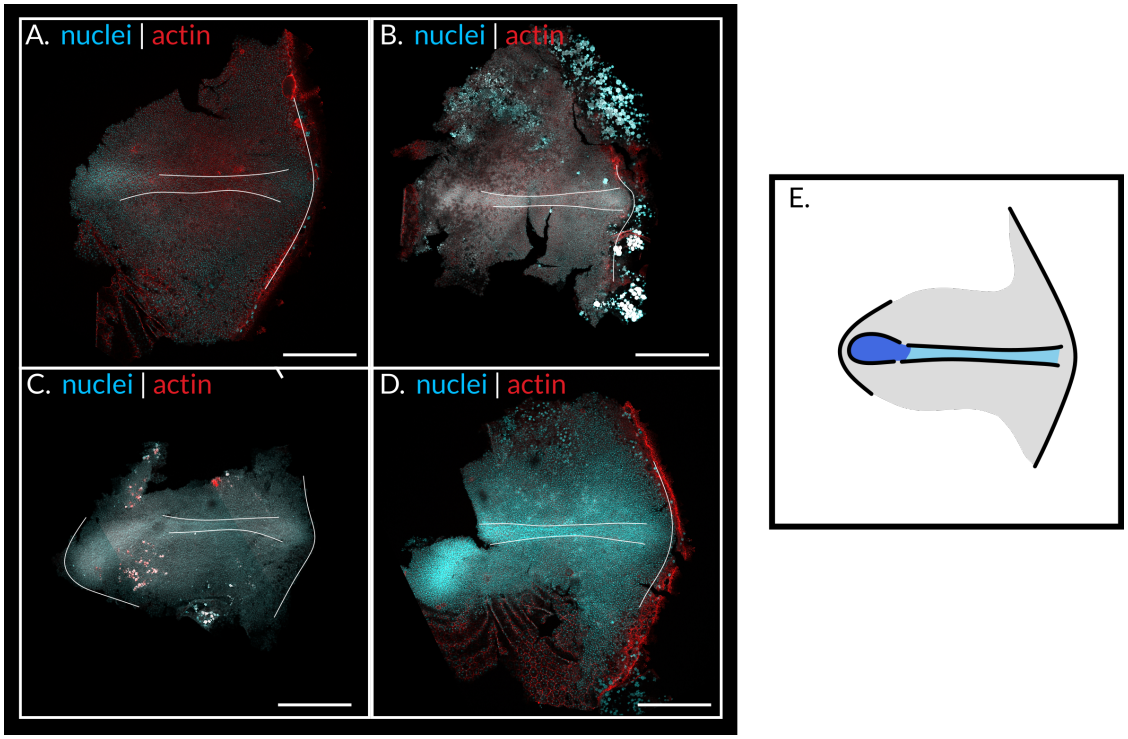


Figure 3.1: Anatomy of the embryo prior to somite formation. A-D: Individual embryos, flat mounted and dissected from the yolk, prior to somite formation. Embryos are *A. calliptera* stained with DAPI (cyan) and phalloidin (red). Note that embryos are extremely fragile at this stage and therefore some are damaged, as follows: A: anterior of the embryo, in the top 50% of the head, has been ripped off. B: Yolk was not fully dissected from the embryo, leading to autofluorescent fat globules, and shadows where fat has blocked signal. The embryo is also slightly torn at the midline. C: Embryo is largely complete; the lateral plate mesoderm at the top of the image has been torn. D: The embryo head has torn and is facing downwards. E: Schematic of embryo anatomy at this stage. All embryos are imaged from the dorsal aspect and presented anterior left. White lines indicate the anterior and posterior boundary of the embryo, and boundaries of the notochord and neural tissues. Scale bar: 500μm.

challenging, as the embryo is difficult to visualize under widefield microscopy (Marconi et al., 2023). The first sign of a body axis I was able to detect under widefield microscopy was a line of transparent, thicker tissue on one side of the embryo. Dissecting embryos at this stage yielded some embryos prior to the onset of segmentation. Note that at this stage, the embryo is extremely fragile (only a few layers of cells thick), and so some embryo damage during dissection is difficult to avoid. Autofluorescence from the yolk made imaging the embryo prior to dissection difficult.

I dissected these embryos, and stained them with DAPI and phalloidin

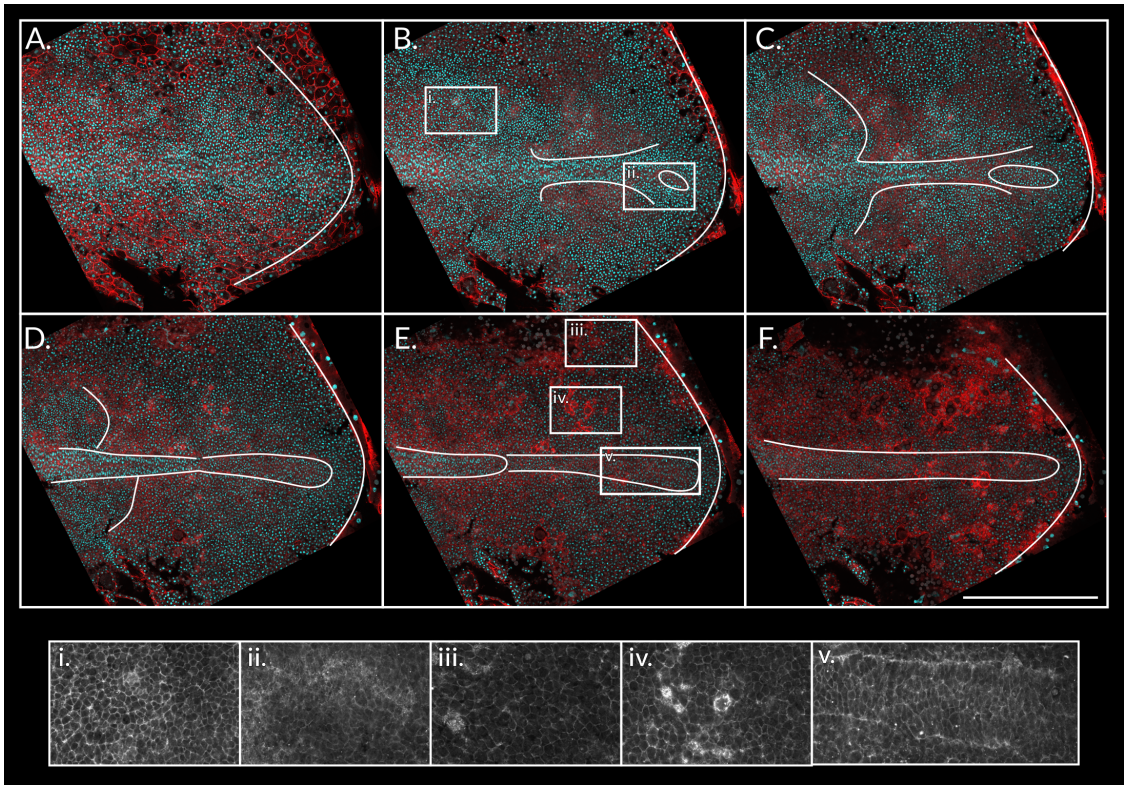


Figure 3.2: Anatomy of the PSM at 0ss. A-F: single optical sections of an *A. calliptera* embryo stained for DAPI (cyan) and phalloidin (red) to visualize nuclei and actin-labelled cell membranes, respectively. Each optical section is taken from evenly spaced regions of the embryo. i-v: insets showing structures of interest at higher resolution. Phalloidin staining is shown in greyscale. i: small, tightly packed neurectoderm cells. ii: Protrusion of the notochord into the neurectoderm layer. iii: Semi-organized cells in the lateral plate mesoderm. iv: semi-organized cells in the paraxial mesoderm. v: Shape of cells in the developing notochord. All embryos are imaged from the dorsal aspect and presented anterior left. White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues. Scale bar: 500μm.

to visualize cell nuclei and membranes, before imaging them under confocal microscopy. Prior to the first somite forming, the embryo is variable in anatomy, likely because we are conflating multiple stages together (Figure 3.1A-D). Even very early embryos have an obvious midline, consisting of the neural tube on top of the notochord. The lateral plate mesoderm is indistinguishable from the pre-somitic mesoderm (Figure 3.1A-D, Figure 3.2). The neurectoderm is forming anteriorly; this structure is obvious because it has smaller, elongate cells that lie on top of the mesoderm (Figure 3.2B, i). These cells express the neural marker *sox2* (Figure 3.3).

The neurectoerm lies dorsal to the rest of the embryo (Figure 3.2A, B). At early stages, this cell layer covers the entire embryo (Figure 3.2A). An actin-rich membrane separates the neurectoderm from the mesoderm (Figure 3.2C). The ectoderm will change in shape dramatically between 0ss and 7ss: it is initially a single sheet of cells covering the entire embryo (though thicker at the anterior, which will form the head). The neural tube then begins to form medially. However the neural tube only has the characteristic tube shape at the center of the embryo (50%EL); at the posterior, it is still very broad, and in fact will retain this "paddle shape" until well into segmentation (7ss, see below).

The notochord is visible at the midline of the embryo. The cellular anatomy of this structure differs to later stages: compared to mid-segmentation embryos, cells are less elongate and have not yet intercalated, and the notochord is multiple cells wide (Figure 3.2E, v). The notochord is also thicker than at later stages, being approx 60μm wide in at 50%EL, and around 150μm wide in the very posterior (Figure 3.2E). The notochord also protrudes into the neurectoderm dorsally: at the posterior tip of the notochord, only a single layer of neurectoderm covers the notochord (Figure 3.2B).

Gene expression at the onset of segmentation

I investigated pattern formation within the early embryo, by staining for neural and mesodermal cell fate markers. In zebrafish and most other vertebrates, *tbxta* marks the notochord, a mesodermal derivative, and the mesoderm in the posterior of the embryo (Schulte-Merker et al., 1992; Bruce & Winklbauer, 2020). In *A. calliptera* at 1ss, *tbxta* is expressed in the notochord and in the germ ring (Figure 3.3C,E). As in zebrafish, *tbx16* is expressed in the germ ring at the posterior of the embryo, and within the PSM proper and the lateral plate mesoderm (Figure 3.3B,C,F) (Griffin et al., 1998; Thisse et al., 2004). In zebrafish, *tbx6* is expressed in a graded expression domain in the PSM, and is not expressed in the lateral plate mesoderm, even at very early stages (80% epiboly) (Nikaido et al., 2002).

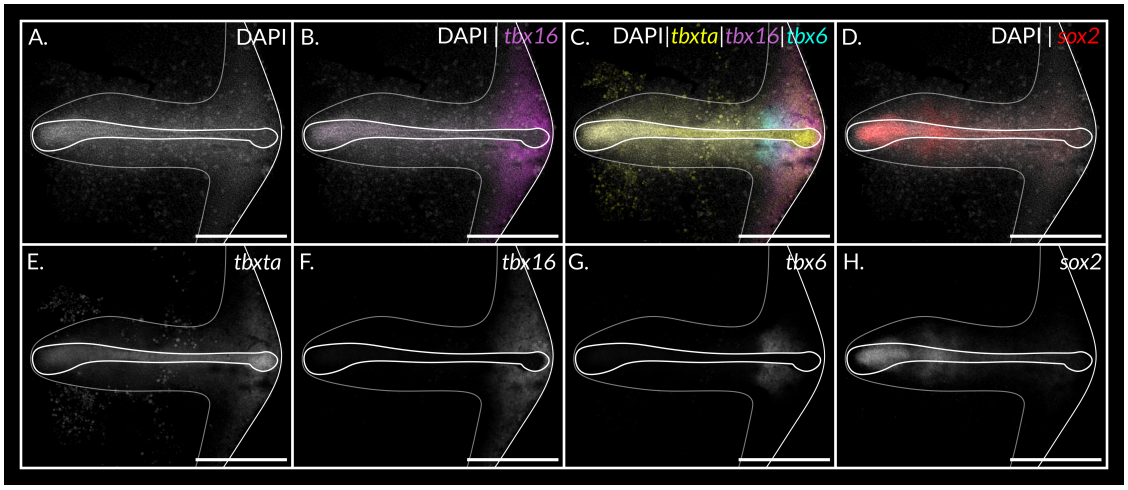


Figure 3.3: Tbox gene expression in the embryo at 1ss. Images are single confocal slices of an *A. calliptera* embryo, stained with DAPI (white), *tbxta* (yellow), *tbx16* (magenta), *tbx6* (cyan), and *sox2* (red). Embryo is oriented anterior left and imaged dorsally. The embryo and the ectoderm/neural tube are outlined in white. Scale bar: 500μm.

Like in zebrafish, *A. calliptera tbx6* is constrained to the presumptive PSM and is not expressed in the lateral plate mesoderm (Figure 3.3G).

At this stage, there limited local gene expression heterogeneity in the mesoderm: though there are gradient expression domains, the majority of cells' expression resemble their neighbours. The exception to this a population of larger, circular cells, expressing *Pouf5f1/oct4* but not *tbx6* (Figure 3.4). This cell population is both morphologically and transcriptionally distinct. As *oct4* is a pluripotency marker, these cells are likely undifferentiated; this notion is further supported by the fact that they disappear after the 2ss stage.

To investigate whether travelling waves of segmentation clock gene expression precede somite formation, as occurs in zebrafish, I then investigated the expression of the segmentation clock gene *delta-c* prior to somite formation (Figure 3.5). In zebrafish, the expression of *dlc* changes depending on the phase of the segmentation clock: during some phases, *dlc* is only expressed as bright puncta within the nuclei of each cell, and at other phases *dlc* is expressed in either the nuclei, cytoplasm, or both structures (Mara et al., 2007). Three domains of *dlc* are visible in the *A. calliptera* embryo. The brightest is a thick expression domain of nuclear and cytoplasmic *dlc*, about half-way through the PSM (Figure

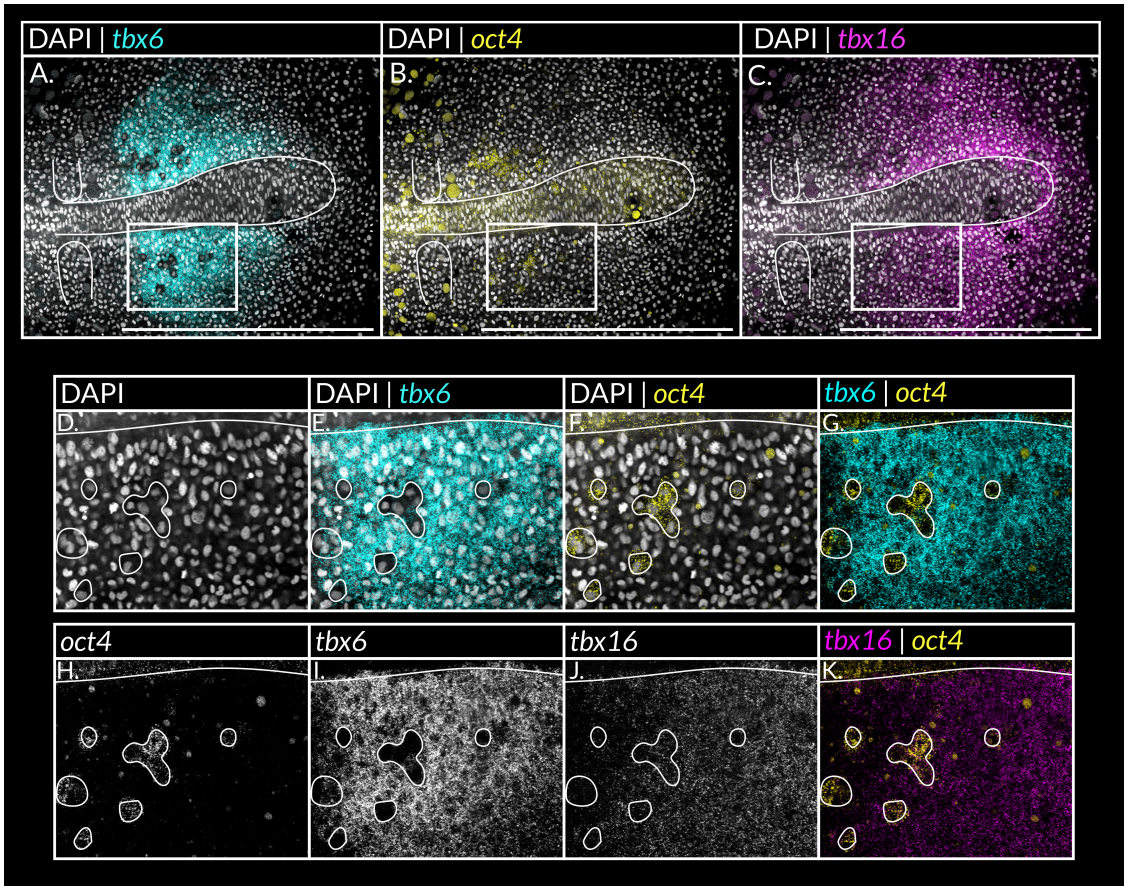


Figure 3.4: Heterogenous *Oct4/Pouf5a* expression in 1ss *A. calliptera*. Images are single confocal slices of embryo, stained with DAPI (white), *tbx6* (cyan), *oct4* (yellow), and *tbx16* (magenta). A-C: depiction of the entire PSM. Note that oct4 was detected using an Alexa-488 flurophore, which also detects large circular fat gobules (B); genuine *oct4* expression is identified via punctate signal in F. D-K: magnifications of the unsegmented mesoderm; greyscale of individual genes is presented in H-J. As above, a white line outlines the neural tube and nascent somite; in D-K, Oct4-expressing cells are also outlined in white.

3.5). Anterior to this domain are a small number of cells expressing *dlc* in the nuclei and cytoplasm (Figure 3.5). The posterior of the embryo is in a different phase of the segmentation clock, as expression there is punctate and restricted to the nucleus (Figure 3.5). Therefore, there are travelling waves of gene expression present in the embryo at very early stages, as expected. As expression of *mespaa*, a marker of somite specification, was not detected in these embryos, it is unclear whether these travelling waves of gene expression represent oscillations of the segmentation clock prior to the first somite forming, or are marking the earliest somite.

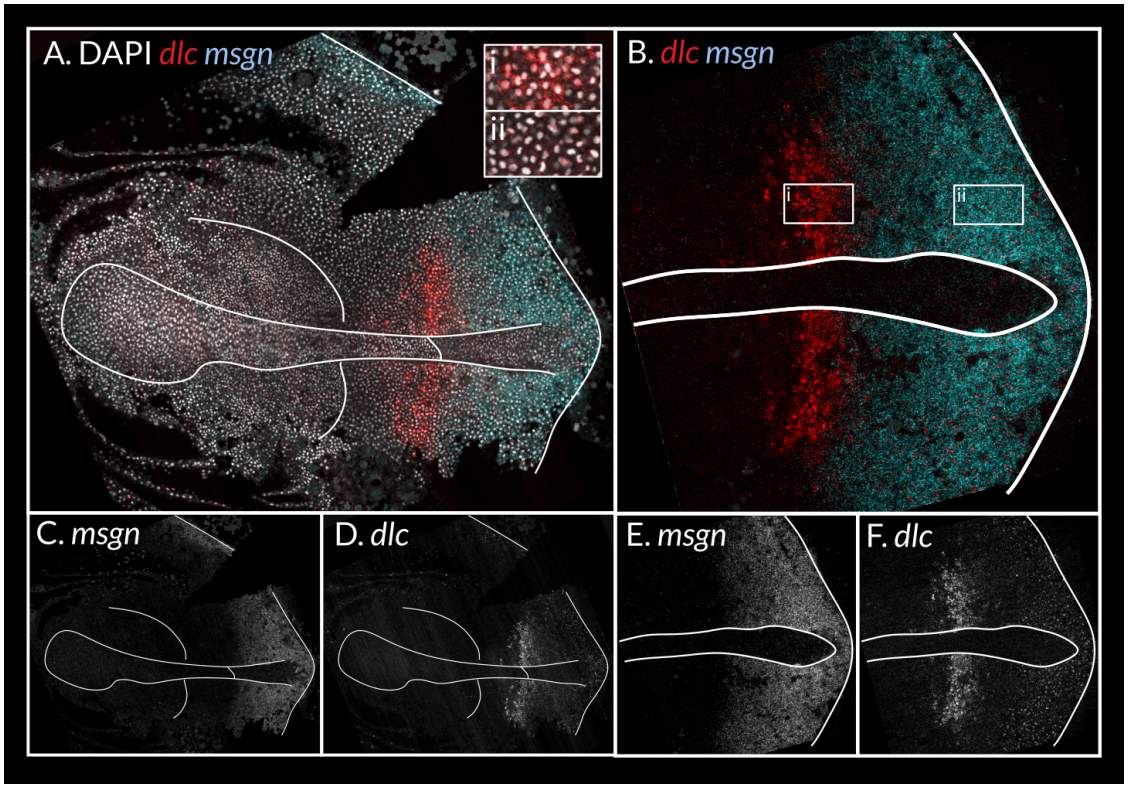


Figure 3.5: Travelling waves of *delta-c* in the early embryo. Images are is single confocal sections of a flat-mounted *A. calliptera* embryo, stained for DAPI (grey), *dlc* (red), and *msgn* (cyan). A: whole embryo. B: Zoom of the PSM. C-F: individual channels. All embryos are imaged from the dorsal aspect and presented anterior left. White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues. i, ii: inset depicting differing subcellular localization of *dlc* (red). White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues, and somite boundary. Scale bar: 500μm.

3.2.2 Structural integrity of the PSM arises at 2-4ss

By 2ss, the PSM has adopted a more complex structure and the majority of structures present at 5ss onwards have formed. The PSM is still covered by the neurectoderm layer, however, and like at the onset of segmentation, the notochord protrudes into the neural tube, being covered by only a single layer of tissue (Figure 3.6B). An actin-rich membrane separates the neurectoderm from the mesoderm (Figure 3.6B,C). The neural tube is better defined, and has formed the distinctive lumen at the midline (Figure 3.6C, D).

At this stage, somites have begun to epithelialize: a clear epithelial layer separates the dorsal and ventral surfaces of neighbouring somites, and a less

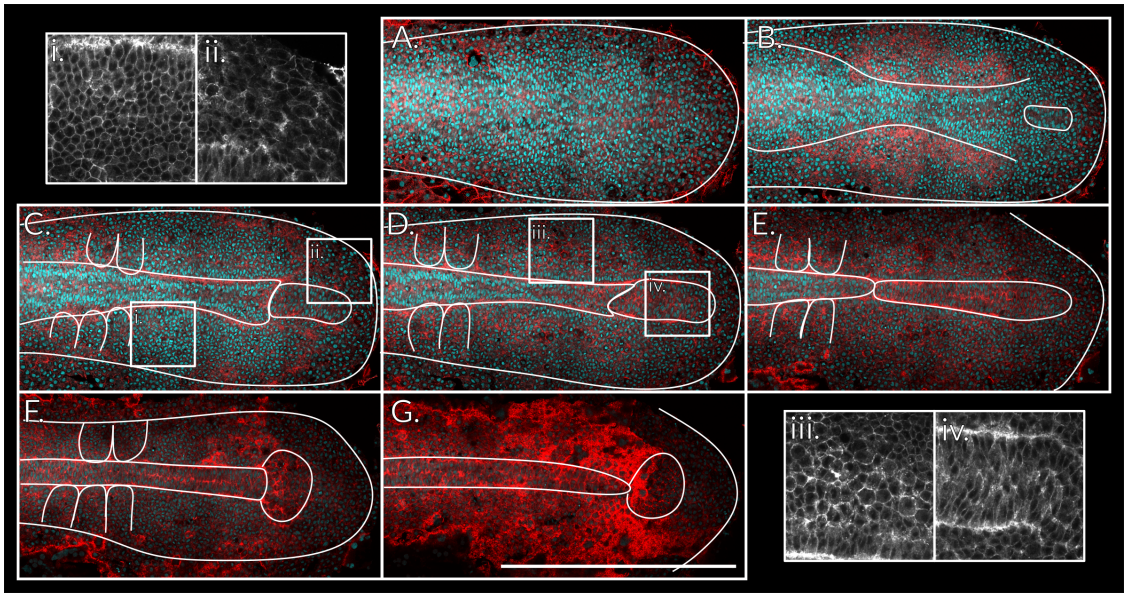


Figure 3.6: Anatomy of the PSM at 2-4ss. A-G: single optical sections of an *A. callipterea* embryo stained for DAPI (cyan) and phalloidin (red) to visualize nuclei and actin-labelled cell membranes, respectively. Each optical section is taken from evenly spaced regions of the embryo. A: most dorsal; G: most ventral. i-iv: insets showing structures of interest at higher resolution. Phalloidin staining is shown in greyscale. i: an organized cell layer surrounds the PSM dorsally and ventrally. ii: Cells in the posterior mesoderm are poorly organized. iii: Cells within the mesoderm are less well organized than the surrounding layer (compare to i). iv: Cell shape in the notochord. All embryos are imaged from the dorsal aspect and presented anterior left. Scale bar: 500μm.

well-organized epithelium separates somites from the notochord and neural tube medially (Figure 3.6C). Laterally, however, somites have not epithelialized, instead remaining open to the lateral plate mesoderm until 10ss (Figure 3.6C). This means that it may be possible for cells from the lateral plate mesoderm to ingress directly into somites, as occurs in zebrafish at early stages (Steventon et al., 2016).

By 2ss, a psuedo-epithelial layer has formed on the dorsal and ventral layers of the PSM (Figure 3.6C, i). This layer of cells is absent laterally (Figure 3.6C, i). Cells in this layer tend to be more organized, with more regular packing, and are elongate. In contrast, cells in the middle layer — that I will afterwards refer to as the *mesenchymal core* — are arranged in a less orderly manner, and are not elongate (Figure 3.6D, iii).

It is unclear if cells in the pseudo-epithelial layer, and the core, are in distinct transcriptional states, whether cells can transition between layers, and whether

cells in either layer have undergone an epithelial to mesenchymal transition (or mesenchymal to epithelial transition). At later stages, patterning of the PSM is similar between tissue layers for all genes so far examined (e.g. Figures 3.8, 3.18). It is clear, however, that the pseudo-epithelial layer demarcates the PSM from external tissue dorsally and ventrally; but at this stage, not laterally (Figure 3.6C, D, F).

There are also a group of cells in the posterior of the embryo, which are distinct in structure from cells in the anterior PSM (Figure 3.6C, iv). Cells in this layer are elongate, with their long axis oriented perpendicular to the midline, in contrast to cells in the pseudo-epitheial layer, which are oriented towards the midline (Figure 3.6C, ii).

3.2.3 The early segmentation stage: 5-15ss

The main morphological transition in the unsegmented region between 5 and 10ss is the retraction of the neurectoderm from a structure covering the entire posterior body, to a paddle-shaped structure at the posterior of the body, to a real neural tube.

The overall structure of the PSM between 5 and 10ss is similar to that at 4ss. At these stages, the embryo is continuing epiboly: that is, the posterior-most margin of cells in the embryo is moving posteriorly as the embryo continues to elongate. In dissected embryos, the embryonic margin is sometimes visible as a bright, actin-rich membrane at the posterior of the embryo, directly behind the posterior boundary of the PSM (eg. Figure 3.7). This close association of the posterior margin of the PSM and the edge of epiboly suggests that in cichlids, the process of epiboly is linked to that of axial elongation, at least until separation of the PSM from the lateral plate mesoderm.

Some morphological differentiation of notochord cells has occurred: cells in the anterior PSM, adjacent to the nascent somite, have intercalated and formed long, thin cells (Figure 3.7). However, cells in the posterior notochord have not yet intercalated and elongated (Figure 3.7).

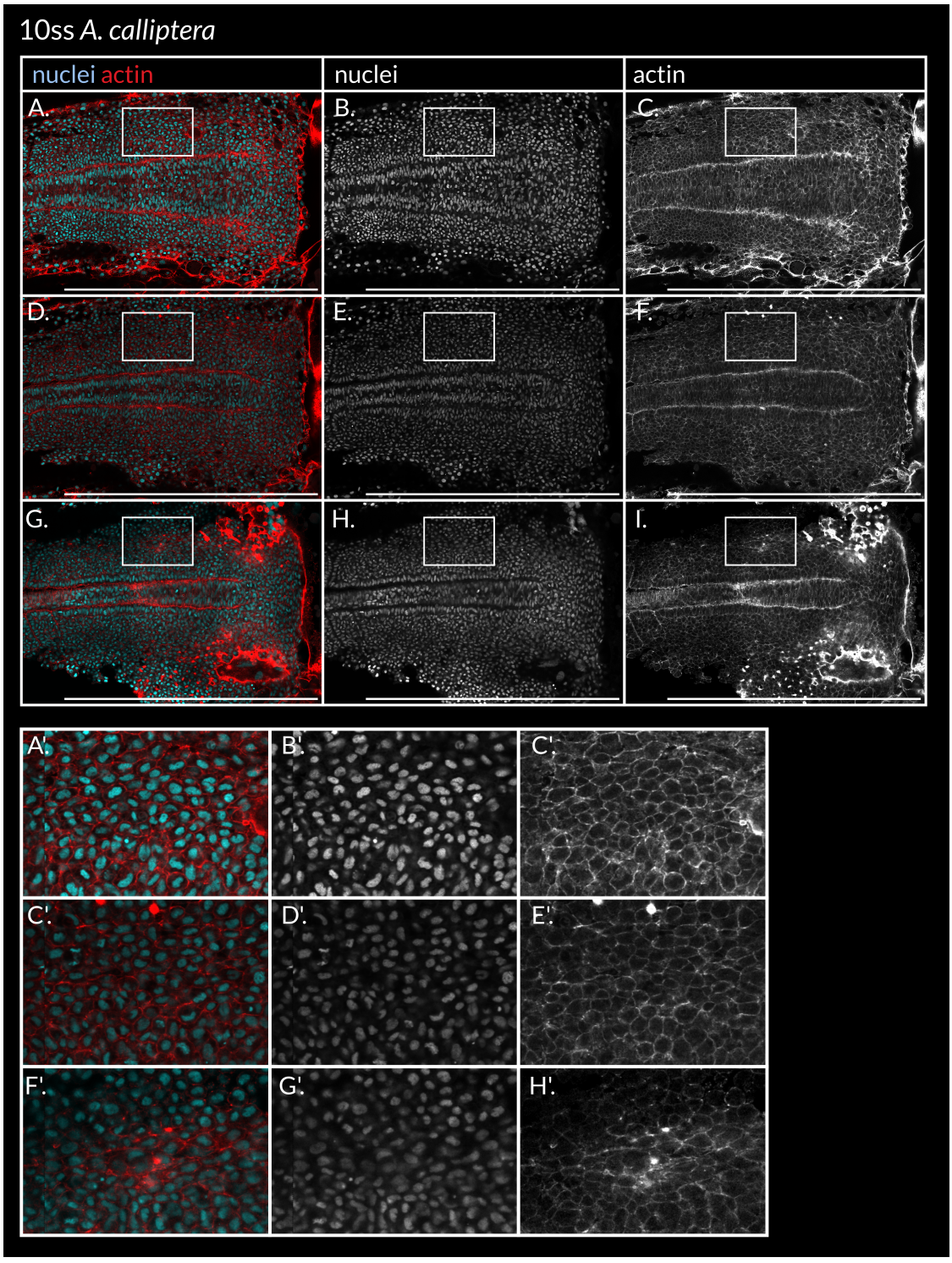


Figure 3.7: Anatomy of the PSM at 10ss. Figure depicts different focal planes of a 10ss embryo. A-C: most dorsal focal plane; through the neural tube. D-F: middle focal plane. G-I: most ventral focal plane, through the bottom half of the notochord. A'-H': Magnification of the region indicated in the top panels. Each panel is a single confocal section of dissected, whole-mount *A. calliptera* embryos, at different z-positions within the embryo. The embryo is stained for DAPI (cyan or white) and phalloidin (red or white). Embryo is imaged from the dorsal aspect and presented anterior left. Scale bar: 500μm.

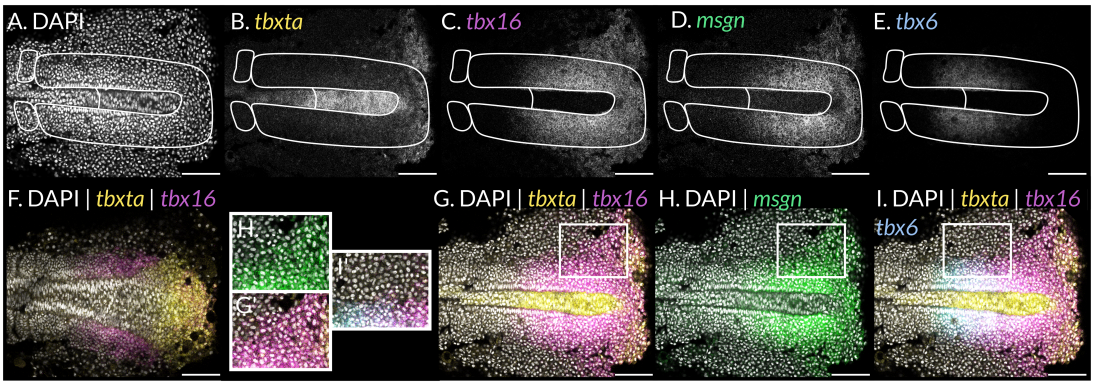


Figure 3.8: Patterning of *A. calliptera* at 5ss. Scale bar: 100μm. White lines outline the unsegmented mesoderm, nascent somite, notochord, and neural tube. Embryo is stained for tbxta, tbx16, msgn, and tbx6. F: different focal plane to illustrate *tbxta* expression into the neural tube. H', I', J': magnifications of corresponding panels.

The expression domains of mesodermal marker genes at 5ss have shrunk compared to expression domains at 1ss. *tbxta* is expressed in the notochord and the posterior PSM. Expression of *tbxta* is also detectable in about five cells adjacent to the posterior margin of the embryo, including into the lateral plate mesoderm (Figure 3.8B, G). There also appears to be neural expression of *tbxta*: at 5ss, the neural tube is a large, paddle-shape structure, with elongate nuclei (Figure 3.8). Nuclei remain elongate up until the posterior of the body embryo, suggesting that the neural tube extends all the way to the posterior of the embryo. *tbxta* is expressed in the posterior neural tube (Figure 3.8F). While the extent of this neural expression of *tbxta* is surprising, in zebrafish, there is some expression of *tbxta* in the neural tube (Toh et al., 2022), and at later stages *tbxta* is expressed in the *A. calliptera* neural tube.

Both *tbx16* and *msgn* are expressed in overlapping domains in the PSM and lateral plate mesoderm (Figure 3.8C,D); however, unlike at earlier stages, gene expression in the lateral plate mesoderm does not extend as far anteriorly as it does in the PSM (Figure 3.8C,D). Expression of *tbx6* is restricted to the PSM proper, much like in earlier stages (Figure 3.8E).

By 10ss, there is still morphological continuity between the PSM and the lateral plate mesoderm, and the pesudo-epithelial layer is yet to form laterally

(Figure 3.7). There is also little qualitative difference between cells in the PSM and those in the lateral plate mesoderm. However, cells in the mesodermal progenitor zone (posterior to the notochord) have a different morphology to those in the PSM proper. This group of cells tend to be more elongate compared to those in the PSM proper, and are more disordered in arrangement (Figure 3.7). The neural tube appears to have reduced in width: at its posterior boundary, it no longer spreads across the entire width of the embryo (Figure 3.7). By 10ss, the newly formed somites have also epithelialized laterally, so are fully demarcated from other tissue at this stage (Figure 3.7).

3.2.4 Separation of the PSM from the lateral plate mesoderm: 14-16ss

In *A. calliptera*, the PSM fully separates from the lateral plate mesoderm between 14 and 16ss (Figures 3.9, 3.10). This marks a key morphological transition: after separation of the PSM from the lateral plate mesoderm, there can be no further addition of cells from the lateral plate mesoderm to the PSM, and volumetric growth of the unsegmented body becomes more difficult. Consistent with this idea, the dynamics of growth also change between species at this time, presumably because this difference in embryological constraint (see Chapter 4, Figure 4.2). Because this transition controls an important constraint on volumetric growth, I will here describe this process in *A. calliptera*.

At 14ss, while the lateral plate mesoderm is still present beside the PSM, it is clear that these tissues have begun to separate. There is significant disorder in the actin boundaries of cells between these two layers, especially in the dorsal-most focal plane (Figure 3.9). Additionally, the posterior of the PSM has become more curved, like the PSM at later stages (Figure 3.9). This is in contrast to the embryo at 10ss, which has a straight posterior boundary. This change in curvature appears to be a signature of separation from the lateral plate mesoderm.

At 15ss, there is a clear separation between lateral plate mesoderm and PSM, with the formation of a pseudo-epithelial layer laterally, which although present remains at this stage fairly disordered (Figure 3.9). There is a region between

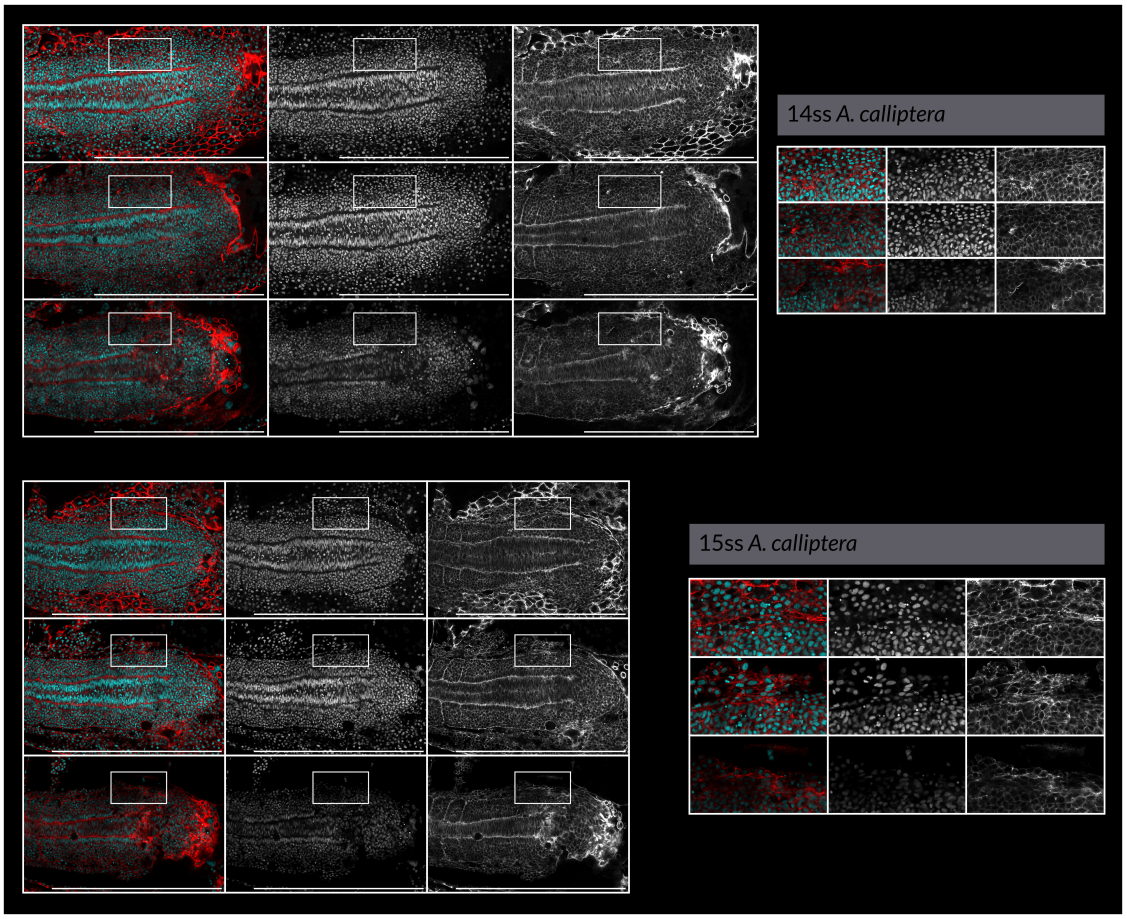


Figure 3.9: Detail of the boundary between the PSM and the lateral plate mesoderm. Images shown are single confocal sections of dissected, whole-mount 14 and 15ss *A. calliptera* embryos, at different z-positions within the embryo. Each embryo is stained for DAPI (cyan or white) and phalloidin (red or white). Embryos are imaged from the dorsal aspect and presented anterior left. Scale bar: 500μm.

the lateral mesoderm and PSM which lacks cell nuclei, and some actin-based membranes between these tissues (Figure 3.9). Some cells in the lateral plate mesoderm have also adopted very round shapes.

By 16ss, the separation between lateral plate mesoderm and PSM is largely complete (Figure 3.10). While there are still cells lateral to the PSM, it is clear that a boundary formed between these two cell populations (Figure 3.10). The lateral psuedo-epithelial layer has also become more regular in shape (Figure 3.10). This structure begins forming in the anterior of the PSM and is present more posteriorly as development progresses.

In one of these embryos, the Kupffer's vesicle is well-preserved. It is present

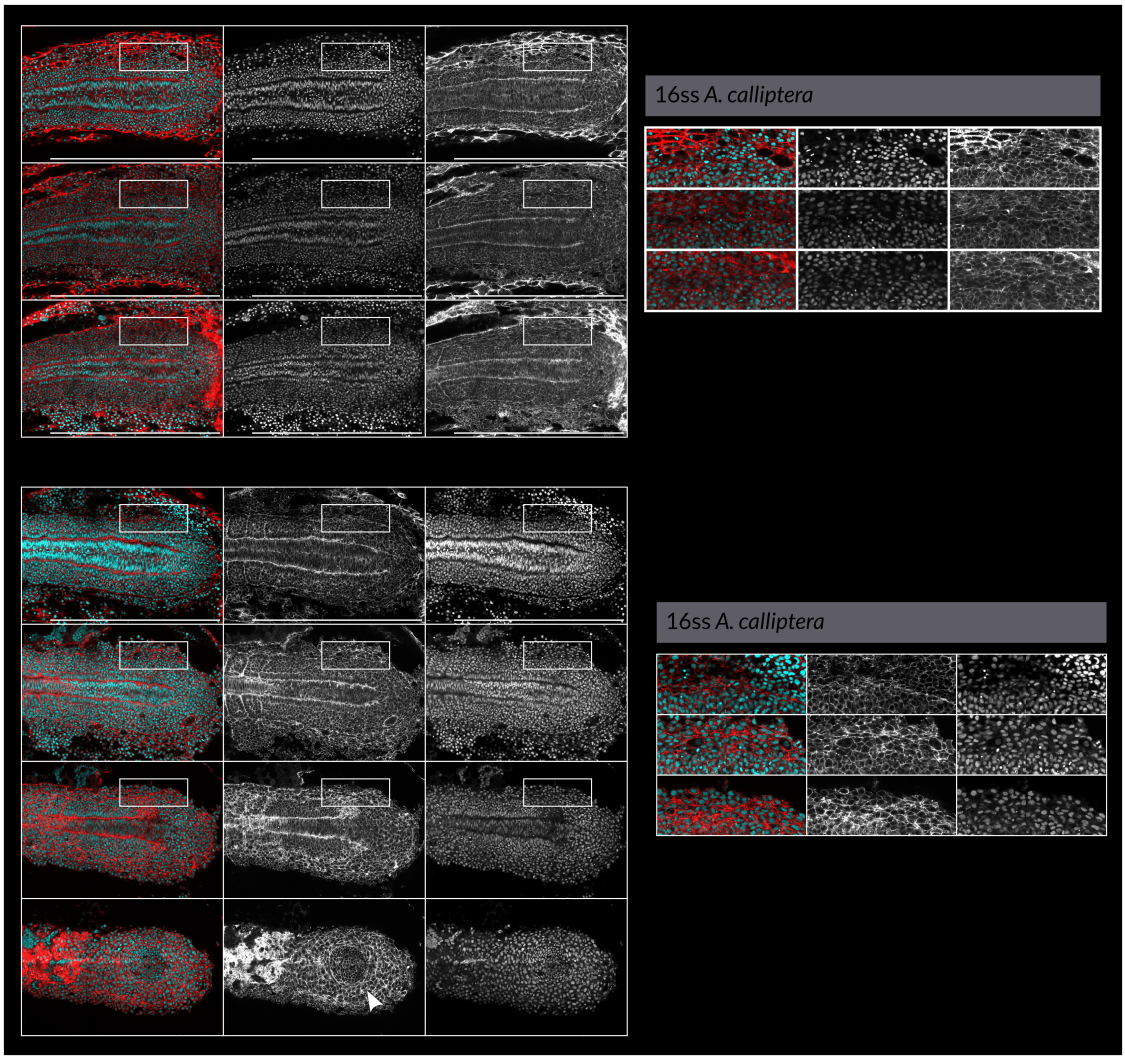


Figure 3.10: Separation of the PSM from the lateral plate mesoderm Images shown are single confocal sections of dissected, 16ss *A. calliptera* embryos, at different z-positions within the embryo. Each embryo is stained for DAPI (cyan or white) and phalloidin (red or white). Embryos are imaged from the dorsal aspect and presented anterior left. Arrow indicates position of Kuppfer's vesicle. Scale bar: 500μm.

at the ventral side of the embryo, as a strikingly geometrically arranged group of cells, with bright, straight cell boundaries (Figure 3.10). This structure is present in early embryos as a transparent, circular blob at the posterior ventral aspect of the embryo; unfortunately it is often lost during yolk removal and therefore not always present to analyse[2].

[2]During dissection, I remove yolk from the PSM by brushing it off with a paintbrush; this agitation can also remove Kuppfer's vesicle.

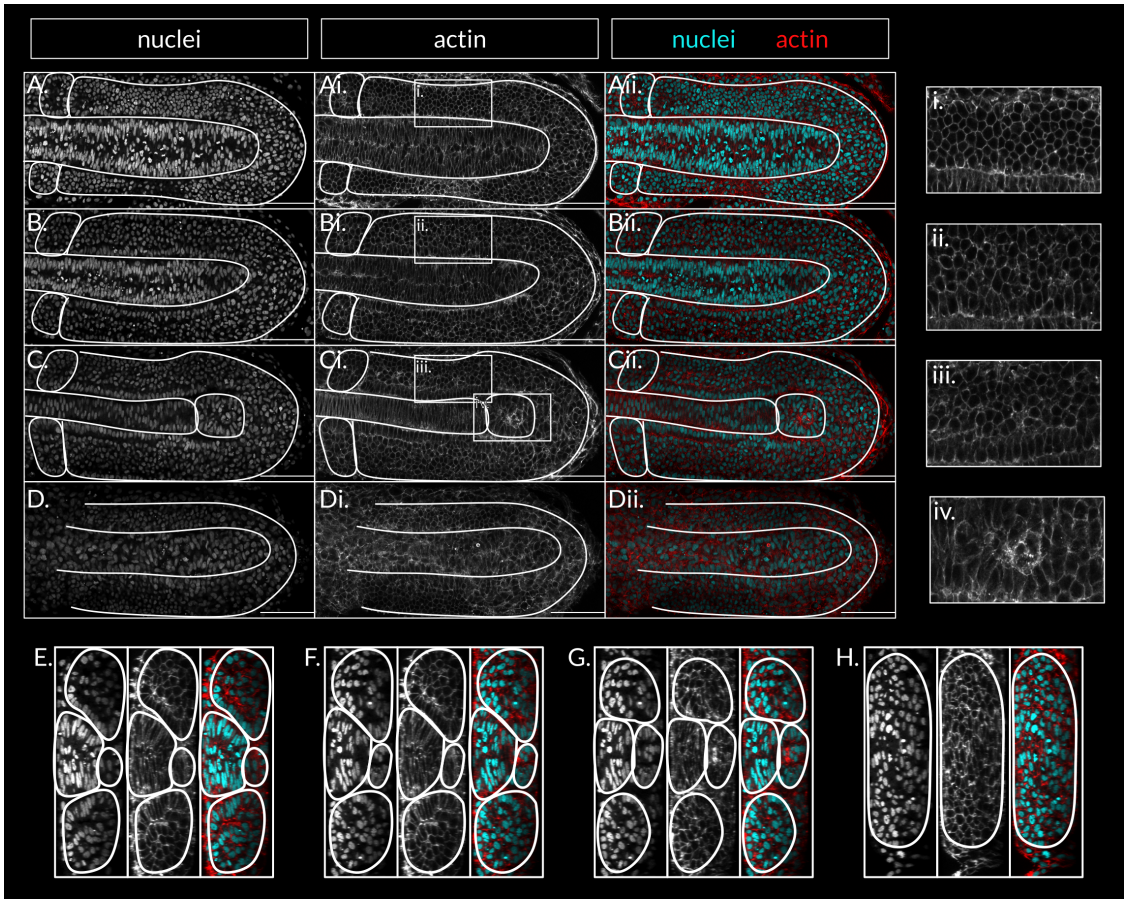


Figure 3.11: Anatomy of the PSM at mid segmentation. A-D: single optical sections taken at different dorso-ventral positions in the PSM. More dorsal focal planes are at the top of the image. Left panel: DAPI (nuclei) staining. Middle panel: phalloidin (actin) staining. Right panel: merged image, with nuclei in cyan and actin in red. i-iv: phalloidin stained insets of features of interest. White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues, and somite boundary. Images A-D and i-iv are arranged anterior left, scale bar represents 100μm. E-H: Reslices of the embryo, showing the dorso-ventral aspect of the embryo at different positions along the anterior-posterior axis. Bilinear interpolation was used between slices to ensure isotropic scaling of pixels. Left panel depicts nuclei, middle depicts actin, and right depicts merged (nuclei: cyan; actin: red).

3.2.5 Distinct cell types in the embryo: 20ss

By 20ss, there are multiple distinct cell types in the PSM. These are detectable both morphologically (by cell shape) and sometimes also by gene expression.

A pseudo-epithelial layer of cells surrounds the PSM dorsally, ventrally, laterally, and medially (Figure 3.11A, E). This pseudo-epithelial layer of cells is better organized than cells in the mesenchymal core, and cells inside it tend

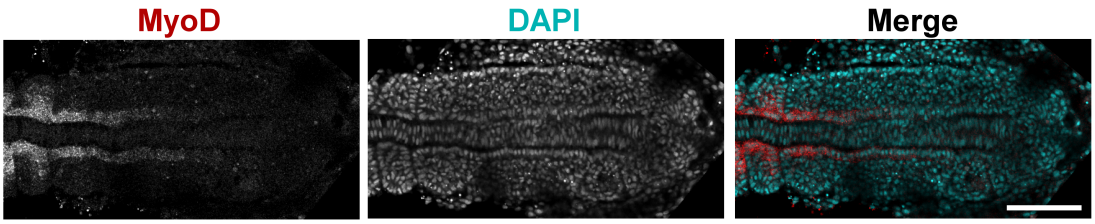


Figure 3.12: MyoD expression in adaxial cells. Left from right, panels depict *MyoD*, DAPI, and merged (DAPI cyan, *MyoD* red). Image is a single optical section of a 19ss *A. calliptrea* PSM through the notochord, oriented anterior left. Scale bar: 100μm.

to be more elongate (Figure 3.11).

In zebrafish, there is an adaxial layer of cells in the PSM directly adjacent to the notochord (Daggett et al., 2007). These cells will ultimately form the slow-twitch muscles of the myotome, and therefore represents a distinct cell type. This cell type is present in *A. calliptera*: in *A. calliptera* there is an adaxial layer of cells directly adjacent to the notochord (Figure 3.11B, C). These cells are more regular, and more elongate, than cells in the PSM core, but resemble cells in the pseudo-epithelial layer (Figure 3.11B, C). They are transcriptionally distinct from the rest of the pseudo-epithelial layer, however, expressing MyoD (Figure 3.12). Therefore, despite the morphological similarities between these cell types they appear to be different cell types.

3.2.6 Late segmentation: 25ss onwards

At 25ss, *A. calliptera* undergo tail eversion, where the tailbud changes morphology: the two halves of the PSM come together, and the tailbud fully separates from the yolk. After this stage, the two halves of the PSM surround the notochord and the neural tube lies dorsally to the rest of the tailbud (Figure 3.13). Despite changes in the arrangement of tissues in the tailbud, cell shapes in the tailbud are similar to that at mid-segmentation stages. As before, there is a pseudo-epithelial layer of cells surrounding the PSM; these cells are more regular and tightly packed than those within the PSM itself (Figure 3.13A, i). There is also an elongate adaxial layer of cells adjacent to the neural tube (Figure 3.13B, ii). Unlike at earlier stages, however, there does appear to be a population of notochord progenitor

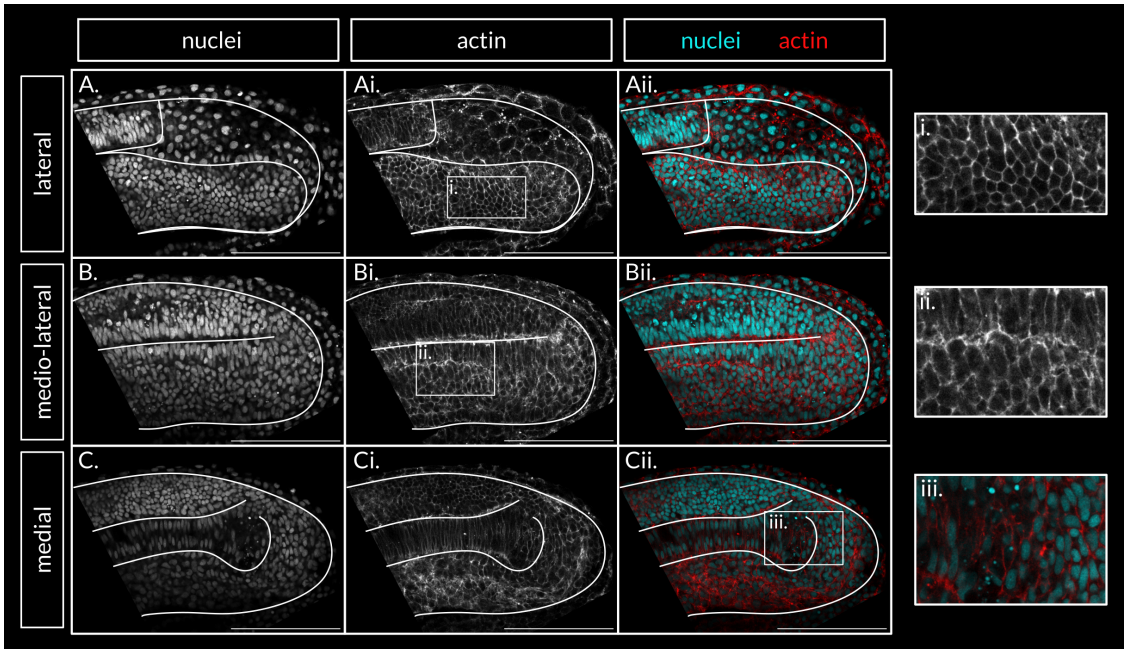


Figure 3.13: Anatomy of the PSM at late segmentation A-C: single optical sections taken at different lateral positions in the PSM. More lateral focal planes are at the top of the image; C depicts a medial slide through the notochord. Left panel: DAPI (nuclei) staining. Middle panel: phalloidin (actin) staining. Right panel: merged image, with nuclei in cyan and actin in red. i-iii: phalloidin stained insets of features of interest. White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues. Images A-D and i-iv are arranged anterior left, scale bar represents 100μm.

cells posterior to the notochord (Figure 3.13C, iii). This population of cells differ in morphology to other cells in the PSM, appearing to have more cytoplasm than other nearby PSM cells. The specification and differentiation of this cell population is described further in Section 3.2.8.

3.2.7 Cell density in the PSM

I then investigated cell density in the *A. calliptera* PSM. Cell density can be used to suggest a potential mechanism of elongation: in chick, a gradient of cell density in the PSM is associated with the incrased cell motility causing axial elongation and extension of the PSM (Bénazéraf et al., 2010). In zebrafish, total cell density increases over time and does not differ over space (Thomson et al., 2021). Therefore, I investigated cell density in the *A. calliptera* PSM over both

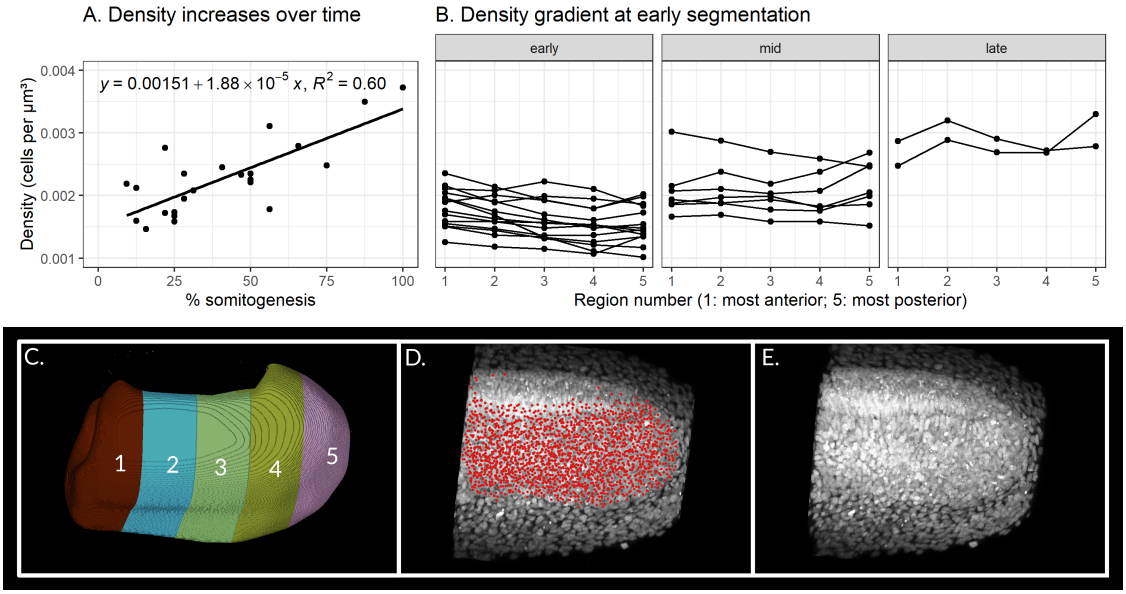


Figure 3.14: Cell density increases over time, and differs over space at early segmentation. A: Cell density in *A. calliptera* over developmental time. Cell density is measured as described in the methods. B: Cell density is higher in the anterior of the PSM, at early segmentation only. Dots represent measurements in different fifths of the PSM (divided anterior to posterior) for individual embryos; measurements for individual embryos are connected by lines. C-E: example of embryo with cells counted. C: subdivision of the PSM. Colours indicate different sections of the PSM, as labelled. D, E: example nuclei detection. Detected spots are in red.

space and time. To compute cell density, I counted the number of nuclei in the PSM using a method described in the Methods section, and divided by the total, manually segmented, volume of the PSM. As in zebrafish, the cell density in the *A. calliptera* PSM approximately doubles over time: at early segmentation, there are approximately 0.002 cells per μm^3; by the end of segmentation, this has risen to 0.004 cells per μm^3. This implies that like zebrafish, *A. calliptera* may be undergoing compaction-extension. I then computed the cell density across space in the PSM. I did this by subdividing the PSM into five evenly spaced regions across the anterior-posterior axis, to measure the cell density in each region. I found that when embryos at all stages are combined, the cell density does not significantly differ over space (p = 0.051, two-way ANOVA also accounting for percent somitogenesis). However, at early segmentation, the *A. calliptera* PSM lies flat on the yolk, more closely resembling the chick embryo, where a cell density

and motility gradient control elongation, Bénazéraf et al., 2010. Therefore I then analysed early segmentation embryos alone. At this stage, the cell density in the anterior-most fifth is 17% higher than that of the posterior-most (0.0018 vs 0.0016 cells per μm^3). This difference is statistically significant and the effect size is moderate (p = 0.0013, $\eta^2 = 0.14$, two-way ANOVA also accounting for percent somitogenesis). There is no statistically detectable difference over space at mid or late segmentation. This suggests that there is a cell density gradient, potentially contributing to axial elongation, at early but not mid or late segmentation stages.

3.2.8 Patterning of the *A. calliptera* midline

I then characterized the patterning and establishment of tissues at the embryonic midline. In zebrafish, a distinct population of "notochord progenitor cells" contribute towards the formation of notochord during segmentation (Talbot et al., 1995; Row et al., 2016; McLaren & Steventon, 2021). This cell population is marked by expression of the homeobox gene *noto* (Talbot et al., 1995; Melby et al., 1997). Therefore, I characterized *noto* expression in *A. calliptera*.

At early segmentation (7ss), *noto* is expressed in the posterior notochord (Figure 3.15A-C). At this stage, the notochord itself is only covered by a thin neurectodermal layer, and *noto* is also expressed dorsally in the tailbud (Figure 3.15A). Ventrally, *noto* is expressed in a population of cells that resemble neither the notochord cells, nor PSM cells (Figure 3.15C). Note that *noto* is only expressed in the posterior of the notochord, not throughout the notochord as in zebrafish (Talbot et al., 1995). Expression of *noto* at mid-segmentation is similar: the gene is expressed in the posterior notochord, and in a population of cells ventral to the notochord (Figure 3.15D, E). By late segmentation, *noto* is no longer expressed in the notochord; being instead expressed in a group of morphologically distinct cells immediately posterior to the notochord (Figure 3.15F). Expression of *noto* therefore migrates in *A. calliptera* from the posterior notochord, to a population of cells immediately behind the notochord.

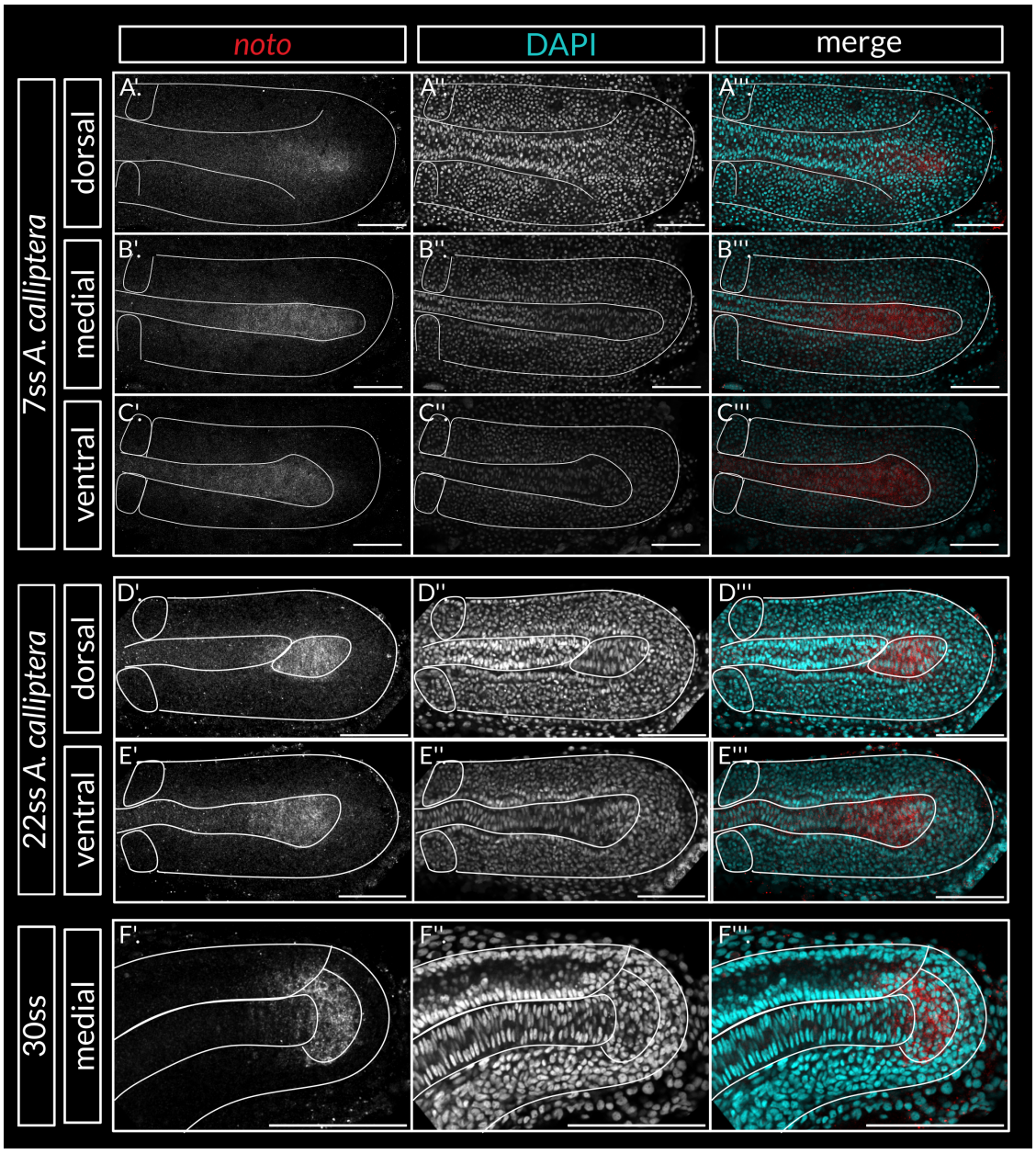


Figure 3.15: Expression of *noto* in *A. calliptera*. Each panel depicts single optical sections of embryos stained for *noto* and DAPI at different dorso-ventral positions in the embryo. Left column: *noto* expression in greyscale, Middle panel: DAPI signal in greyscale. Right panel: DAPI (cyan) and *noto* (red) expression merged. A-C: *Noto* expression at 7ss. D-E: lateral view of *noto* expression at 22ss. F: *noto* expression at 30ss. All embryos are presented anterior left. Scale bar: 100μm. Panels A-E imaged from dorsal aspect. Panel F imaged from lateral aspect. White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues, and somite boundary.

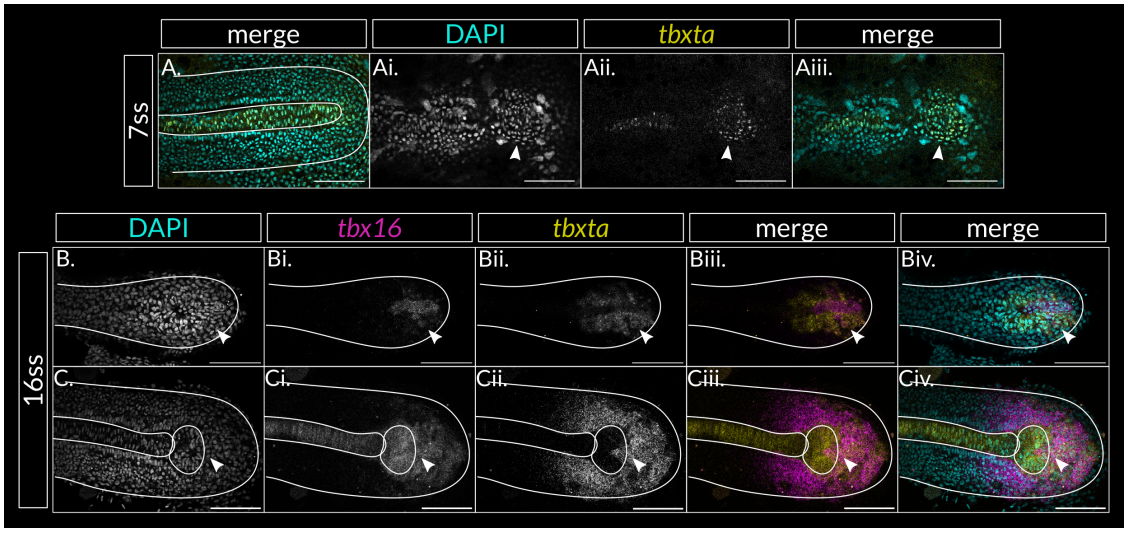


Figure 3.16: Gene expression within Kuppfer's vesicle. A.: *tbxta* expression (cyan) at 7ss. Left-most panel depicts DAPI (cyan) and *tbxta* expression through the notochord. Ai-Aii depict gene expression at the most ventral aspect of the embryo, through Kuppfer's vesicle (arrow). B, C: Expression of *tbxta* (yellow) and *tbx16* (magenta) at 16ss. B: Gene expression at the most ventral position, through Kuppfer's vesicle (arrow). C: Gene expression through the notochord. Kuppfer's vesicle is again indicated with an arrow. White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues. All embryos are oriented anterior left; scale bar 100μm.

In zebrafish, Kuppfer's vesicle is a fluid-filled vesicle of epithelial cells required for left-right patterning of the body axis (Essner et al., 2005; Ikeda et al., 2022). At 14-16ss (the time of tail eversion in zebrafish), this structure begins to invaginate into the PSM, where it will later contribute to mesodermal tissue (PSM and somites), as well as the notochord (Ikeda et al., 2022). In *A. calliptera*, the Kuppfer's vesicle is visible as a well-organized, epithelialized, structure at 16ss (Figure 3.10). This structure appears to express both *tbx16* and *tbxta* at 16ss; and expresses at least *tbxta* at 7ss (Figure 3.16). The significance of this gene expression is unclear, but may relate to the future differentiation of cells from Kuppfer's vesicle into the PSM and notochord.

3.2.9 Oct4 patterning of the *A. calliptera* neural tube

Patterning of the neural tube differs between *A. calliptera* and zebrafish. In zebrafish, *oct4* is expressed in the posterior neural tube, from the posterior boundary of the neural tube to about the end of the PSM (Yuikawa et al., 2021;

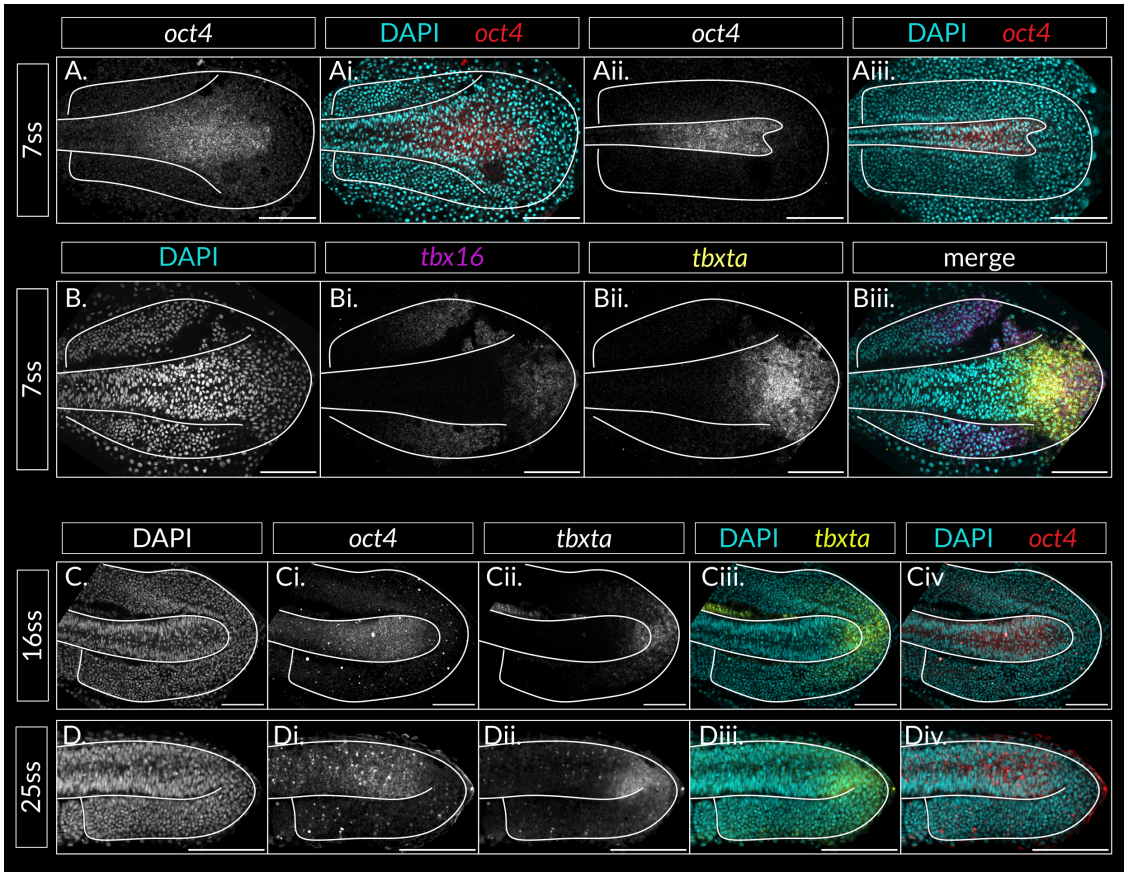


Figure 3.17: Patterning of the neural tube of *A. calliptera*. A: expression of *oct4* in two different focal planes of an early segmentation (7ss) embryo. B: expression of *tbxta*, but not *tbx16*, in the neural tube of an early segmentation (7ss) embryo. C: expression of *tbxta* and *oct4* in the neural tube of a mid segmentation (16ss) embryo. This embryo is a little squashed so the morphology is inaccurate. D: A late segmentation embryo (25ss) stained with *tbxta* and *oct4*. Each image depicts a single optical section of an embryo, imaged through the neural tube, oriented anterior left. Embryos A-C are imaged from the dorsal side and D is imaged laterally. White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues. Scale bar: 100 μm.

Toh et al., 2022). There is also a limited amount of *tbxta* expression within the neural tube (Yuikawa et al., 2021; Toh et al., 2022). In contrast, in *A. calliptera*, *tbxta* is expressed within the neural tube, and expression of *oct4* ceases slightly anteriorly to the posterior of the neural tube (Figure 3.17). This is the case at early, mid, and late segmentation (Figure 3.17). Note that unlike *tbxta*, *tbx16* is never expressed in the neural tube (Figure 3.17Bi). It is unclear why this occurs. If *oct4* is a reliable marker of neural fate in *A. calliptera* as it is in zebrafish, it would imply that cells retain a pluripotent/NMP identity further into the neural

tube in *A. calliptera* than in zebrafish.

3.2.10 Gene expression patterning of the PSM

I then investigated gene expression patterning of the PSM, characterizing the expression of genes involved in development, morphogenesis, and specification of the PSM. I first characterized expression of the tbox genes, a well-conserved set of genes required for PSM morphogenesis and development. *tbxta* has a well-conserved role in specifying the posterior body, and is expressed in the posterior PSM and notochord in zebrafish (Schulte-Merker et al., 1994; Martin & Kimelman, 2008, 2010; Bruce & Winklbauer, 2020). *tbx16* is required for proper cell movement and morphogenesis in the PSM, and in zebrafish is expressed in a posterior to anterior gene expression gradient (Ruvinsky et al., 1998; Griffin et al., 1998; Manning & Kimelman, 2015). Somite specification is achieved via *mesp-aa* class genes, and *tbx6* (Sawada et al., 2000; Nikaido et al., 2002; Windner et al., 2015). Here, I characterize expression of these genes in *A. calliptera*.

Retraction of tbox patterning over time

I characterized expression of the tbox genes *tbxta*, *tbx16*, and *tbx6* in *A. calliptera*. Expression of *tbxta* is very similar to zebrafish: *tbxta* is highly expressed in the notochord, and is expressed in the posterior of the embryo, in a gradient from posterior to anterior (Figure 3.18A′, B′, C). Notably, at early and mid segmentation, there is no *tbxta* expression indicative of notochord progenitor cells immediately posterior to the notochord (Figure 3.18A′, B, see Section 3.2.8). Expression of *tbx16* is also very similar to zebrafish: as in zebrafish, *A. calliptera* express *tbx16* in a gradient expression domain, with expression decreasing towards the anterior of the embryo (Figure 3.18A″, B″, C). Unlike zebrafish, however, the anterior boundary of *tbx16* retracts posteriorly over time: this is obvious both when comparing embryos at 7ss and 17ss (the anterior boundary of *tbx16* is at ∼75% AP position at 7ss and 50% at 17ss, after correcting for total PSM length, Figure 3.18A″, B″), and when examining quantifications of gene expression over

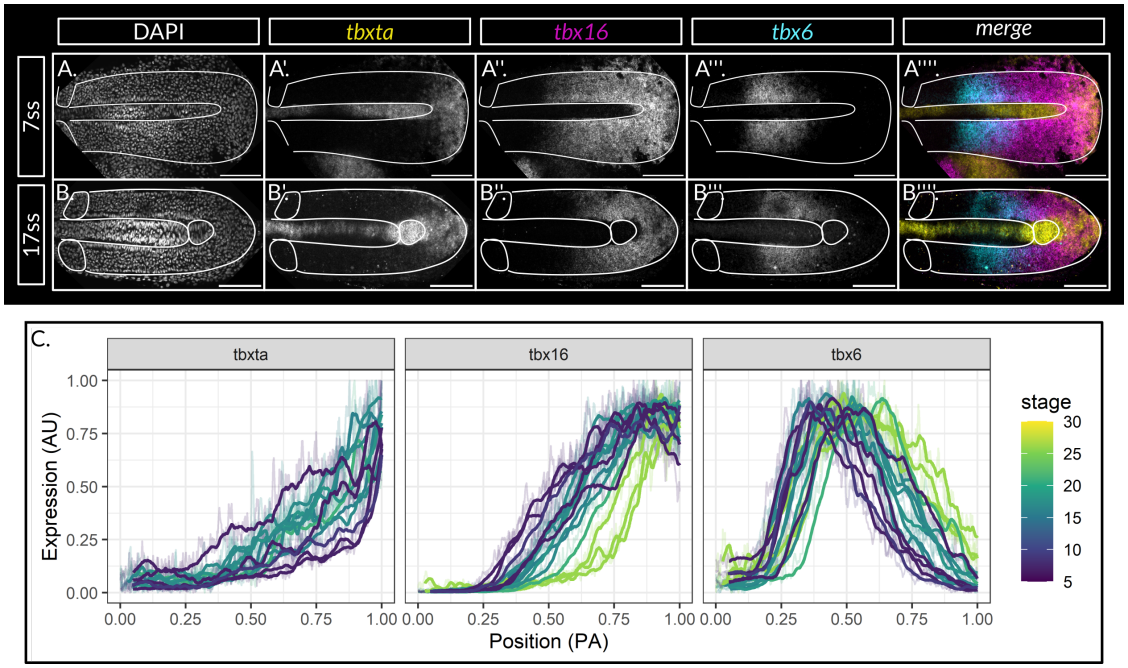


Figure 3.18: Expression of the tbox genes (*tbxta*, *tbx16*, and *tbx6*) retracts over time. A, B: Expression of tbox genes in early (7ss, A) and mid (17ss, B) stage embryos. Panels depict averages across 10 optical sections/12μm in the middle of the embryo. Left from right, panels show *DAPI*, *tbxta*, *tbx16*, *tbx6*, and merged gene expression. Merged panel shows *tbx6* in cyan, *tbx16* in magenta, and *tbxta* in yellow. Embryos are imaged from the dorsal aspect and presented anterior left. White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues. Scale bar: 100μm. C: Quantifications of gene expression across the anterior-posterior axis. Quantification was performed as in the Methods. Gene expression is normalized 0 to 1 in the anterior-posterior dimension (0 representing the posterior tip of the nascent somite, and 1 representing the posterior boundary of the notochord), and from 0 to 1AU.

space (Figure 3.18C). The expression of *tbx6* likewise retracts over time. At 7ss, the posterior-most extent of *tbx6* expression is at ~60% AP position, well anterior to the posterior of the notochord, while at 17ss *tbx6* expression extends to the notochord boundary (Figure 3.18A''', B''', C). The overall retraction of tbox expression in *A. calliptera* represents a significant difference to zebrafish, where expression of tbox expression scales with the PSM over time (Fulton, 2021; Spiess et al., 2024).

Change in relative expression of *tbx6* and *mesp-aa*

I also characterized the expression of *mesp-aa*, which marks the somite determination boundary in zebrafish. In *A. calliptera*, *mespaa* is expressed appoximately

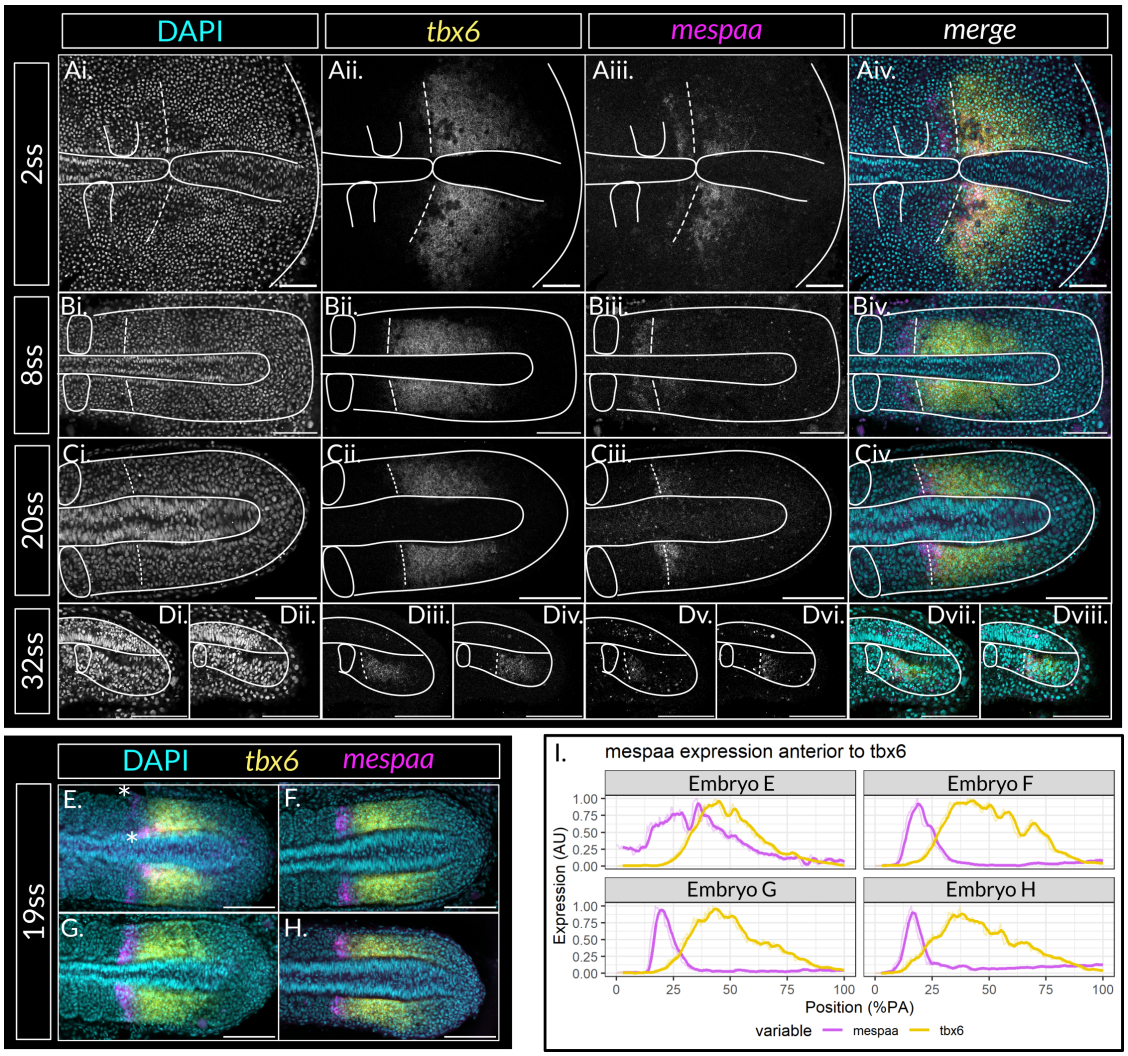


Figure 3.19: *A. calliptera mespaa* is consistently anterior to *tbx6*, in contrast to zebrafish. A-C: single confocal sections of embryos stained for *mespaa* and *tbx6*. Left panel (i): DAPI expression. Panel ii: *tbx6* expression. Panel iii: *mespaa* expression. Panel iv: merged gene expression, with DAPI in cyan, *tbx6* in yellow, *tbx16* in magenta. Embryos are imaged from the dorsal aspect and presented anterior left. Scale bar: 100μm. D: embryo is laid out as for A-C, except that it is imaged from the lateral aspect (presented anterior left), and two different focal planes are shown. White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues, and the somite. Dashed lines indicate the anterior boundary of *tbx6* expression. E-H: Average intensity projections of embryos from the same clutch, stained and presented as above. I: Quantifications of gene expression in E-H.

one somite length from the nascent somite, anterior to *tbx6* expression (Figure 3.19A-D). This pattern is consistent across early (8ss), mid (20ss), and late (32ss) segmentation (Figure 3.19B-D). However, at the onset of segmentation (2ss), *mespaa* expression is significantly more disorganized: *mespaa* is expressed in two domains, one thin stripe anterior to the *tbx6* anterior boundary, and a thicker expression domain within the *tbx6* stripe (Figure 3.19A). This expression was present in all three embryos analysed at this stage. Note also that the circular population of cells expressing *oct4* but not *tbx6* are present at this stage; *mespaa* is not expressed within this cell population (Figure 3.19A). This differing expression pattern could be due to an increased rate of somitogenesis resulting in greater numbers of nascent somites forming within the embryo, or early patterning of the PSM still being established at this stage.

Expression of *mespaa* is dynamic within the embryo. In most embryos imaged, *mespaa* is expressed in one stripe directly anterior to *tbx6* (Figure 3.19F-H). However, rarely, two *mespaa* stripes are present: one, fading, anterior to the *tbx6* boundary and the other within the *tbx6* boundary (Figure 3.19E, I, asterisks). Presumably, in these embryos, the fading, anterior, *mespaa* stripe represents the somite that has already been specified and is undergoing the epithelial-to-mesenchymal transition and somite specification, while the posterior-most *mespaa* stripe marks a newly specified somite.

The relative expression of *tbx6* and *mespaa* differs between zebrafish and *A. calliptera*. In zebrafish, multiple *mespaa* stripes are expressed within the *tbx6* expression domain, not anterior to it (Nikaido et al., 2002) [3]. In *A. calliptera*, however, *mespaa* is consistently expressed just anterior to the *tbx6* stripe, and except for newly-formed *mespaa* stripes (eg. Figure 3.19E), never within the *tbx6* stripe (Figure 3.19). Additionally, except at very early stages, *A. calliptera* only ever have one *mespaa* stripe, perhaps due to the slower place of somitogenesis clock relative to axial elongation in this species (Figure 3.19).

[3] In the paper cited, *tbx6* is called *tbx24* and *tbx16l* is called *tbx6*.

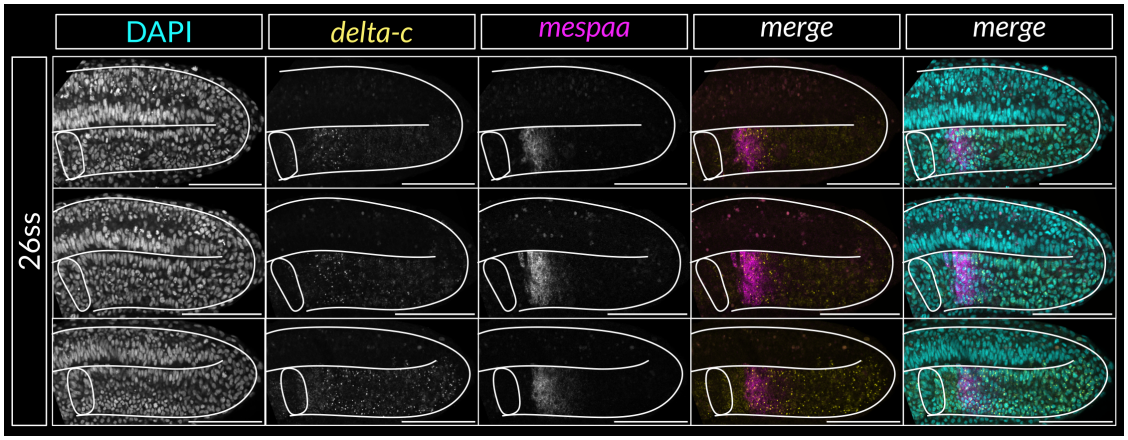


Figure 3.20: *mespaa* marks somite specification. Each row depicts single confocal slices of embryos from the same clutch at different phases of the segmentation clock. Left to right, panels depict DAPI expression, followed by *delta-c*, *mespaa*, then *delta-c* (yellow) alongside *mespaa*, followed by DAPI (cyan), *mespaa* (magenta), and *delta-c* (yellow). Embryos are imaged from the dorsal aspect and presented anterior left. White lines indicate the posterior boundary of the embryo, and boundaries of the notochord and neural tissues, and the somite. Scale bar: 100μm.

Given the difference in *mespaa* expression between zebrafish and *A. calliptera*, I verified that *mespaa* is indeed likely to mark the nascent somite. In zebrafish, *delta-c* expression varies based on the phase of the segmentation clock: nascent transcripts are detectable as bright punta of expression, and levels of cytoplasmic transcript vary (Mara et al., 2007). Expression of *delta-c* is similar in *A. calliptera*; some embryos exhibit a broad domain of punctate *delta-c* expression, while in others the expression is cytoplasmic and lower (Figure 3.20). Crucially, in all embryos examined, *delta-c* expression switches from punctate to cytoplasmic precisely at the anterior boundary of *mespaa* expression. This implies that *mespaa* does indeed mark formation of the nascent somite in *A. calliptera*.

3.3 Discussion

In this chapter, I performed a detailed characterization of PSM morphology during axial elongation and segmentation in *A. calliptera*. In general, *A. calliptera* closely resemble zebrafish with regards to both morphology of the PSM during development, and patterning of the PSM. However, unlike zebrafish, *A. calliptera*

have a density gradient of cells within the PSM, implying that the mechanism of segmentation could potentially differ between these species at early segmentation. In addition, patterning of the neural tube does slightly differ between species, with a retraction of tbox expression domains, a change in the relative expression of *mespaa* and *tbx6*, and a difference in patterning of the neural tube.

3.3.1 The morphology of the *A. calliptera* PSM changes over time

The morphology of the *A. calliptera* PSM changes over time. I divide different stages of segmentation into early, mid, and late stages, as distinguished by different arrangements of the PSM and the extraembryonic tissues. At early segmentation, the paraxial mesoderm is continuous with the lateral plate mesoderm: morphologically, there is no difference between cells in the paraxial mesoderm and lateral plate mesoderm, and mesodermal markers *tbx16* and *tbx6* are expressed into the lateral plate mesoderm (Figure 3.3). At 15ss, a layer of elongate cells begin to surround the PSM, which I call the psuedo-epithelial layer (Figures 3.9, 3.10). At this stage, the PSM is lying flat onto the yolk, is morphologically distinct from the extraembryonic tissues and is separated from extraembryonic tissues by a membrane. I call this stage "mid-segmentation". Finally, at 25ss, the *A. calliptera* embryo undergoes tail eversion, where the tailbud lifts from yolk and the relative arrangement of tissues in the tailbud change.

It appears that similar stages exist in other species. In zebrafish, the earliest somites are produced while epiboly is ongoing and the blastopore has not yet closed (Kanki & Ho, 1997; Steventon et al., 2016). In addition, prior to 12ss, lateral ingression of cells from the lateral plate mesoderm and dorsal tissues into the PSM and somites is possible (Lawton et al., 2013; Steventon et al., 2016). Expression of the mesodermal marker *tbx16* is confined to the PSM by 6ss, potentially earlier, implying that the overall shape of the PSM has been specified by this time (Osborn et al., 2020). In zebrafish, tail eversion also occurs at 15-16ss (or 50% somitogenesis); after this time, the morphology of the tailbud resembles that of *A. calliptera* after tail eversion (Kanki & Ho, 1997; Steventon et al., 2016). Therefore,

zebrafish appear to go through phases of development resembling the so-called early, mid, and late segmentation stages in *A. calliptera*.

Chick also appear to undergo different phases of segmentation. In this species, somitogenesis occurs from stages HH8 - HH25 (Hamburger & Hamilton, 1951; Tenin et al., 2010). The tailbud appears continuous with the lateral plate mesoderm until HH15 (24-27ss / $\sim 50\%$ somitogenesis). By HH17 (30ss/57% somitogenesis) there is a clear separation between the PSM and lateral plate mesoderm. Tail eversion occurs by HH20 (40-43ss/$\sim 75\%$ somitogenesis) (Hamburger & Hamilton, 1951; Tenin et al., 2010). This is the same as *A. calliptera*, where mid-segmentation begins at 15ss / 47% somitogenesis and late segmentation begins at 25ss/75% somitogenesis. This means that there is a heterochronic shift in the timing of tail eversion between species: *A. calliptera* undergo tail eversion at 78% somite production, zebrafish at 50% somite production, and chick at 75% somite production. This heterochrony may be important in the evolvability of axial growth.

3.3.2 *A. calliptera* differs to zebrafish in patterning

There is also evidence that the GRNs patterning the cichlid PSM have evolved between *A. calliptera* and zebrafish. Gene expression patterning within the PSM differs between species. That is, expression of the tbox genes *tbx16* and *tbx6* retracts across the PSM over time in *A. calliptera*. This retraction does not occur in zebrafish, where the *tbx6* domain peaks at $\sim 50\%$ PSM length, and *tbx16* expression peaks at $\sim 80\%$ PSM length. Similarly, the relative expression of *mesp-aa* and *tbx6* has changed between *A. calliptera* and zebrafish: in zebrafish, multiple *mesp-aa* stripes are expressed in the PSM, well within the *tbx6* mRNA expression domain (Nikaido et al., 2002; Cutty et al., 2012; Windner et al., 2015). In contrast, *A. calliptera* tend to have one *mespaa* stripe, which is expressed at, or close to, the anterior boundary of *tbx6*. This could represent a change in the regulation between *tbx6* and *mespaa* in *A. calliptera*, or it could represent a change in *tbx6* autoregulation: in zebrafish, the Tbx6 protein domain is expressed anterior to

tbx6 mRNA, and autoregulation of Tbx6 is required for correct segmentation (Windner et al., 2015; Ban et al., 2019). It is possible that the extent of Tbx6 autoregulation has changed in cichlids, instead of *mesp-aa* regulation. Given the high degree of conservation in the *mesp-aa* enhancer, it is more likely that Tbx6 regulation autoregulation is has evolved in *A. calliptera*, but this question warrants further research (Terasaki et al., 2006; Yasuhiko et al., 2008).

Patterning of the neural tube also differs between *A. calliptera* and zebrafish. In zebrafish, *tbxta* is expressed slightly into the neural tube, and *oct4* is expressed throughout the entire neural tube. In contrast, in *A. calliptera*, *tbxta* is expressed well into the neural tube, and *oct4* expression does not extend to the very posterior of the neural tube. This could occur because specification of neural cells does not complete until after they have entered the neural tube, or it could be due to a delay in *tbxta* decay.

3.3.3 *A. calliptera* elongate via compaction-extension

Quantifying cell density in the PSM provided some insight into potential mechanisms of elongation in *A. calliptera*. I found that the cell density in the PSM approximately doubles during *A. calliptera* axial elongation. This is consistent with compaction-extension of the PSM during axial elongation: in the zebrafish PSM, cells do not undergo directed rearrangement and intercalation, but instead decrease in size (increase in density) and converge to the midline (Thomson et al., 2021).

I therefore investigated cell density across the *A. calliptera* PSM, over time. Chick embryos are thought to elongate via a cell motility gradient: cells in the posterior are more motile and therefore the body axis extends posteriorly (Bénazéraf et al., 2010). This mechanism is also associated with a cell density gradient; cells in the posterior are less dense (and therefore more motile). I found that in *A. calliptera*, there is a slight cell density gradient at early segmentation: the anterior-most fifth of the PSM is, on average, 17% more dense than the posterior-most fifth. However there is no difference in cell density across the PSM at mid or

late stages, or indeed when all stages are analysed together. This suggests that at early segmentation, *A. calliptera* may elongate via a cell density/motility gradient, and moreover that the mechanisms of elongation used by *A. calliptera* may differ over time. However, without live imaging and tracking of cells in the cichlid PSM at different developmental stages, it is impossible to confirm this idea.

3.4 Conclusions

In this chapter, I characterize *A. calliptera* PSM development during axial elongation. I find that, in general, *A. calliptera* closely resemble zebrafish in anatomy and gene expression patterning in the PSM, with a number of differences. Firstly, there is a heterochonic shift in the timing of tail eversion between zebrafish and *A. calliptera*. Secondly, tbox expression domains retract in *A. calliptera*, and the relative expression of *mesp-aa* and *tbx6* differ, suggesting there may have been evoution in the GRN controlling axial elongation. Finally,there is preliminary evidence that *A. calliptera* may elongate via a diffusion gradient at early segmentation but not at later stages, representing a change in the mechanism of morphogenesis over time. This chapter also lays the groundwork for the detailed comparison of *A. calliptera* and *R. chillingali* axial elongation performed in Chapter 4, and will be a resource for future evo-devo research in teleost fish.

Chapter 4: Evolving somite counts without evolving the segmentation clock

Contents

4.1 Introduction **73**
4.2 Results **77**
4.2.1 PSM size differs at the onset of somitogenesis between *R. chillingali* and *A. calliptera* 77
4.2.2 Structural anatomy of the tailbud is similar between species . . 81
4.2.3 Molecular anatomy of the tailbud is similar between species . . 84
4.2.4 Anatomy of the body axis and the PSM is indistinguishable between species at the onset of segmentation 86
4.3 Discussion **87**
4.3.1 Cichlids may have evolved somite counts by modifying the initial conditions of somitogenesis 87
4.3.2 A unique mechanism of epiboly in African cichlids may contribute to morphological evolution 89
4.3.3 Evolution via boundary conditions of embryogenesis 91
4.4 Conclusions **92**

4.1 Introduction

How phenotypic diversity is generated and maintained is a fundamental question in developmental biology. One example of phenotypic diversity is vertebrae number. This varies hugely between species: for example snakes can form hundreds of vertebrae, while fish form dozens (Gomez et al., 2008). Vertebral precursors are called somites, and are laid down during a process called somitogenesis, where new somites are produced from a posterior structure called the presomitic mesoderm (PSM) (Oates et al., 2012). Somitogenesis has been extensively studied in traditional vertebrate model species (eg. Gomez et al., 2008; Bénazéraf et al., 2010, 2017; Steventon et al., 2016; Lawton et al., 2013). However, the evolvability of somitogenesis remains poorly understood.

The process of somite production is well-explained by the Clock and Wavefront model. According to this model, the segmentation clock is comprised of a set of autonomously oscillating genes (such as *delta-c*, *her11*, and *lunatic fringe*); one somite is produced for each oscillation of the segmentation clock (Mara et al., 2007; Krol et al., 2011; Oates et al., 2012). Segmentation clock oscillations are controlled by a wavefront of FGF signalling, which controls where in the PSM oscillations cease, and where the nascent somite is positioned (Oates et al., 2012). There are therefore a number of different ways somite counts can evolve (Oates et al., 2012). Firstly, the rate of the somite production could change. As the length of the nascent somite is controlled by the rate of the segmentation clock, and the wavefront, this would result in more, smaller somites, being formed from the same total tissue volume (Oates et al., 2012). This mechanism likely explains the increased number of vertebrae in snakes (Gomez et al., 2008). Another way to evolve the number of somites would be to increase the duration of somitogenesis: a species could segment for longer, and therefore produce more somites (Oates et al., 2012). Alternatively, more somites could be formed simultaneously at the onset of segmentation (Oates et al., 2012). Note that despite previous reports, neither chick nor zebrafish form segments simultaneously

at the onset of segmentation, though the rate of somite production in chick is slower at the onset of segmentation (Kimmel et al., 1995; Schröter et al., 2008; Maia-Fernandes et al., 2024).

In vertebrates, the dynamics of volumetric growth of the body axis, and somite production, vary. This could contribute to the evolvability of somite counts by changing the total amount of tissue available to produce somites. Some species, including chick, mouse, and some snakes undergo extensive total growth of both the PSM and the whole body axis (Gomez et al., 2008; Bénazéraf et al., 2017; Steventon et al., 2016). Zebrafish, on the other hand, never exhibit overall growth the PSM or of the somitic mesoderm; indeed, the PSM and nascent somite increase in density over time, leading to overall compression of the body axis alongside elongation (Steventon et al., 2016; Thomson et al., 2021). Modifying the extent of tissue growth during somitogenesis and axial elongation could increase somite number without a corresponding change in somite size.

While the mechanisms of somite production and axial elongation are relatively well-understood in zebrafish and chick, the large evolutionary distance and different embryological strategies make it difficult to compare species directly, and therefore understand how somite numbers can evolve. Therefore, the ideal system to study the evolvability of somite counts would be a species group where species are similar developmentally and genetically, but vary in somite number. This would allow detailed and quantitative comparison of the process of somite production, and ultimately enable the developmental basis of somite number evolvability to be explained.

Lake Malawi cichlids provide such a system with which to study the evolution of somite and vertebrae number. This species group has undergone an explosive radiation in the last 800 000 years (Turner, 2007; Malinsky et al., 2018; Santos et al., 2023). In this time, 800 different cichlid species have evolved drastically different phenotypes (varying in craniofacial morphology, body colouration and patterning, and body shape), without extensive genetic divergence: the average nucleotide divergence between species is only 0.1-0.25% (Malinsky et al.,

2018). Remarkably, there can be more genetic variation within a single individual than between species, meaning that genetic divergence does not explain the phenotypic diversity in this species flock (Malinsky et al., 2018). Of relevance to our study, Lake Malawi cichlids differ in vertebrae number, varying from 28 to 42 vertebrae (Bucklow et al., 2024). This extensive phenotypic divergence, accompanied the by extremely limited nucleotide differences between species, makes Lake Malawi cichlids an excellent system to study the evolvability of vertebrae and somite number, as well as many other phenotypic traits (Turner, 2007; Malinsky et al., 2018; Santos et al., 2023).

In this work, I investigate the origins of somite number variation in cichlids, by characterizing somitogenesis in two species: *A. calliptera* and *R. chillingali*, which form a mean of 32 and 38 somites respectively (Marconi et al., 2023; see also Figure A.1). I chose to focus on these two species because their development has already been characterized (Marconi et al., 2023), because they vary 25% in somite counts, and because they represent the ancestral (*A. calliptera*) and one of the most phenotypically extreme (*R. chilingali*) fish in the lake (Malinsky et al., 2018). I aim to understand the developmental basis of this difference in somite count. An overview of *A. calliptera* and *R. chillingali* development is in Figure 4.1.

A change in the rate of somitogenesis does not appear to explain the increased somite count in *R. chillingali* compared to *A. calliptera*. Marconi et al., 2023 have characterized the rate of somite production between species, and found that *R. chillingali* segment 8% faster than *A. calliptera*. This increased rate would account for two of the six somite difference between species. This study characterized the number of segments by measuring the number of hours after the onset of segmentation, which could introduce error in the measurements. James Hammond has since performed a more detailed analysis, measuring the number of somites formed at different hours post fertilization (hpf). This revealed that the timing of the segmentation onset differs between species, with *R. chillingali* beginning and ending somitogenesis later than *A. calliptera*. *A. calliptera* begin somitogenesis at 31hpf and end at 67hpf, while *R. chillingali* begin segmentation

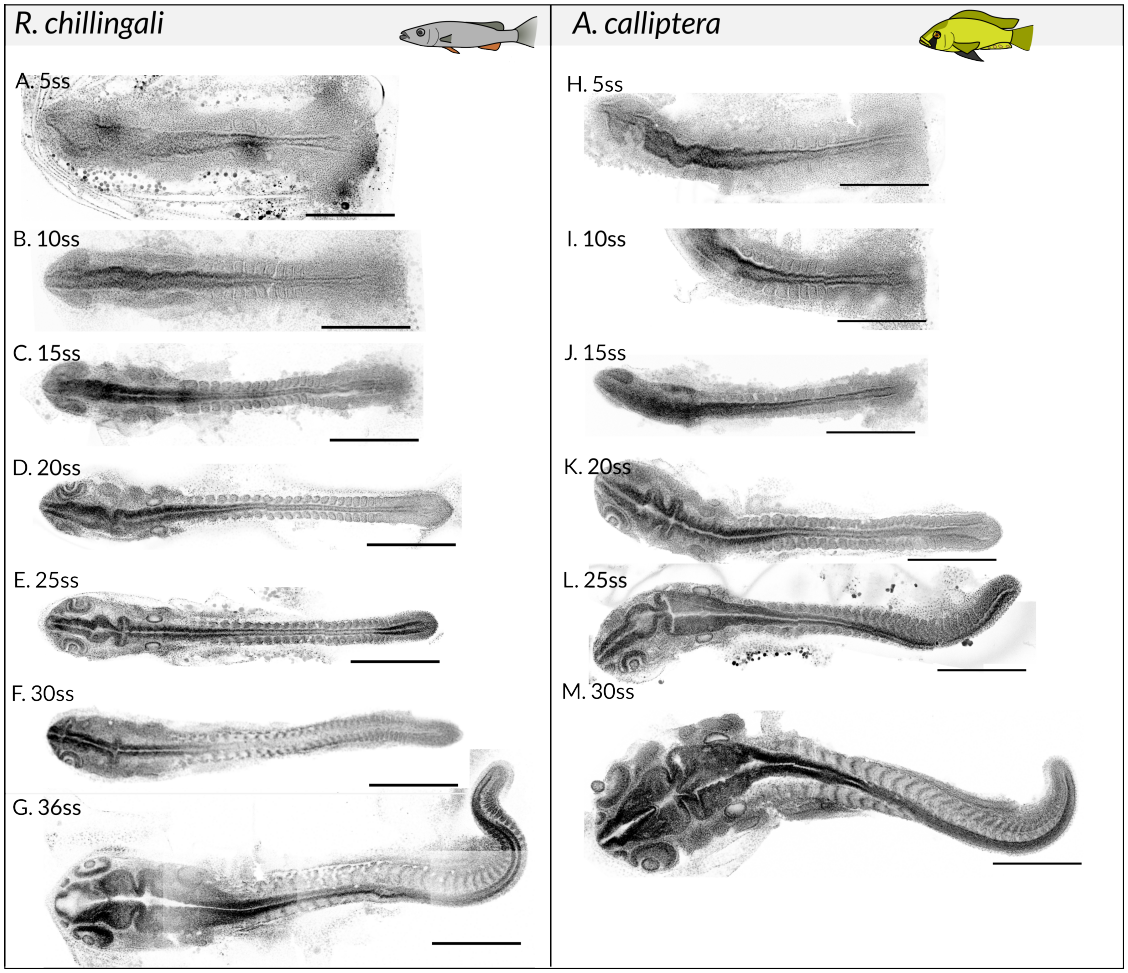


Figure 4.1: Overview of cichlid somitogenesis. A-M: single optical section of flat-mounted of DAPI-stained (black) cichlid embryos at different stages. All embryos are a dorsal view, oriented anterior left, to the same scale. Scale bar: $500\mu m$.

at 36hpf and complete it at 82hpf (James Hammond, unpublished data). The duration of segmentation is therefore longer in *R. chillingali*, taking 46 hours in *R. chillingali* compared to 36 hours in *A. calliptera*. This dataset also implies that *A. calliptera* segment faster than *R. chillingali*, forming one somite every 66 minutes compared to 73 minutes in *R. chillingali*. Therefore, *R. chillingali* appear to evolve increased somite counts by evolving the duration, not the rate, of somitogenesis (James Hammond, unpublished data).

As an increase in the rate of somite production does not fully explain the difference in somite number between *R. chillingali* and *A. calliptera*. There must be a difference in the dynamics of growth between species. Here, I characterize these

differences. I characterize the morphodynamics of the PSM over time, to uncover where the excess tissue required for *R. chillingali* to produce more somites arises from. I then perform a detailed morphometric characterization of anatomy of the PSM, and characterize patterning of the PSM, to identify whether there is any evidence for a change to the developmental process of segmentation between species. Finally, I characterize the onset of segmentation in detail.

4.2 Results

4.2.1 PSM size differs at the onset of somitogenesis between *R. chillingali* and *A. calliptera*

An increased rate of somite production does not explain the increased somite count of *R. chillingali*. Instead, *R. chillingali* produce segments for longer during embryogenesis (James Hammond, unpublished data). Here, I attempt to understand where the excess tissue producing the increased number of somites arises from, developmentally speaking. It could be that there is no excess tissue; instead the nascent somites are smaller in *R. chillingali* than *A. calliptera*, meaning that more somites can be produced from the same tissue volume. Alternatively, the PSM could be larger in *R. chillingali* at the early stages of segmentation, meaning that more somites can be produced with tissue present early in segmentation. Finally, the *R. chillingali* PSM could be able to grow more, or replenish lost tissue faster, than *A. callitera*. Here, I attempt to establish which mechanism is operating by quantifying the size of the nascent somite, and the PSM/posterior unsegmented mesoderm, in both species and over developmental time.

I found that the dynamics of growth change over time, particularly before and after 15ss, which is when the PSM separates from the lateral plate mesoderm. Prior to 15ss, there is no statistical evidence that the PSM is decreasing in length ($p = 0.07$, two-way ANOVA also accounting for species). After 15ss, the PSM is decreasing in length at a rate of $16.3 \pm 0.7 \mu m$ for each somite produced ($p < 0.001$, η^2=0.74, two-way ANOVA also accounting for species). I see a similar result for

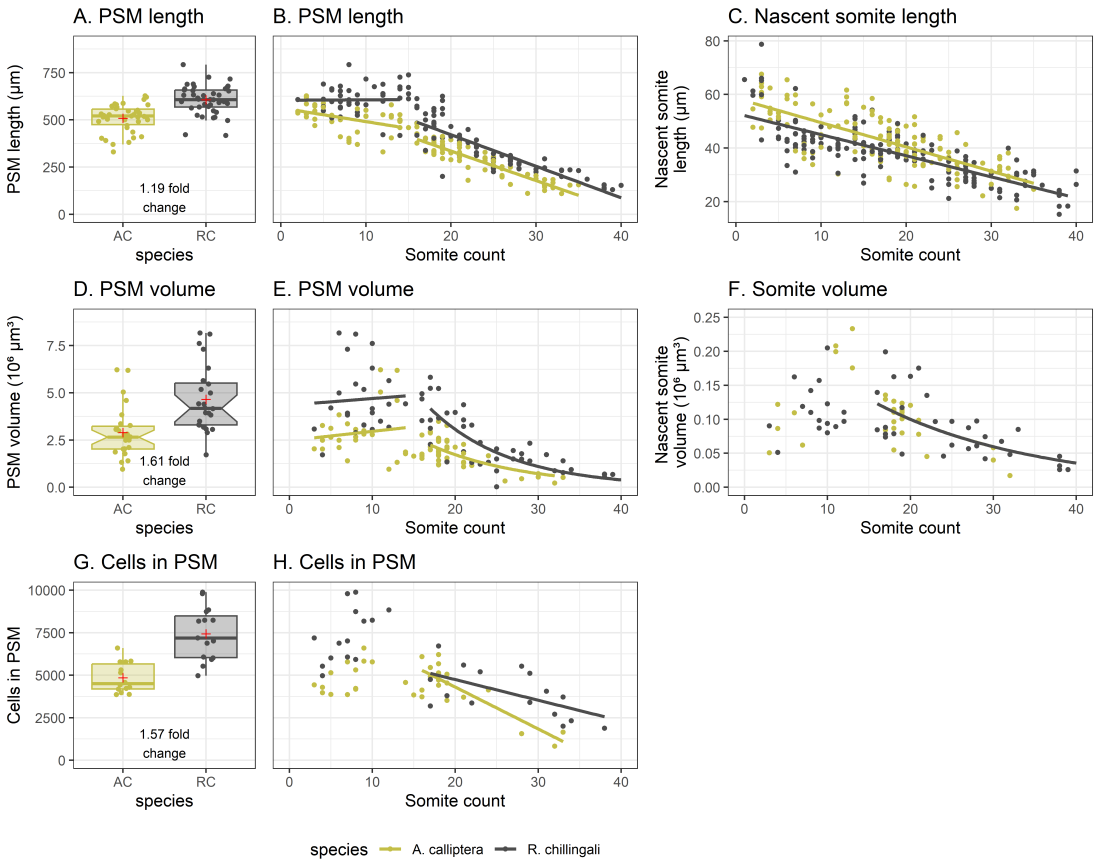


Figure 4.2: PSM size differs only at the onset of segmentation. A, D, G: quantification of PSM size for both species prior to 15ss. Boxplots depict the upper and lower quartiles of the data, with the solid line depicting the median of the data. Red cross depicts the mean of the data. Whiskers represent the largest value no further than 1.5 times the interquartile range of the plot. A, B: PSM length over developmental time. PSM length is measured from scans of whole embryos taken under lower magnification and is defined as the distance from the posterior tip of the PSM to the posterior aspect of the nascent somite; measured along the embryo midline. The different lines of best fit are computed using ordinary least squares for data subsetted to before, and after, 15ss. C: Nascent somite length over time. Somite length, like PSM length, is measured on images taken under lower magnification. The line of best fit is computed using ordinary least squares, aggregating data before and after 15ss as there was no evidence for a change in the line of best fit at this timepoint. D, E: PSM volume over time. PSM volume is measured manually as described in the Methods section. The trend line was computed to an exponential decay function using nonlinear least squares. F: Somite volume, measured and trendline computed as for E. G, H: Number of cells the PSM over time. Cells were counted using a custom automatic script defined in the Methods.

PSM volume: after 15ss, the data are well-represented by an exponential decay curve (Figure 4.2C). However, prior to 15ss there is no evidence for an association between PSM volume and somite count; that is, the PSM is not changing in volume over time ($p = 0.54$; two way ANOVA testing the interaction between PSM volume and somite count alongside species). After 15ss, the number of cells in the PSM is well approximated by a linear regression linking cell number to somite count, but I do not have enough data to compare the dynamics before and after 15ss. Altogether, this shows that the dynamics of PSM growth change over time: prior to 15ss, the PSM remains constant in length and volume, but after 15ss, it decreases in length, volume, and cell number (Figure 4.2). Note that prior to 15ss, tissue is being lost to the nascent somite but the PSM is not changing overall in length and volume. This means the body axis as a whole is undergoing volumetric growth to replenish tissue lost to the somites.

I then compared the total size of the PSM and nascent somite between species. The length of the nascent somite differs between species: *A. calliptera* have longer somites than *R. chillingali* at the same somite stages (Figure 4.2C). The intercept of the line of best fit is larger in *A. calliptera* than *R. chillingali* ($p < 0.001$, $\eta^2=0.07$, two-way ANOVA accounting for species and somite stage), but the slope does not differ statistically ($p = 0.099$; *R. chillingali* nascent somite is reducing in size at a rate of $0.77\mu m$ per somite formed; *A. calliptera* at a rate of $0.9\mu m$). This means that this difference in somite length is established at the onset of segmentation, and that there is no statistically detectable difference in the rate of change of somite length between species.

R. chillingali have a significantly longer PSM than *A. calliptera* at all stages (Figure 4.2). At early segmentation (prior to 15ss), the *R. chillingali* PSM is on average 19% larger than that of *A. calliptera* (Figure 4.2A). The rate of change of the PSM does not differ measurably between species: while the *A. calliptera* PSM is decreasing in length at a rate of $7.2\mu m$ per somite produced, and the *R. chillingali* PSM is increasing in length at a rate of $0.2\mu m$ per somite produced, this difference in rate is not statistically significant ($p = 0.12$). Similarly, after 15ss, the

R. chillingali PSM reduces in length at a rate of 16.7μm per somite produced, and *A. calliptera* at 15.6μm/somite, a statistically indistinguishable difference (p=0.47). Thus, *R. chillingali* have a longer PSM than *A. calliptera* throughout somitogenesis and there is no statistically detectable difference in the rate of shrinkage of the PSM, implying that this difference is established prior to segmentation.

R. chillingali also have a PSM which is larger in volume than *A. calliptera*. Prior to 15ss, the *R. chillingali* PSM is on average 61% larger in volume than that of *A. calliptera* (Figure 4.2E; $p < 0.001$, $\eta^2 = 0.25$, two way ANOVA as above). After 15ss, the exponential decay trendlines describing both species' development show that the *R. chillingali* PSM is always larger than *A. calliptera*. Unlike the PSM volume, the nascent somite does not differ detectably in volume between species ($p = 0.82$, two way ANOVA also accounting for somite stage), though it does decrease over time (Figure 4.2F; $p < 0.001$; η^2=0.32).

Finally, I quantified the number of cells in the PSM. As with PSM length and volume, species differ in the number of cells in the PSM at early segmentation: *R. chillingali* have 57% more cells than *A. calliptera* prior to 15ss ($p <0.0001$, $\eta^2 = 0.63$, Figure 4.2J). The number of cells in the PSM is also larger in *R. chillingali* after 15ss ($p = 0.047$; $\eta^2 = 0.12$, Figure 4.2J).

Altogether, these data show that *R. chillingali* have a larger PSM than *A. calliptera*, with regards to PSM length, PSM volume, and the number of cells in the PSM. This difference is particularly obvious at early segmentation, where the *R. chillingali* PSM is 19, 61, and 57% larger than that of *A. calliptera*, with regards to length, volume, and cell number respectively. There is limited difference in somite size between species: while the nascent *R. chillingali* somites are slightly shorter than those of *A. calliptera*, the volume does not differ statistically significantly between species. Moreover, I was unable to detect any differences in the rate of change in PSM size (length and volume) over time. The tissue required to produce more somites is present at the onset of segmentation in *R. chillingali*, implying that this species may have evolved increased somite numbers by evolving the initial conditions of somitogenesis.

4.2.2 Structural anatomy of the tailbud is similar between species

Above, I showed that at early stages in segmentation, the *R. chillingali* PSM is larger than *A. calliptera* with regards to length, volume, and cell number. This implies that *R. chillingali* may have evolved an increased somite count by evolving the initial conditions of somitogenesis. However, I wished to know whether there is any evidence that the process of segmentation has changed between species, presenting an alternate explanation for the evolution of somite counts. Therefore, I then characterized the morphology of the cichlid unsegmented region in more detail, measuring the PSM length, width, depth, notocord length, length of the putative mesodermal progenitor region, nascent somite size, and the ratios between these values (11 metrics in total, Figure 4.3F). I found that within each species, all traits are significantly associated with somite stage (Table A.1). After correcting for multiple testing and accounting for somite stage, only PSM length, notocord length, and MPZ length differ between species (Figures 4.3A-D, F, A.4, Tables A.1, A.2). If these differences represented some change in the segmentation process, rather than a difference in initial conditions, I would expect the slope of the line of best fit to differ between species. This is not the case for any of the statistical interactions between variable and species, except for PSM width and PSM width:length (Figure A.4, Table A.2); only the intercept of the regression differs between species. Therefore, it appears that these differences, like the PSM length, volume, and number of cells, are established prior to the onset of segmentation.

The ratio of PSM length to somite length differs between species and over time (Figure 4.3D). At early stages of segmentation, the PSM is 10-15 times larger than the nascent somite; by the end of segmentation it is only around seven times larger than the nascent somite (Figure 4.3D). As the period of somitogenesis does not change over time, and the length of the nascent somite is controlled by the position of the wavefront of the segmentation clock within the embryo, this implies that the position of the wavefront is changing over time in both species.

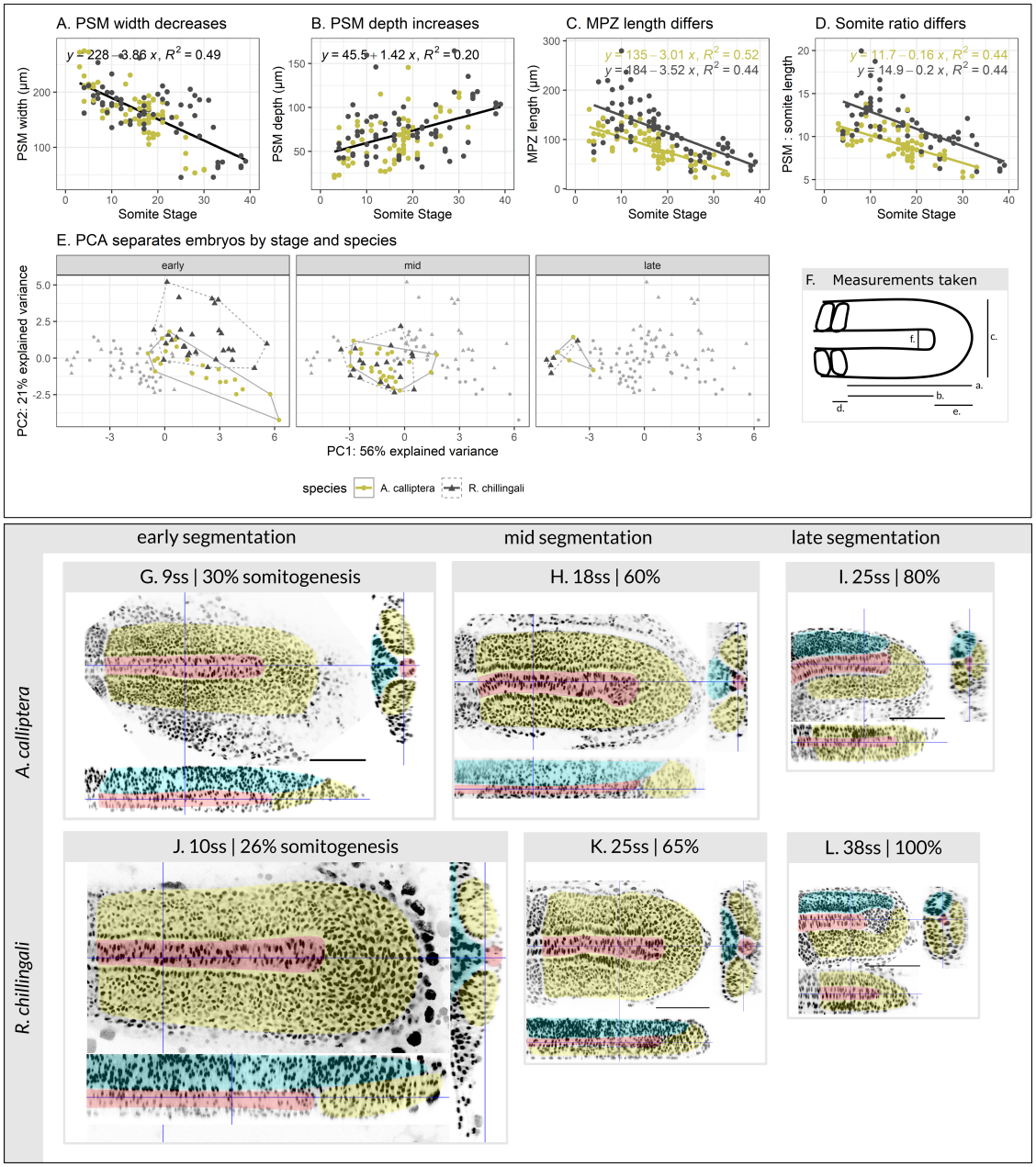


Figure 4.3: PSM anatomy differs only at the onset of segmentation. A-D: Change in selected traits over time. All other metrics are presented in Figure A.4. Only one regression line shown for metrics which do not differ between species. E: PCA of given variables separates species only at the early stage of segmentation. F: Schematic of variables used for PCA. a: PSM length. b: Notochord length. c: PSM width. d: nascent somite length. e: MPZ length. f: notochord length. Also measured were PSM depth, PSM length to width ratio, PSM length to depth ratio, PSM width to depth ratio, and PSM length to somite length ratio. Grey points indicate measurements from other stages. G-L: Single confocal slices showing the anatomy of the PSM over time. The middle panel depicts a sagittal slice of the PSM at the level of the notochord; the bottom panel depicts a dorso-ventral slice of the PSM through the notochord, from neural tube to notochord, and the right panel depicts a dorso-ventral slice of the PSM oriented medio-laterally, at the middle of the PSM. Tissues are false-coloured as follows. PSM: yellow. Notochord: red. Neural tube: cyan. DAPI signal is shown in greyscale. Embryos are oriented anterior left; scale bar 100μm.

As for other metrics, the relationship between variables does not differ between species, meaning that while the position of the wavefront differs between species, this difference is again established at the onset of segmentation (Figure 4.3D).

In summary, the shape of the PSM is the same between species in most dimensions. Exceptions include PSM length, notochord length, and the length of the MPZ. This matches what is seen qualitatively, where the anatomy of the PSM is very similar between species (Figure 4.3G-L); the only visually obvious difference is in the size of the PSM at early segmentation (Figure 4.3G, J).

I then performed a principle component analysis (PCA) on this dataset, to identify whether species differ at one stage more than another. PCA separates embryos by stage: early stage embryos lie at the middle/middle right of the graph when graphed by PC1 and PC2; while mid-segmentation embryos are in the middle and late segmentation to the left (Fig 4.3G). I found that while *A. calliptera* and *R. chillingali* cluster together at mid and late stages of segmentation, they do not cluster at early segmentation (Fig 4.3G). Instead, *R. chillingali* lie at the far right of the plot while *A. calliptera* are located more centrally. Thus, morphometric analysis reveals that the only distinguishable interspecific differences are those present at the early stages of segmentation.

Taken together, this analysis shows that our two cichlid species are measurably different only at the earliest stages of segmentation. Even traits, such as MPZ size and PSM length, that do differ between species, differ in the starting value (the intercept of the trendline), not the slope of the trendline. This implies that while the starting point of the developmental process differs between species, the process of axial elongation during somitogenesis itself is operating similarly in both cichlid species. This suggests that *R. chillingali* may have evolved an increased number of somites by evolving the size of the PSM at the onset of segmentation, rather than by altering the process of segmentation.

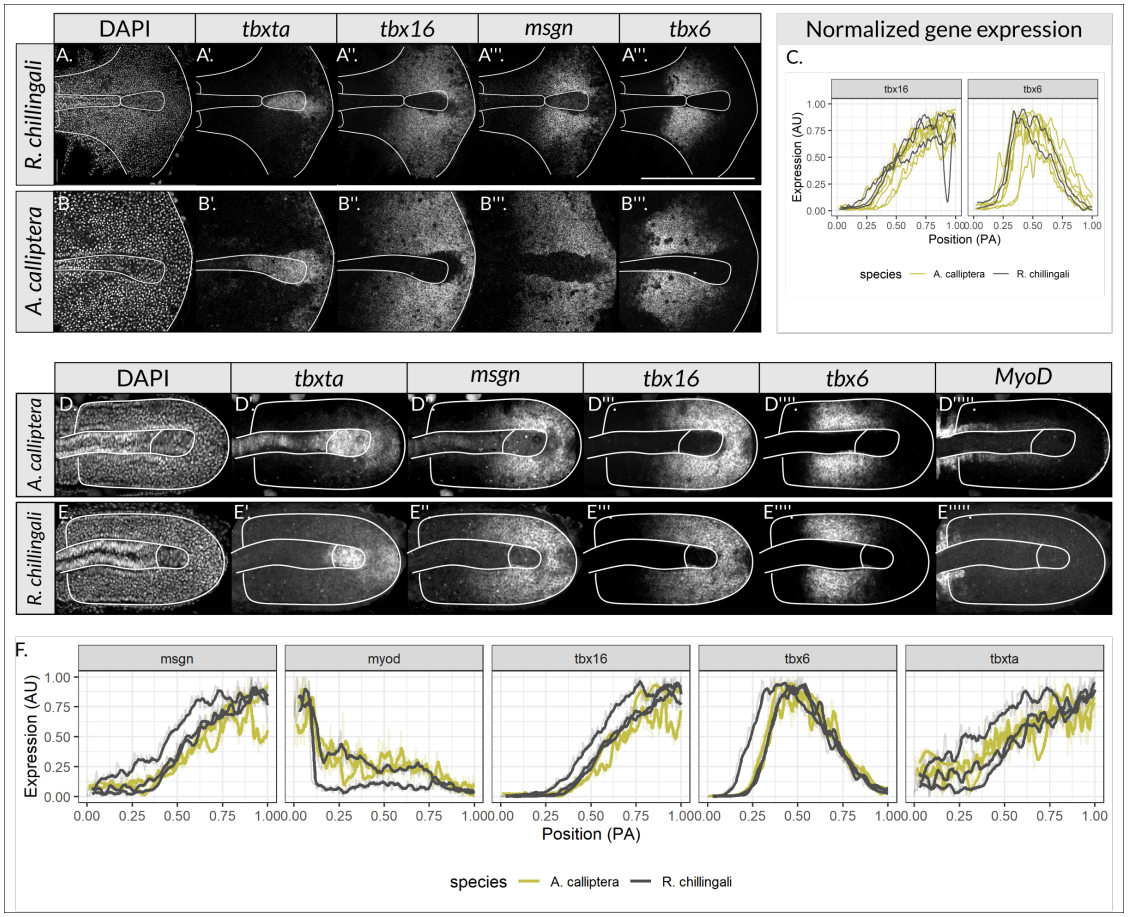


Figure 4.4: Patterning of the PSM is identical between species. A-B: single confocal sections of very early segmentation (3ss) embryos stained for DAPI, *tbxta*, *tbx16*, *msgn*, and *tbx6*. Scale bar: 500um. Thick white lines outline the nascent somites, notocord, neural tube, and unsegmented region boundary. C: Normalized gene expression measured across multiple embryos. Normalization performed as described in the Methods; *tbx6* and *tbx16* expression is measured along the middle of the PSM, across the same line for each gene. N_{AC} : 6. N_{RC} : 2. D-E: Single confocal sections of mid-segmentation embryos stained for DAPI, *tbxta*, *tbx16*, *msgn*, *tbx6*, and *MyoD*. Thick white outlines outline the PSM, notocord, and neural tube. F: quantifications of patterning for each gene at mid segmentation, normalized as described in the methods. Expression of *tbxta*, *tbx16*, *msgn*, and *tbx6* are measured along the same line, across the midline of the embryo. Expression of *MyoD* is measured along the adaxial layer of cells. $N_{AC} = 2$. N_{RC} : 3. All embryos are oriented anterior left, and the same embryo is shown in each row, except for B'''.

4.2.3 Molecular anatomy of the tailbud is similar between species

I then attempted to further test the notion that the process of segmentation has not evolved between these cichlid species. I did this by further characterizing expression of genes required for axial elongation and segment production in other

species, to investigate whether *A. calliptera* or *R. chillingali* express either of these genes differently. I therefore characterized expression of genes well-conserved in the process of axial elongation in vertebrates. The gene *tbxta* is required for both specification of the notochord, and specification of the posterior body, in vertebrates. As in other species, *tbxta* is expressed in the notochord in *A. calliptera* and *R. chillingali*, as well as in the germ ring in early embryos (Figure 4.4A', B'). At mid segmentation, *tbxta* is expressed in the notochord and in the very posterior of the PSM. Expression is similar between species, and very similar when quantified at mid segmentation (Figure 4.4A', B', D', E', F). The genes *tbx16* and *msgn* are required for correct cell migration in the zebrafish tailbud, and are expressed in a posterior-to-anterior gradient (Manning & Kimelman, 2015). This expression is conserved in cichlids, and the slope and position of the gradient, when quantified, is very similar between species at both early and mid-segmentation (Figure 4.4A"', A", B", B"', C, D", D"', E", E"', F). The gene *tbx6* is required for somite maturation after each somite has been specified during somitogenesis. In both *R. chillingali* and *A. calliptera*, as in zebrafish, *tbx6* is expressed in a broad band of gene expression, with a sharp gradient anteriorly and a more shallow gradient in the posterior (Figure 4.4A"", B"", D"", E""). As with other genes patterning the PSM, expression of *tbx6* when quantified is very similar between species, both at mid and late segmentation. Finally, *MyoD* is involved in muscle development in the newly formed somites, and is expressed in the nascent somite and in a single row of cells directly adjacent to the notochord, the adaxial cells, in both species (Figure 4.4D""', E""', F). When quantified, expression of this gene, like the other genes characterized, is likewise indistinguishable between species.

Altogether, then, *A. calliptera* and *R. chillingali* do not differ in patterning of the PSM, qualitatively or quantitatively, at early or mid segmentation stages. I am thus unable to identify a change to the process of segmentation or axial elongation which could contribute to the evolution of somite numbers in these species.

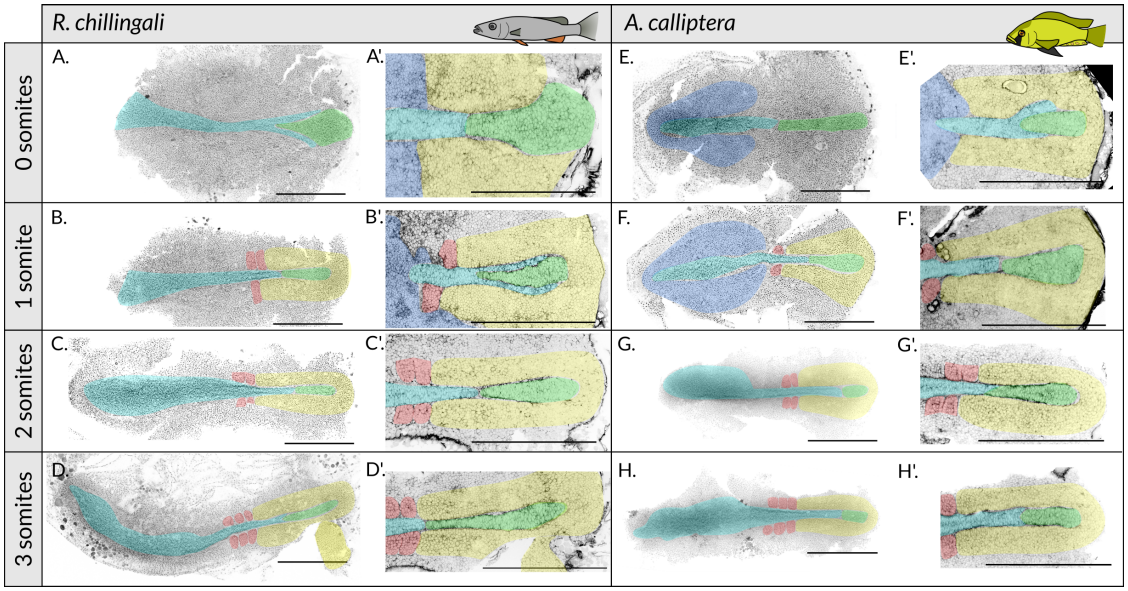


Figure 4.5: Anatomy of the early embryo and unsegmented region at the onset of segmentation. A-H: whole mount, dissected cichlid embryos stained with DAPI (nuclei) and phalloidin (actin). Images to the left (A-H) are shown complete, with nuclei in grey; images to the right (A'-H') magnify the unsegmented region, and show actin/phalloidin in grey. Individual tissues are false-coloured as follows: neural tube: cyan; notocord/notocord projenitors: green; somites: red; unsegmented mesoderm: yellow; ectoderm: dark blue. Scale bar: 500μm. All embryos are oriented anterior left.

4.2.4 Anatomy of the body axis and the PSM is indistinguishable between species at the onset of segmentation

I then wished to know whether the differences between species are established at the onset of segmentation, or prior to segmentation. I therefore characterized the onset of segmentation in greater spatial and temporal detail, fixing and staining embryos at hourly intervals at the onset of segmentation. For both species, I was able to obtain embryos with 0-4 somites, indicating that neither species forms significant numbers of somites simultaneously, consistent with findings in zebrafish and chick (Figures A.2, A.3, 4.5) (Schröter et al., 2008; Maia-Fernandes et al., 2024).

Despite the fact that *R. chillingali* embryos at the onset of segmentation (0-3ss) are older than their *A. calliptera* counterparts, the anatomy of the embryos is identical between species. Compared to later embryos, the PSM appears morphologically continuous with the lateral plate mesoderm, and somites have

not yet epithelialized laterally (compare Figures 4.5 and 4.3G-L). However, the only obvious difference between species is in the overall size of the PSM. This suggests, alongside other data, that the difference between species is established prior to the onset of segmentation.

4.3 Discussion

Altogether, these results imply that *R. chillingali* have evolved an increased number of somites by evolving the initial conditions of somitogenesis, rather than by modifying the process of somitogenesis. James Hammond has shown that *R. chillingali* and *A. calliptera* differ in the duration of somitogenesis, and that the slight change in the rate of somitogenesis observed does not explain the difference in somite number between species. Moreover, at early segmentation, the *R. chillingali* PSM is 19% longer, 61% greater in volume, and has 57% more cells than that of *A. calliptera* (Figure 4.2). In addition, the nascent somite does not differ in volume between both species, though the *A. calliptera* somite is greater in length (Figure 4.2). Therefore, the excess tissue required to produce more somites in *R. chillingali* is likley present at the onset of segmentation. There is also no evidence for a difference in the segmentation process between species, as species only differ in PSM shape at the early stage of segmentation (Figure 4.3). Finally, I investigated genetic patterning of the PSM at early and mid segmentation, to investigate whether more subtle genetic process could be impacting the process of segmentation and axial elongation, and found no difference between species. Altogether, this shows that *R. chillingali* have likely evolved an increased somite count by modifying the initial conditions of somitogenesis.

4.3.1 Cichlids may have evolved somite counts by modifying the initial conditions of somitogenesis

Our data have shown that the species differences in PSM size are established prior to the onset of segmentation. Despite the difference in PSM size between species, the anatomy and patterning of the PSM is very similar between species even at the

very earliest stages of segmentation, 1-3ss (Figures 4.4, 4.5). Additionally, there is no evidence for a difference in the dynamics of PSM size at the early stages of segmentation/prior to 15ss, as the rate of change of PSM length, and PSM volume, does not differ between species prior to 15ss (Figure 4.2). The timing of the onset of segmentation also differs between species: *R. chillingali* commence segmentation slightly later, at 36hpf, while *A. calliptera* begin segmenting at 31hpf (James Hammond, unpublished data). It is possible that *R. chillingali* therefore have more time for cell division during gastrulation and an associated increase in size of the embryonic body axis prior to the onset of segmentation.

In addition, I found no evidence for a change in the somitogenesis process itself. If the process of somitogenesis and axial elongation did differ between species, we would expect to see either a difference in the molecular or structural anatomy of the PSM between species. We would also expect to see a difference in the rate of change of anatomical measurements between species. This was not observed. Only a small number of anatomical measurements differ between species: namely, PSM length, somite length, length of the MPZ, notochord length (which is strongly associated with PSM length), and the ratio between the length of the PSM and the nascent somite (Figures 4.3, A.4). Even for the metrics exhibiting species-specific differences, the rate of change is the same over developmental time, it is the starting conditions which differ between species. This implies that the developmental processes driving morphogenesis are the same between species. Further supporting this idea, expression of key genes controlling axial elongation is indistinguishable between species.

Recent work provides further support for this idea. Marconi et al., 2024 perform RNA sequencing on embryos of different stages in both *A. calliptera* and *R. chillingali*. Though they do identify consistent species differences in transcriptomic profiles, the largest species differences are present at 4ss; very early in segmentation (Marconi et al., 2024). This finding supports our data, as again, the most sizeable transcriptomic differences occur at very early stages of

segmentation. It also implies that differences in other phenotypic traits may also be establised at early segmentation or prior to segmentation.

Altogether, this suggests that *R. chillingali* may have evolved an increase in somite number by evolving the initial conditions of somitogenesis.

4.3.2 A unique mechanism of epiboly in African cichlids may contribute to morphological evolution

A. calliptera and *R. chillingali* exhibit an unusual embryological trait: the completion of epiboly is delayed until late in segmentation (>28ss in *A. calliptera*) (Marconi et al., 2023). I performed a literature review of *Cichlidae* species development, and was able to find descriptions of development for *Cichlasoma dimerus*, Blood parrot cichlid (*Cichlasoma synspilum* x *C. citrinellum* hybrid), *Amphilophus xiloaensis*, *Thorichthys meeki*, *Oreochromis niloticus*, *Etroplus suratensis*, *Amatitlania nigrofasciata*, *A. calliptera*, and *Haplochomis velvet black*. Every one of these species completes epiboly early in segmentation (Meijide & Guerrero, 2000; Fujimura & Okada, 2007; Zhijing et al., 2014; Kratochwil et al., 2015; Bindu & Padmakumar, 2016; Jung et al., 2018; Gültekin & Yilmaz, 2023). The exceptions are *A. calliptera*, *R. chillingali* (Lake Malawi cichilds), and *Haplochomis "velvet black"* (from Lake Victoria): all haplochromine cichlids from Lake Malawi or Lake Victoria (Marconi et al., 2023; Fleig, 1993; de Jong et al., 2009). Figure 4.6 depicts cichlid development in more detail. I have also been unable to find an example of a non-cichlid teleost that does not complete epiboly early in segmentation. Therefore, it appears that delayed completion of epiboly is unique to haplochromines. I hypothesize that delayed completion of epiboly has contributed to the extreme evolvability of vertebral count in African cichlids. Delaying completion of epiboly until later in segmentation may have allowed more flexibility and adaptability during early development, when key tissues such as somites and neural crest cells are established. Ultimately this could have contributed to the stunning evolution of phenotypic diversity in African cichlids.

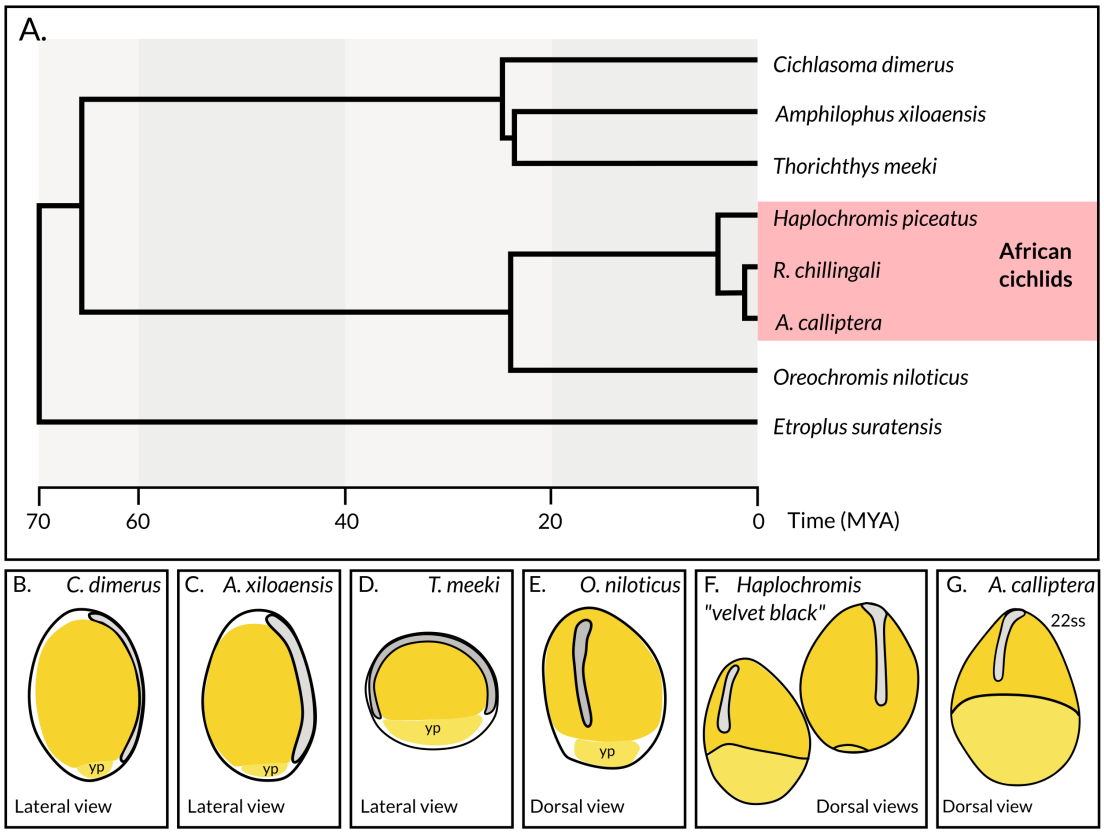


Figure 4.6: Delayed completion of epiboly is unique to African cichlids. A. Phylogeny of cichlid species where development has been described. Divergence times are taken from the timetree project (Kumar et al., 2022). B-G: Cartoons of late-epiboly stage cichlids, traced from images in journal articles. B: *C. dimerus* form a tailbud immediately before completion of epiboy (Meijide & Guerrero, 2000). Specimen was imaged using widefield microscopy. C: *A. xiloaensis* (from the Midas cichlid flock) have six to eight somites at 90% epiboly (Kratochwil et al., 2015). Specimen was imaged using widefield microscopy. D: *T. meeki* (firemouth cichlid) undergo complete epiboly prior to signficant numbers of somites forming (Gültekin & Yilmaz, 2023). Specimen was imaged using widefield microscopy. E. *O. niloticus* (Nile tilapia) have eight somites at the completion of epiboly (Fujimura & Okada, 2007). Specimen traced was stained using hematoxylin. F. *Haplochromis "velvet black"* has an obvious body axis prior to the completion of epiboly (Fleig, 1993). Left specimen was stained with Osmium; right imaged under a scanning electron microscope. de Jong et al., 2009 characterize development of *Haplochromis piceatus* in more detail but unfortunately do not depict the embryo developing onto the yolk. G. *A. calliptera* does not undergo complete epiboly until the end of segmentation (>28ss) (Marconi et al., 2023). Embryo shown is a 22ss embryo, injected with Texas Red to better visualize embryonic tissues (Marconi et al., 2023). Not shown is the development of the Blood Parrot hybrid cichlid, *Amatitlania nigrofasciata*, and *Etroplus suratensis* which all undergo complete epibboly early in segmentation (Zhijing et al., 2014; Jung et al., 2018; Bindu & Padmakumar, 2016). yp: yolk plug, which forms at the end of epiboly.

4.3.3 Evolution via boundary conditions of embryogenesis

Our findings suggest a simple mechanism to achieve evolution of phenotypic traits. Instead of modifying a complex developmental process, such as segmentation or axial elongation, it may be simpler to instead evolve the initial conditions or the boundary conditions of that process. These, potentially subtle, changes early on in development could cascade and produce much more dramatic phenotypic consequences. In Lake Malawi cichlids, these initial conditions are the initial size of the body axis and modifying the size of the body axis appears to allow somite number evolution.

Other species also achieve developmental innovations by modifying development prior to the onset of segmentation. *Austrofundulus linneaus* (killifish) are able to disaggregate and reaggregate all cells within the embryo between the completion of epiboly and segmentation (Podrabsky et al., 2017). This remarkable process allows this species to arrest development and survive periods without water (Podrabsky et al., 2017). At the very least, this shows that teleost development can be robust to dramatic perturbations prior to segmentation. It also provides an example of a life-history adaptation, much like the evolution of somite counts, which is achieved by modifying development immediately prior to segmentation, as occurs in Lake Malawi cichlids. The two-tailed goldfish, in addition, is thought to have evolved a second tail by altering dorso-ventral patterning at the gastrula and bud stage (Abe et al., 2014). This species also exhibits delayed blastopore closure relative to single-tailed goldfish, and has a larger PSM, though this does not seem to result in evolution of somite counts (Li et al., 2019). Again, then, a developmental innovation is achieved by modifying the initial conditions of somitogenesis.

Athropods may also evolve phenotypic diversity by evolving the boundary conditions of segmentation. Different arthropod species achieve segmentation using different methods: some use a method similar to the segmentation clock, others (such as *Drosophila*) appear to pre-pattern body segments, while others

use an intermediate approach. These different methods of segmentation can all be produced using the same GRN but by modifying the boundary conditions of the segmentation process, in this case by altering expression of so-called timer genes (Clark & Akam, 2016; Clark, 2017; Clark & Peel, 2018; Taylor & Dearden, 2022). Therefore, evolution via initial or boundary conditions may occur in species other than Lake Malawi cichlids.

4.4 Conclusions

Our data show that *R. chillingali* are likely to have evolved increased somite counts by modifying the initial conditions of segmentation, not the process of segmentation. *R. chillingali* have a significantly larger PSM than *A. calliptera* at the onset of segmentation, providing the excess tissue used to produce more somites. Moreover, there is no evidence for a change in the segmentation process between species, as morphometric analysis of the anatomy of the PSM, and gene expression patterning of the PSM are very similar between *R. chillingali* and *A. calliptera*. This therefore presents a very simple mechanism to achieve phenotypic evolvability: simply altering the initial conditions of development can result in dramatic phenotypic changes, without needing to modify a complex developmental process.

Chapter 5: Approximated gene expression trajectories for gene regulatory network inference on cell tracks

Contents

5.1 Introduction 94
5.1.1 Studying pattern formation in the absence of morphogenesis . . 95
5.1.2 Zebrafish tbox patterning is an ideal system to study the interaction of patterning and morphogenesis 96
5.1.3 Live modelling overview and prior work 100
5.2 Results 102
5.2.1 Different AGET construction methods yield similar results. . . . 102
5.2.2 Optimal fits can be obtained using as few as 10% of AGETs for reverse engineering. 102
5.2.3 Network Validation 107
5.2.4 Parameter sets reproduce gene expression at the single cell level 108
5.2.5 Parameter sets recapitulate tissue-level gene expression patterning 110
5.2.6 Parameter sets predict known genetic interactions 112
5.2.7 Parameter sets recapitulate some genetic perturbations 115
5.2.8 Parameter sets can reproduce upregulation of *tbx6 in silico* . . . 117
5.3 Discussion 119
5.3.1 Parameter sets recapitulate key biological behaviours of the tbox GRN 119
5.3.2 "Live modelling" allows one to investigate patterning in a tissue undergoing morphogenesis 120
5.3.3 Future work to model the tbox GRN in zebrafish 122
5.4 Conclusions 124

5.1 Introduction

How spatial patterns are established and maintained in developing tissues is a central question in evolutionary developmental biology. An extensive body of work investigates pattern formation during embryonic development, for example during gap and pair rule pattern formation in the *Drosophila* blastoderm, and dorso-ventral patterning of the vertebrate neural tube (Jaeger et al., 2004; Manu et al., 2009a; Verd et al., 2017; Zagorski et al., 2017). In both these systems, pattern formation is occuring in the absence of cell movements. Pattern formation in the *Drosophila* blastoderm occurs prior to gastrulation and the accompanying cell rearrangements (Jaeger, 2011; Schroeder et al., 2011; Clark et al., 2019), while patterning of the dorso-ventral axis of the vertebrate neural tube occurs after the neural tube has formed (Frith et al., 2024). However, such systems are unusual; many developmental systems achieve pattern formation and cell differentiation at the same time as morphogenesis. Examples include the vertebrate limb bud, and patterning of the vertebrate body axis during segmentation (McQueen & Towers, 2020; Fulton et al., 2022a). Therefore, new methods are required to infer and simulate the gene regulatory networks (GRNs) achieving patterning in tissues undergoing morphogenesis.

Live Modelling presents a framework with which to achieve this (Fulton et al., 2022a). In this methodology, the gene regulatory networks controlling patterning are inferred and simulated on cell movements tracked by live-imaging a developing tissue. This allows cell movements to be explicitly included in a model of pattern formation (Fulton et al., 2022a).

In this chapter, I outline my efforts to establish, troubleshoot, and validate a protocol to reverse-engineer GRNs using Live Modelling. This work has been published as a research article in *iScience*, which I am co-first author on (Speiss, Taylor et al. 2024, also included in Appendix **??**). In this chapter, I describe my own contributions to this work.

5.1.1 Studying pattern formation in the absence of morphogenesis

How GRNs achieve patterning has been extensively studied in tissues where cells are not rearranging. For example, in the *Drosophila* blastoderm, patterning of the body axis occurs prior to gastrulation and in the absence of morphogenesis and cell rearrangement (Jaeger, 2011; Clark et al., 2019). An extensive body of work reverse-engineers the GRNs controlling pattern formation in the *Drosophila* blastoderm (e.g. Jaeger et al., 2004; Manu et al., 2009a; Jaeger, 2011; Crombach et al., 2012, 2016; Verd et al., 2019). In order to reverse-engineer GRNs controlling pattern formation, an understanding of the gene expression trajectories experienced by individual cells in the tissue of interest is required. For the gap gene system, the spatial pattern of protein or mRNA expression of genes involved in gap gene patterning of the *Drosophila* anterior-posterior axis are quantified (Jaeger et al., 2004; Crombach et al., 2012). The boundaries of gene expression for each gap gene are then identified at each stage (Crombach et al., 2012). Because cells are not moving or rearranging in the *Drosophila* blastoderm, it is possible to infer the expression of genes in individual cells by simply interpolating between different timepoints of gene expression (Crombach et al., 2012). Gene expression models, formulated as coupled ordinary differential equations (ODEs) can then be paramaterised by reverse-engineering to these interpolated gene expression trajectories (Jaeger et al., 2004; Crombach et al., 2012). This approach has been very fruitful in understanding the development and evolution of gap gene pattern formation in *Drosophila* and other invertebrates (Jaeger et al., 2004; Manu et al., 2009a; Verd et al., 2019).

However, in most animal systems, morphogenesis is occurring alongside pattern formation. For example, in zebrafish embryos, extensive cell rearrangement is occurring in the PSM alongside patterning (Kanki & Ho, 1997; Lawton et al., 2013; Mongera et al., 2018; Thomson et al., 2021) and establishment of the embryonic body axis involves extensive convergence and extension cell movements during gastrulation (Kimmel et al., 1995; Shah et al., 2019). Therefore,

modelling and inference methodologies are required which are able to incorporate morphogenesis into models of gene regulatory network (Fulton et al., 2022a). This will allow us to understand the development of a broader range of tissues and organs. It will also allow us to understand a potential generative role of cell movements in achieving pattern formation (Fulton et al., 2022a). Live modelling approaches this problem by allowing gene regulatory networks to be reverse-engineered in a tissue undergoing morphogenesis and cell rearrangements.

5.1.2 Zebrafish tbox patterning is an ideal system to study the interaction of patterning and morphogenesis

Zebrafish tbox patterning represents an ideal system to establish a methodology to reverse-engineer the GRNs achieving patterning during morphogenesis. There are a number of reasons for this. Firstly, the tbox GRN has been well-characterized in zebrafish, meaning that while the GRN still needs to be parameterised, the sign of some genetic interactions are known and genetic perturbations are available and provide a way to validate, or exclude, the parameter sets obtained (see Table 5.1, Figure 5.2). Secondly, expression of the tbox genes scales with the PSM length over time in zebrafish, simplifying the processes of AGET construction (Fulton, 2021). That is, *tbx6* expression peaks consistently at 50% PSM length, even though the PSM is changing in length over time, and *tbx16* expression similarly peaks at about 15% PSM length (Fulton, 2021, see Figure 5.1). Expression of signalling factors regulating tbox pattern formation, FGF and Wnt, likewise scale with PSM length (Fulton, 2021).

Extensive experimental work has been done to characterize the tbox GRN during zebrafish morphogenesis [1]. Two strands of experimental evidence have been used to characterize tbox GRN topology: demonstration of a direct interaction

[1] A brief note on gene nomenclature in zebrafish: the tbox genes have been given different names at different times. For clarity I am using only their modern names *tbxta*, *tbx16*, and *tbx6*.
tbxta is also known in the literature as *Brachyury*, *No-tail/ntl*, *T*, etc. *tbx16* has also been known as *spadetail/spt*. The naming of *tbx6* is particularly confusing: the gene currently named *tbx6* in zfin is also known as *tbx24* and *fss*. In the past, the gene *tbx16l* has also been called *tbx6*; *tbx16l* is more similar in homology and function to *tbx16* than *tbx6* orthologues.

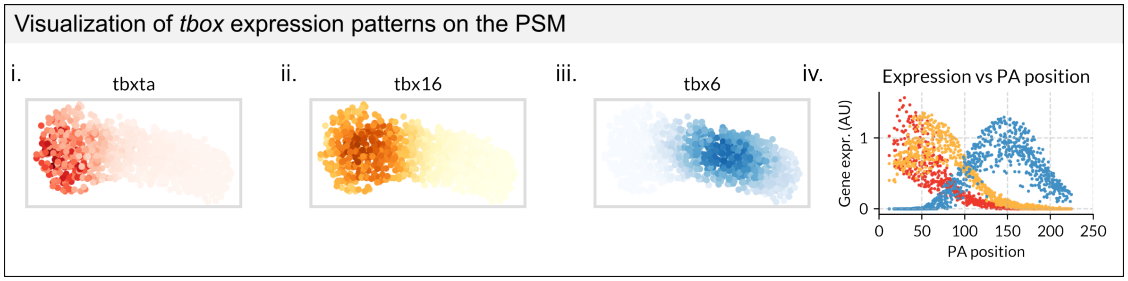


Figure 5.1: Visualization of *tbox* gene expression in the PSM. i, ii, iii: *tbxta*, *tbx16*, and *tbx6* expression plotted against cell position in the tailbud. vi: *tbxta* (red), *tbx16* (yellow), and *tbx6* (blue) expression against posterior-anterior (PA) position in the tailbud, in μm.

between genes via an enhancer, and showing a (potentially indirect) interaction based on the change in expression of one gene following perturbation of the other. See Figure 5.2 for a summary of known interactions in the tbox GRN.

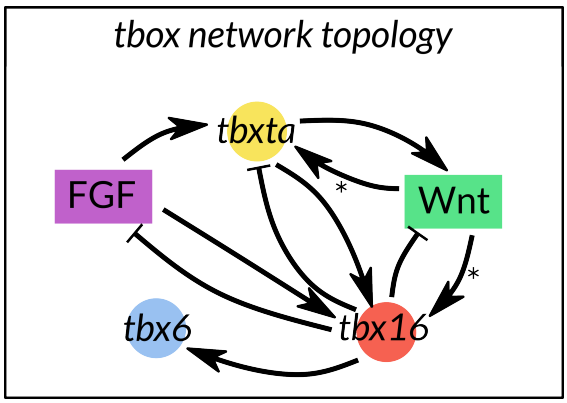


Figure 5.2: Known genetic interactions in the tbox GRN. Asterisks indicate interactions known to be direct.

There are two genetic interactions proven to be direct in the zebrafish tailbud: Wnt activates *tbx16* via a Wnt-binding promoter (Bouldin et al., 2015), and Wnt activates *tbxta* via Wnt binding promoter(s) (Martin & Kimelman, 2008; Harvey et al., 2010; Ye & Kimelman, 2020; Ye et al., 2021).

Other genetic interactions have been inferred via analysis of gene expression following gene perturbations. For example, *tbxta* is thought to activate Wnt, meaning that *tbxta* and Wnt signalling together form a self-organising feedback loop, with both genes upregulating each other (Goering et al., 2003; Martin & Kimelman, 2008; Harvey et al., 2010; Ye & Kimelman, 2020; Ye et al., 2021). Other suspected regulators of *tbxta* include *tbx16*: *tbxta* is upregulated in *tbx16* mutants, meaning that *tbx16* likely directly or indirectly inhibits *tbxta* (Warga et al., 2013).

Table 5.1: Evidence for known genetic interactions in the tbox GRN. Genetic interactions are shown by up- or down-regulation of one gene in response to loss of another, and could be indirect (mediated by another gene). Interactions known to be direct, as evidenced by identification of a promoter element, are indicated.

Interaction	Evidence
Wnt activates *tbxta* (via promoter)	Martin and Kimelman, 2008 show a genetic interaction between Wnt and tbxta. Harvey et al., 2010 identify a Wnt-binding enhancer in *tbxta*. Ye et al., 2021 and Ye et al., 2021 further characterize Wnt-binding enhancers in *tbxta*.
tbxta activates Wnt (genetic interaction)	Loss of *tbxta* results in decreased *wnt8* expression (Goering et al., 2003)
Wnt activates *tbx16* (via promoter)	Bouldin et al., 2015 identify a Wnt-driven promoter in *tbx16*, and show decreased *tbx16* levels result after loss of Wnt signalling.
tbxta activates *tbx16* (genetic interaction)	Loss of *tbxta* results in loss of *tbx16* (Griffin et al., 1998)
tbx16 activates *tbx6* (genetic interaction)	Bouldin et al., 2015 show that upregulating *tbx16* increases *tbx6* expression and Fior et al., 2012 show *tbx16* mutants have lower levels of *tbx6* expression.
tbx16 inhibits *tbxta* (indirect; via Wnt)	Bouldin et al., 2015 show overexpression of *tbx16* causes decreased levels of *tbxta* in the tailbud, but that this is rescued by overexpression of Wnt, suggesting an indirect interaction via Wnt.
FGF inhibits *tbxta* (genetic interaction)	Goto et al., 2017 show that following loss of FGF signalling, *tbxta* is upregulated.
FGF activates *tbx16* (genetic interaction)	Goto et al., 2017 show that loss of FGF signalling results in loss of *tbx16* expression.
tbx16 inhibits FGF (genetic interaction)	FGF is upregulated in *tbx16* (*spt*) mutants (Warga et al., 2013)
tbx16 inhibits Wnt (genetic interaction)	*wnt8* is downregulated in *tbx16* mutants (Goering et al., 2003)
tbx16 activates *tbx6* (genetic interaction)	Upregulation of *tbx16* results in increased *tbx6* expression; loss of *tbx16* results in decreased *tbx6* expression (Fior et al., 2012; Bouldin et al., 2015)
FGF activates *tbxta* (genetic interaction)	Loss of FGF results in expansion of *tbxta* expression (Goto et al., 2017)

In turn, *tbxta* expression is thought to be required for *tbx16* expression; with loss of *tbxta* leading to a loss of *tbx16* expression (Griffin et al., 1998). This suggests *tbxta* may activate *tbx16*. Both *tbxta* and *tbx16* are also thought to be reliant on FGF signalling (Griffin et al., 1998). Expression of *tbx16* is thought to inhibit both FGF and *tbxta*; the latter interaction likely being mediated via Wnt (Warga et al., 2013; Bouldin et al., 2015). Finally, *tbx16* is thought to upregulate *tbx6* (Fior et al., 2012; Bouldin et al., 2015). These suspected genetic interactions provide a way to validate the reverse engineering methodology I will develop in this chapter: we expect parameter sets obtained to reproduce many of these genetic interactions.

The morphogenesis of the zebrafish PSM is also well-understood. It is unclear how robust tbox patterning is achieved alongside morphogenesis in this tissue. Zebrafish achieve somitogenesis and axial elongation in the absence of overall growth of the body axis: cell division is minimal, and the majority of tissue required for somitogenesis is present in the PSM at the onset of axial elongation (Kanki & Ho, 1997; Steventon et al., 2016; Thomson et al., 2021). There are different patterns of cell movement along the zebrafish PSM: cells in the posterior are more motile/fluid like, and cells in the anterior PSM tend to swap neighbours less and are less motile (Lawton et al., 2013; Das et al., 2017; Thomson et al., 2021). This is thought to be due to a fluid-to-solid like transition in the PSM: cells in the posterior are more motile, acting like a fluid; and cells in the posterior are in a solid-like state (Lawton et al., 2013; Mongera et al., 2018). This notion is backed up by direct measurement of the pressure within the PSM (Mongera et al., 2018). A side effect of this heterogeneity in cell movement is that cells take different amounts of time to transit the PSM (Fulton, 2021). This presents a challenge to our understanding of pattern formation, to tackle using the live modelling methodology: how is coherent tbox pattern maintained in the presence of cell rearrangements and such differences in cell motility (Fulton et al., 2022a)?

5.1.3 Live modelling overview and prior work

To reverse-engineer a GRN incorporating cell movements, we required an approximation of gene expression trajectories in individual cells as they transit the PSM. These are currently impossible to measure directly: while it is possible to track expression of single genes in individual cells in the zebrafish PSM (Soroldoni et al., 2014; Venzin et al., 2023), it is not yet possible to measure five genes plus a nuclear marker simultaneously. We therefore require a method to approximate gene expression trajectories onto cell movements, using fixed samples stained for multiple genes of interest. A method to achieve this was developed by Kay Speiss: in short, fixed embryos stained for genes of interest are aligned with embryos that were live-imaged while stained with a nuclear marker (Spiess et al., 2024). Embryos are aligned via their nuclear position, using the iterative closed point algorithm (ICP). This process is repeated for each timepoint in the live-imaging dataset, yielding approximated gene expression trajectories (AGETs) describing gene expression in moving cells over time. To smooth the resulting heterogeneity in gene expression, a local average of gene expression is taken for each cell in each embryo, before embryos are combined via averaging.

Parameter sets are reverse-engineered using Markov Chain Monte Carlo (MCMC). MCMC attempts to produce parameter sets which can closely reproduce AGET gene expression, thus reverse-engineering the GRN. As Bayesian methods, MCMC algorithms produce distributions of parameter sets (the "posterior distribution"), allowing a variety of different parameter sets which well-reproduce the data to be obtained.

Before I began work on this project, Kay Speiss had developed a method to construct AGETs by aligning fixed embryos stained for gene and protein products with cell tracking data. He had also begun the reverse-engineering process, establishing a method to run MCMC on small numbers of AGETs. My contribution, described here, has been to validate AGET construction, and extensively troubleshoot, optimize, and validate the reverse engineering procedure.

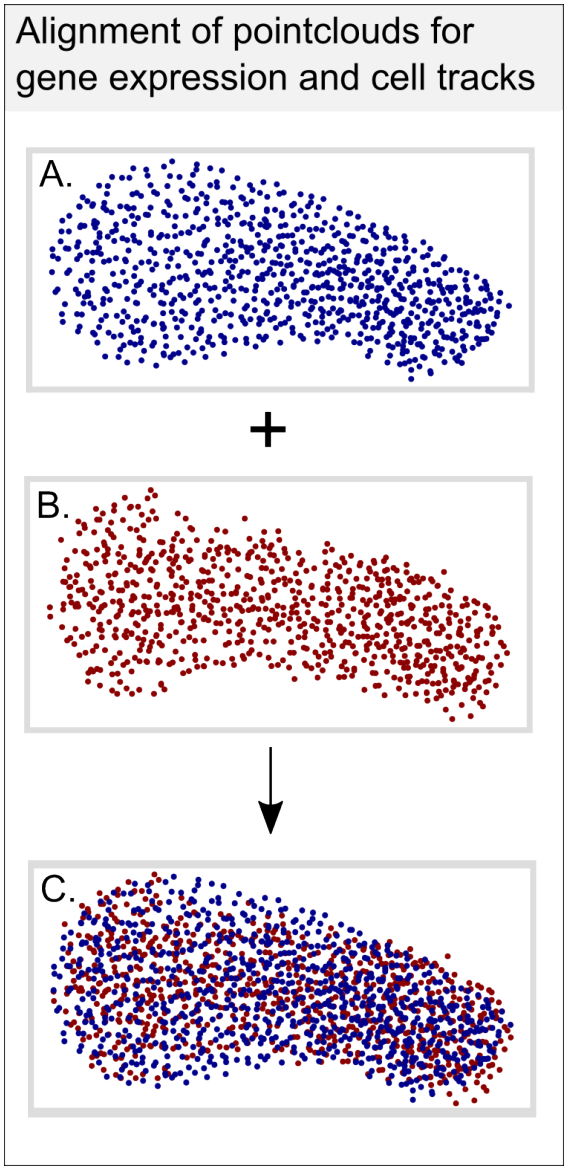


Figure 5.3: AGET construction: pointcloud alignment. Panel A. depicts a pointcloud of nuclei positions from a fixed, stained embryo. Panel B. depicts nuclei positions in the live imaged embryo. Panel C depicts the two embryos overlaid using ICP.

This work is described in this chapter. I first established the best parameters used to construct AGETs. I then optimize the MCMC procedure, and show that using approximately 10% of all AGETs for reverse-engineering provides the best compromise between MCMC efficiency and global goodness of fit, while also leaving large numbers of AGETs free for network validation. Finally, I carefully validate the reverse-engineered networks, showing that these networks are able to faithfully reproduce tbox patterning at the tissue level, recapitulate known genetic interactions within the tbox GRN, and finally reproduce biological perturbations. Altogether, this demonstrates, for the first time, a method to

reverse-engineer GRNs alongside morphogenesis.

5.2 Results

5.2.1 Different AGET construction methods yield similar results.

While a method to construct AGETs had been developed before I began this project, we did not know robust AGET construction was. The key parameter to optimize when constructing AGETs is the amount of spatial averaging to perform: that is, the number of nearby cells' gene expression to consider when constructing AGETs. I therefore tested this parameter, computing the mean expression across five neighbours, the median gene expression across five neighbours, median expression across ten neighbours, and gene expression without any spatial averaging (using a single neighbour) (Figure 5.4). I found that the number of neighbours used does not significantly effect the resulting AGETs: gene expression trajectories from AGETs constructed using different methods are very similar in overall trend (e.g. whether expression is increasing, decreasing, or stable) (Figure 5.4). Using larger numbers of cells for averaging does lead to a smoother time course: compare the red line (median of 10 nearest neighbours) to the green line (no spatial averaging) in Figure 5.4: using ten cells for spatial averaging leads to a smoother AGET than using one cell. Furthermore, gene expression is indistinguishable when graphed relative to position within the PSM for AGETs constructed using different methods (Figure 5.5). Altogether, then, using different strategies to construct AGETs does not signficantly impact the gene expression trajectories produced. Moving forwards, we chose to use AGETs constructed by taking the median of the nearest five cells.

5.2.2 Optimal fits can be obtained using as few as 10% of AGETs for reverse engineering.

Having established the best AGET construction methodology, I then optimized the MCMC procedure, aiming to identify the MCMC parameters producing

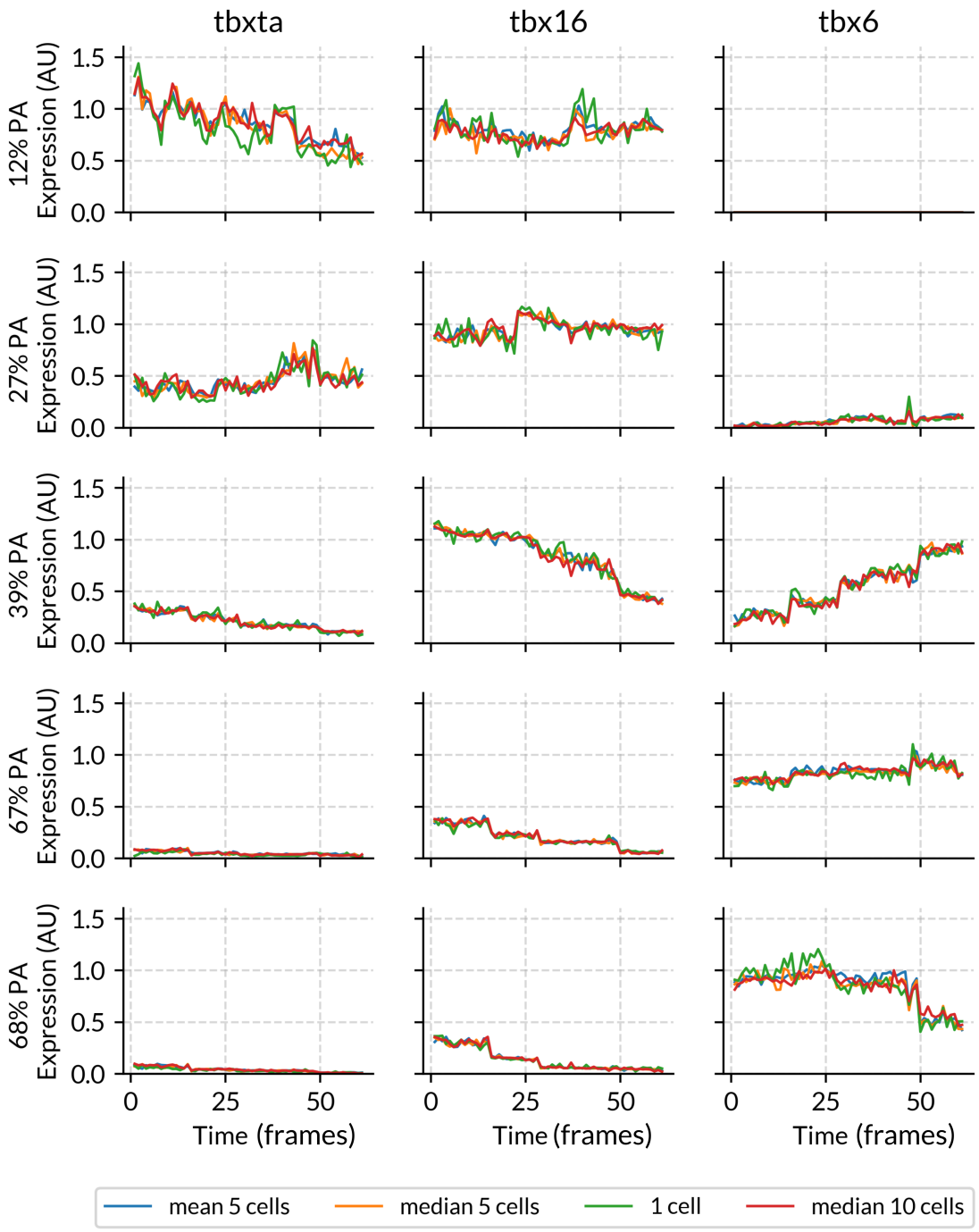


Figure 5.4: Different strategies to construct AGETs result in near-identical outcomes at the cellular level. Panels show approximated gene expression trajectories for different cells, using different construction methods. Columns indicate the different genes (left, *tbxta*; middle, *tbx16*; right, *tbx6*) and rows depict different cells. Position of cells within the embryo is indicated to the left. Colour represents different construction methods: taking the mean of the nearest 5 neighbours ("mean 5 cells"), median of nearest five neighbours ("median 5 cells"), median of 10 neighbours ("median 10 cells"), and the single closest cell ("1 cell").

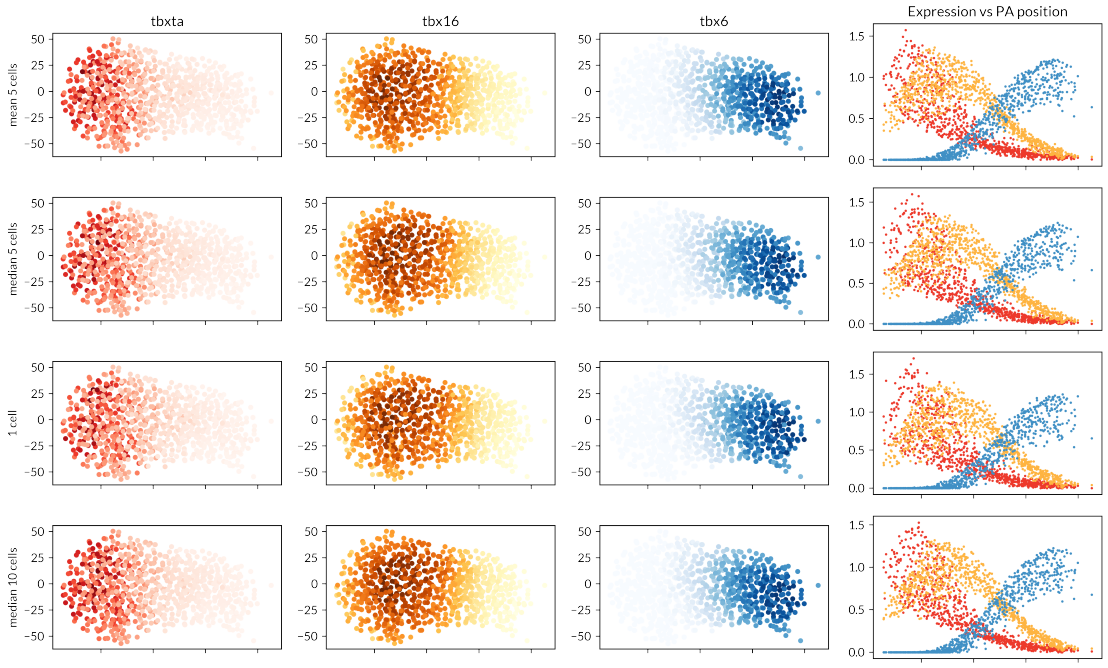


Figure 5.5: Different strategies to construct AGETs result in near-identical outcomes at the cellular level. Rows indicate results from different construction methods: taking the mean of the nearest 5 neighbours ("mean 5 cells"), median of nearest five neighbours ("median 5 cells"), median of 10 neighbours ("median 10 cells"), and the single closest cell ("1 cell"). The first three columns show gene expression graphed onto the embryo: that is, cell's AP position graphed onto X axis, dorso-ventral position graphed onto Y, and cells colourcoded by gene expression of *tbxta*, *tbx16*, and *tbx6* respectively. The rightmost column graphs gene expression (in AU) against posterior-anterior position, for the final timepoint of data.

the best compromise between computational efficiency and goodness-of-fit for the resulting parameters.

The number of AGETs used for reverse-engineering has a significant effect on MCMC performance. To quantitatively compare goodness of fit of the parameter sets between MCMC runs, I took the parameter sets obtained at the final iteration of MCMC, and computed the likelihood score for all AGETs in the embryo. This provides a way to directly compare MCMC runs using different numbers and combinations of AGETs and identify the generalizability of parameter sets. The likelihood score measures the mean squared distance between simulated gene expression trajectories and the AGET (see the Methods for more details). I also computed the acceptance fraction: or the number steps proposed by the sampler

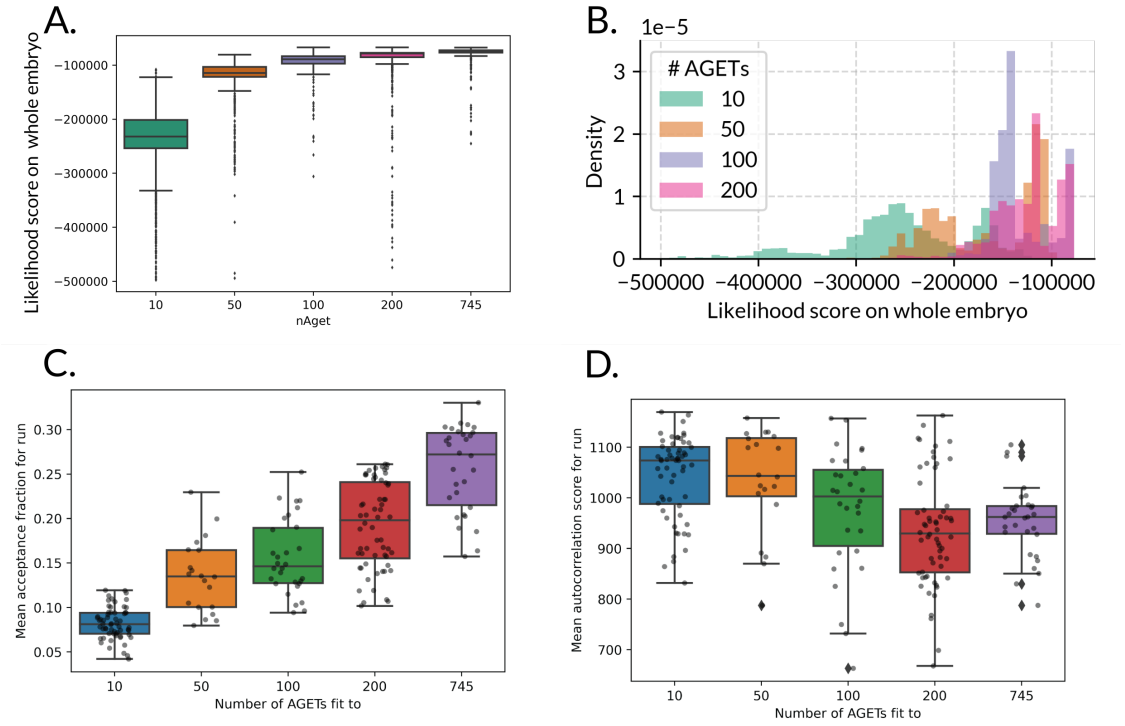


Figure 5.6: Using 200 AGETs for reverse-engineering is the best balance between computation efficiency and network quality. A: Boxplots of likelihood scores (as computed for all AGETs) from parameter sets after using 10, 50, 100, 200, and 745 AGETs for reverse engineering. Results shown are for initial tests. B: Histogram of likelihood scores when reverse engineering to 10, 50, 100, and 200 AGETs. C: Mean acceptance fraction per run increases with the number of AGETs used for fitting in initial tests. D. Mean auto-correlation score per run decreases as the number of AGETs increases until 200 AGETs, and stabilizes thereafter.

that lead to an increase in the likelihood score. For `emcee`, this value should lie between 30 and 50% (Foreman-Mackey et al., 2013). Finally, I computed the autocorrelation time for each run. This metric measures whether runs have converged to a stationary distribution; shorter autocorrelation times mean that the sampler will require fewer iterations to complete (Foreman-Mackey et al., 2013).

I trialled reverse-engineering using different numbers of AGETs, namely 10, 50, 100, 200, and 745 AGETs (of a total of 1903). Increasing the number of AGETs used for reverse-engineering improves the likelihood score, acceptance fraction, and the autocorrelation time (Figure 5.6A-D). Note that increasing the number of AGETs for reverse engineering leads to diminishing returns: the distributions for 200 and 745 AGETs are very similar (Figure 5.6A). Additionally,

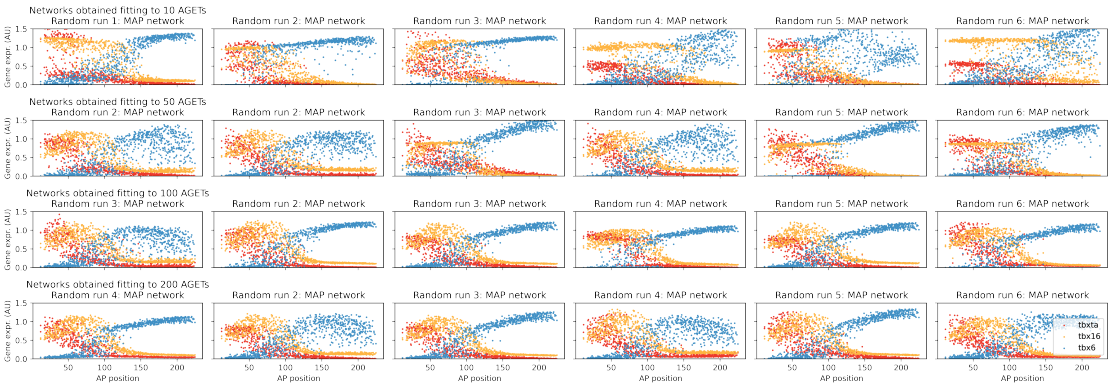


Figure 5.7: Using 200 AGETs for reverse engineering leads to the most robust patterns. Each panel shows the final timepoint of simulations from a randomly chosen parameter set, from the final step of MCMC, for different numbers of AGETs used for reverse engineering. Simulated gene expression in individual cells (Y axis) is graphed against PA position (X axis). Rows depict simulations reverse-engineered using different numbers of AGETs; and columns represent different randomly chosen parameter sets. Parameter sets are taken from the best-scoring network for randomly chosen runs. Note that score is computed for the AGETs used for reverse-engineering, not all AGETs.

it is possible to obtain good networks even when using smaller numbers of AGETs: the maximum likelihood score is similar when using 50-745 AGETs for reverse-engineering, but the spread of scores is tighter when using more AGETs (Figure 5.6A). Thus, using more AGETs for reverse engineering increases both performance and robustness of the reverse engineering procedure.

I then investigated gene expression patterning of the PSM using different numbers of AGETs for reverse engineering. Increasing the number of AGETs used for reverse-engineering leads to more robust patterning (Figure 5.7). When reverse-engineering with ten AGETs, the final pattern obtained is highly variable: while *tbxta* and *tbx16* are typically highly expressed in the posterior and *tbx6* is highly expressed in the anterior, gene expression patterns do not exhibit the smooth gradients seen in the embryo (Figure 5.7). The parameter sets obtained using 200 AGETs are much more consistent: *tbxta* expression gradually decreases from posterior to anterior, there is a distinct peak of *tbx16* expression, and *tbx6* gradually increases expression from posterior to anterior. Patterning from parameter sets obtained using 50 and 100 AGETs are less consistent than those obtained using 200 AGETs. While some parameter sets obtained using 50 or 100

AGETs are able to closely reproduce correct patterning in the PSM, others are not. For example the decrease in *tbxta* and *tbx16* expression in the PSM is too sharp for the MAP networks from Random Runs 4 and 5 (reverse engineered using 100 AGETs) (Figure 5.7). This shows that increasing the number of AGETs used for reverse engineering leads to better and more consistent results.

I chose to use 200 AGETs for reverse-engineering. Using 200 AGETs for reverse-engineering led to significantly better results than using 10, 50, or 100 (Figures 5.6, 5.7). While using 745 AGETs for reverse-engineering did improve the overall likelihood score for obtained parameter sets, the improvement was marginal, and computation was unfeasibly slow (Figure 5.6). In addition, using a smaller proportion of AGETs for reverse-engineering left more AGETs to use for validation and further study. This meant that I could more robustly evaluate the quality of the parameter sets obtained. Therefore, all following experiments have been done on parameter sets obtained from reverse-engineering using 200 AGETs.

5.2.3 Network Validation

Having established an effective MCMC procedure to reverse-engineer parameter sets to 200 AGETs, I then validated parameter sets, by investigating whether they met the following conditions:

1. Do gene expression trajectories simulated from different parameter sets match the corresponding AGETs?

2. Can the simulated trajectories reproduce gene expression patterning across the PSM?

3. Can parameter sets recapitulate known genetic interactions;

4. Do parameter sets recapitulate response to perturbations; and

5. Can simulations from these parameter sets autonomously upregulate *tbx6*; as demonstrated by Fulton, 2021 in Ben Steventon's lab.

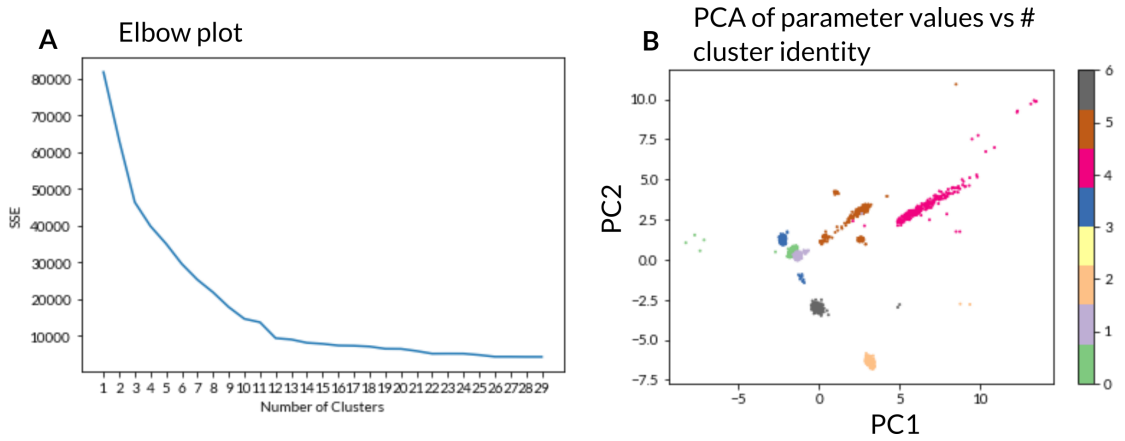


Figure 5.8: GRN Clustering. A. Elbow plot of SSE vs cluster size. B. PCA of parameter values, colour coded by cluster

To obtain candidate networks for future study, I took the parameter sets from the final iteration of MCMC. Because `emcee` tends to produce a small number of poorly fitting walkers each run, I computed the likelihood score for all AGETs for these parameter sets, and removed poorly fitting parameter sets (likelihood scores less than -1500). I also removed parameter sets with unrealistic parameter values ($> \pm 100$). This left 3600 parameter sets. I then clustered networks into 13 clusters, using k-means clustering (the number of clusters was determined using an elbow plot, Figure 5.8), ignored all clusters representing less than 5% of all parameter sets, and then took the median parameter set from each cluster as a cluster representative. This produced seven clusters, numbered 1-7, representing the variety of parameter sets obtained during reverse engineering.

Here I will describe validation of, firstly, the best fitting network ("maximum a posteriori", or MAP, network), and secondly, representatives of the seven clusters obtained.

5.2.4 Parameter sets reproduce gene expression at the single cell level

The MAP network is defined as the network with the best likelihood score, when the likelihood score is computed based on all AGETs. I first investigated the ability of the MAP network to recapitulate gene expression at the single-cell level, by investigating whether simulated gene expression trajectories match AGET

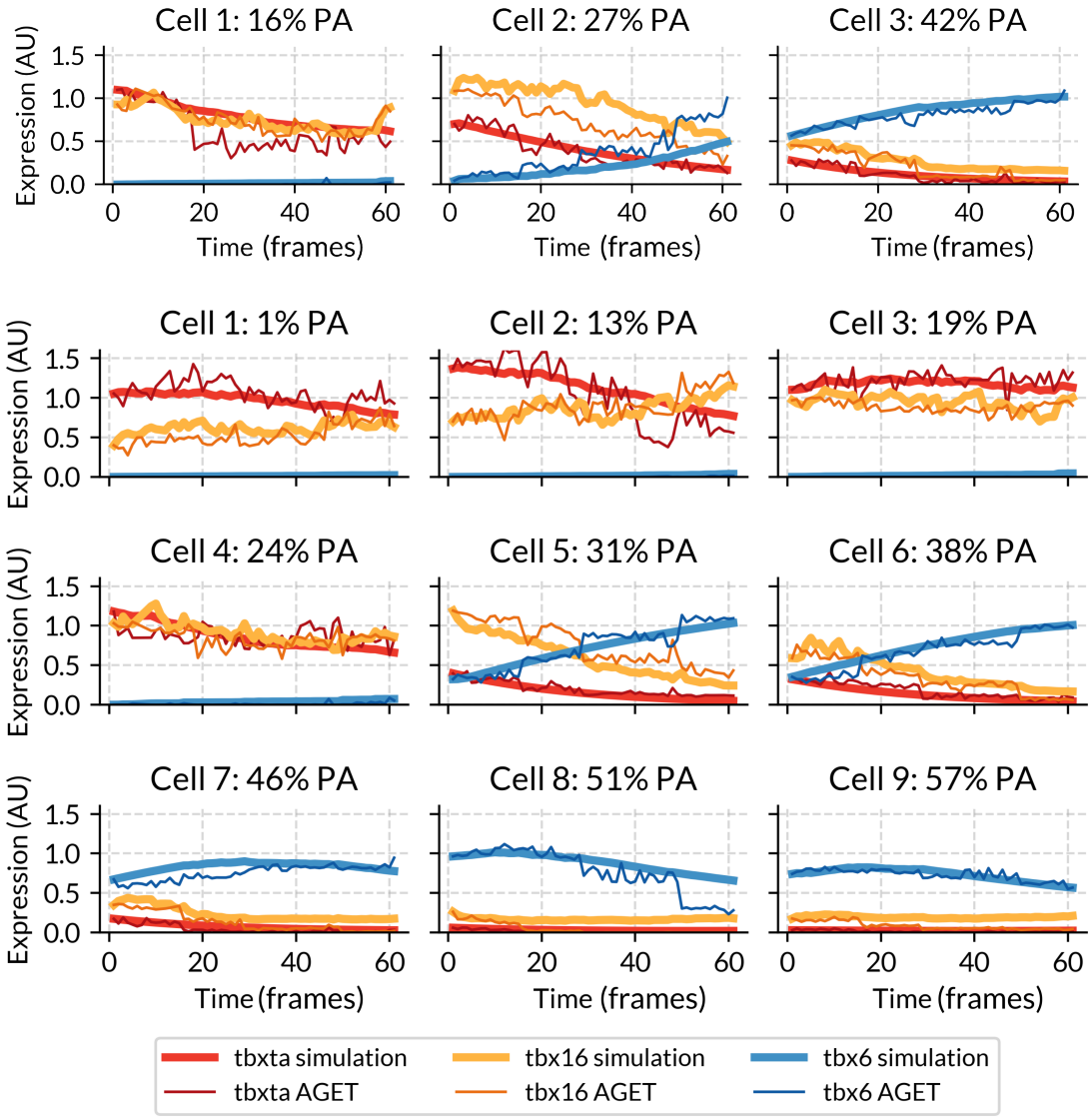


Figure 5.9: Simulated gene expression from MAP network is very close to AGETs. Top row: AGETs used for reverse engineering. These correspond to cells at different posterior-anterior (PA) positions in the embryo, as labelled in the subplot title. Bottom rows: AGETs not used for reverse engineering. Thick lines represent simulated gene expression and thin lines represent assigned gene expression (AGETs).

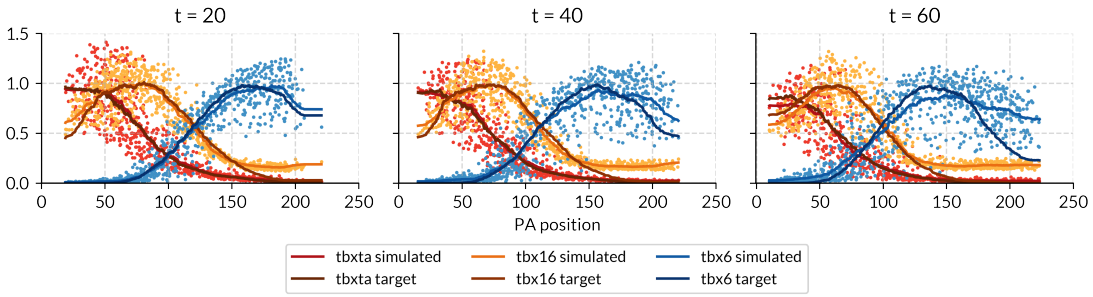


Figure 5.10: For the MAP network, simulated gene expressions across the embryo closely recapitulate the known gene expression patterns. Simulation is shown for different timepoints (20, 40, and 60AU). Dark lines represent a rolling average for the assigned/target/AGET gene expression, light lines represent rolling average of the simulated gene expression, and individual points represent individual cells.

gene expression. The MAP network is able to recapitulate the gene expression dynamics of AGETs used for fitting, and those not used for fitting: simulated gene expression trajectories track the overall AGET expression, though they do not recapitulate the fluctuations in gene expression well (Figure 5.9). This inability to reproduce fluctuations in gene expression trajectories is unsurprising, as the modelling framework used is deterministic. Altogether this demonstrates that the MAP network, as expected, is able to reproduce the target/AGET gene expression trajectories, even for AGETs not used for reverse engineering.

5.2.5 Parameter sets recapitulate tissue-level gene expression patterning

Having shown that the MAP parameter set recapitulates target gene expression at the single cell level, I investigated the ability of this parameter set to recapitulate gene expression pattern at the tissue level. As tissue-level pattern was not explicitly used for reverse-engineering, this represents an independent validation of the reverse-engineering methodology.

The MAP network is able to closely reproduce gene expression patterning at the tissue level. I measured the rolling average of simulated, and target, gene expression for all genes, and found that these match closely at different timepoints (Figure 5.10). Subtle differences in gene expression do exist: specifically, the simulated *tbxta* peak is slightly lower than target gene expression, the lowest value of

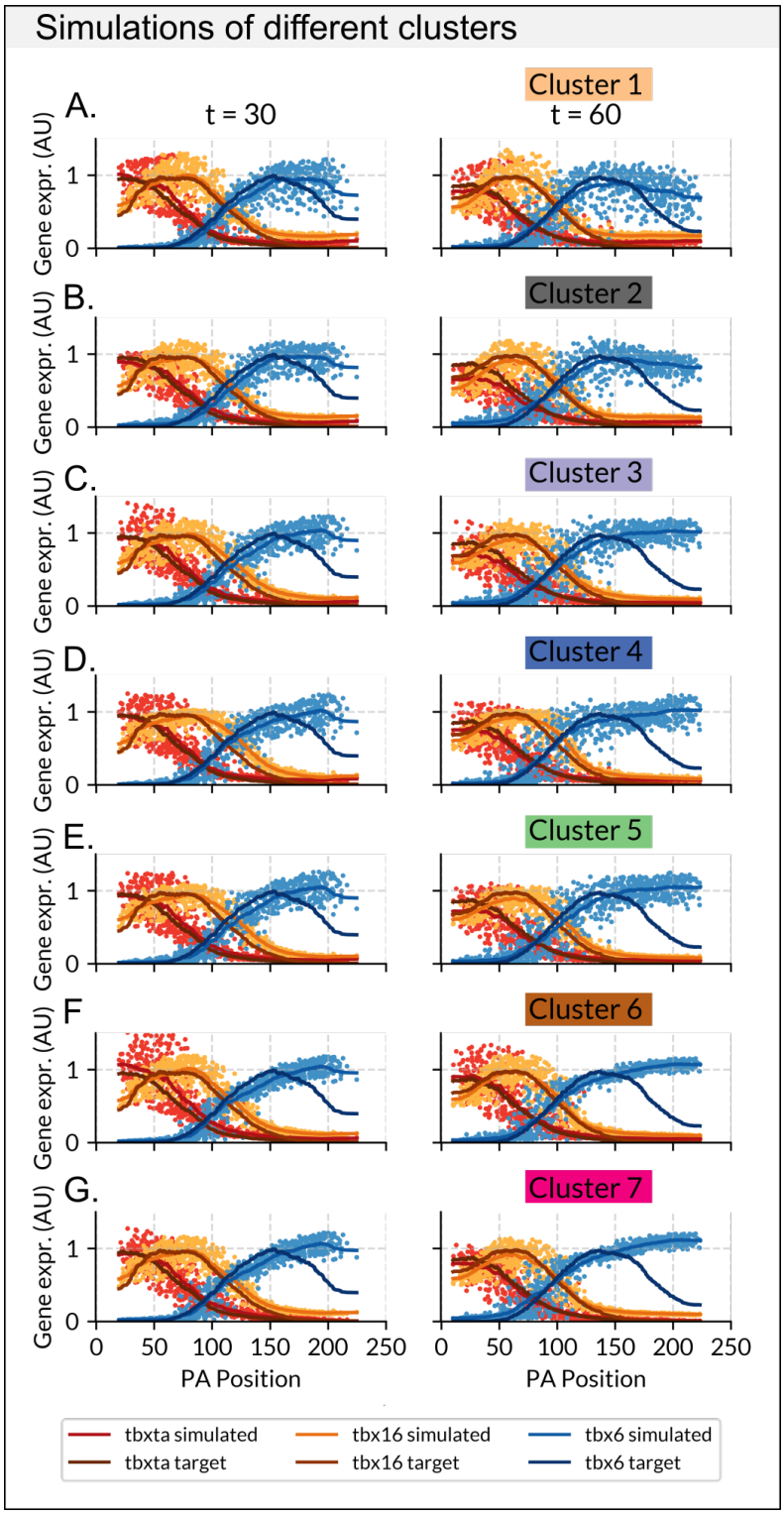


Figure 5.11: Other clustered parameter values recapitulate pattern well. Simulation is shown for different timepoints (30 and 60AU). Dark lines represent a rolling average the AGET gene expression, light lines represent rolling average of the simulated gene expression, and individual points represent individual cells.

simulated *tbx16* expression is slightly above zero, and *tbx6* is not downregulated in the anterior (Figure 5.10). The lack of anterior *tbx6* downregulation is expected, as I intentionally did not use AGETs from the anterior of the PSM for fitting. This was because *tbx6* is expression is inhibited by *ripply* in the anterior PSM, a gene not included in our formulation (Kawamura et al., 2005; Windner et al., 2015; Venzin et al., 2023). Altogether, then, the MAP network very closely recapitulates patterning at the tissue level.

Cluster representatives also reproduce tbox spatio-temporal gene expression patterns closely (Figure 5.11). As for the MAP network and in the PSM, for all clusters, *tbxta* is expressed in a posterior-to-anterior gradient (Figure 5.11). Expression of *tbx16* peaks at about 60μm from the posterior of the PSM, before decreasing in expression, while *tbx6* peaks at around 175μm from the posterior tip of the PSM (Figure 5.11). As with the MAP network and as expected, most cluster representatives do not reproduce the downregulation of *tbx6* in the anterior PSM observed in the embryo (Figure 5.11). There are subtle differences in gene expression between clusters: for example expression of *tbx6* is less heterogeneous in Clusters six and seven than in other clusters and in the MAP network (Figure 5.11). This indicates that parameter sets may use different mechanisms to achieve patterning. Overall, however, all clusters very closely reproduce the gene expression pattern of tbox genes across the PSM, showing that this reverse-engineering methodology is able to produce parameter sets reproducing tbox gene expression pattern in the PSM.

5.2.6 Parameter sets predict known genetic interactions

Having shown that parameter sets recapitulate gene expression well at the single cell and tissue level, I then investigated how well our parameter values agree with genetic interactions known from the literature.

For instance, in zebrafish, Wnt directly activates *tbx16* via a Wnt-binding promoter (Bouldin et al., 2015). Despite not being used as a prior in MCMC, this result is recapitulated in our filtered parameter sets, where all of our parameter

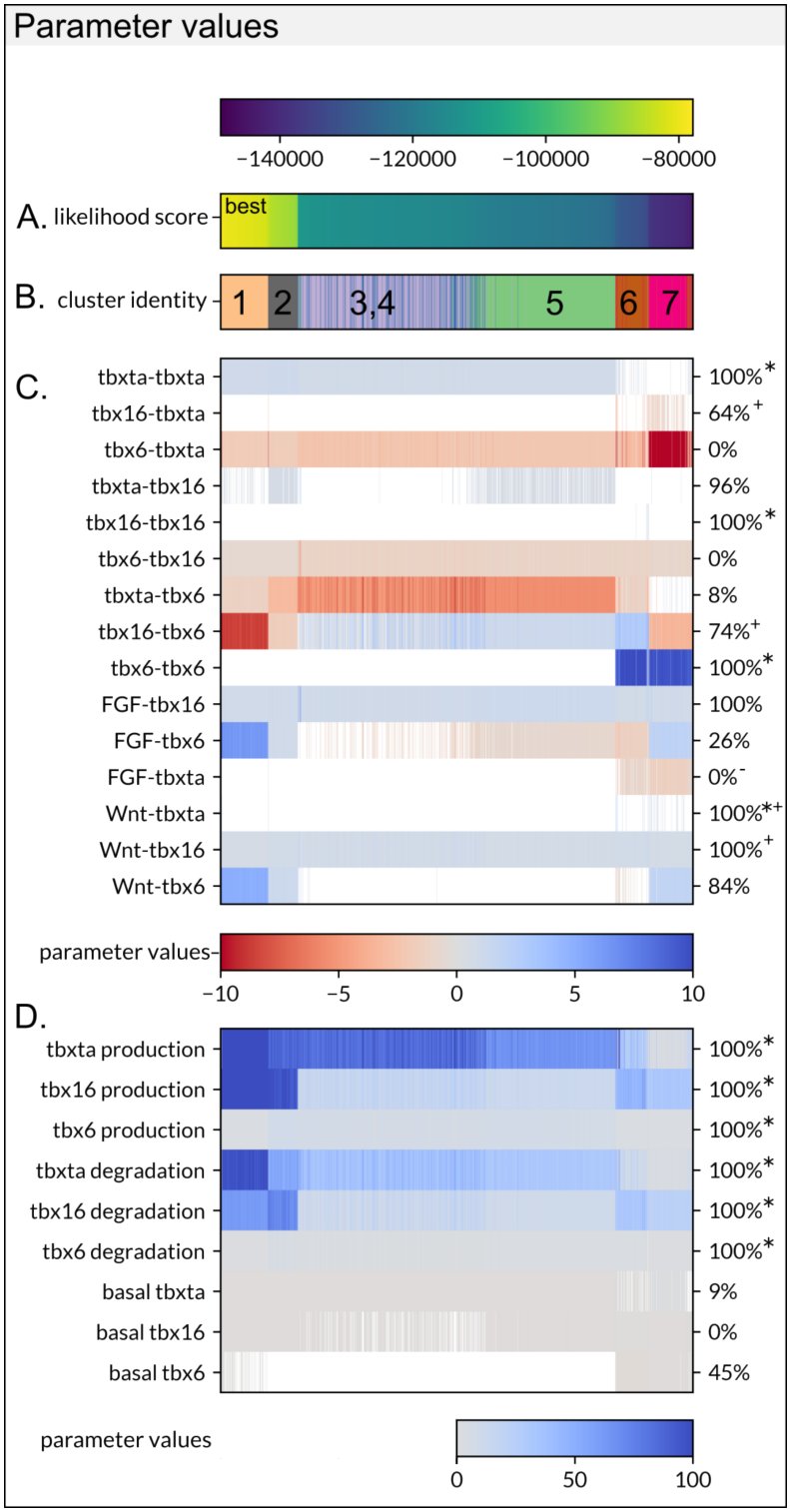


Figure 5.12: Heatmap of parameter values, ordered by likelihood score. A: likelihood score, as measured across all AGETs. B: cluster identify. C: parameter values of interactions, cropped to ± 10. D: parameter values for production, degradation, and basal production rates, cropped to $0 - 100$. Parameters indicated by an asterisk () are positive priors. Parameters indicated by a + or - sign are thought to be positive or negative (respectively) from the literature. White depicts parameters very close to zero (less extreme than ± 0.5)

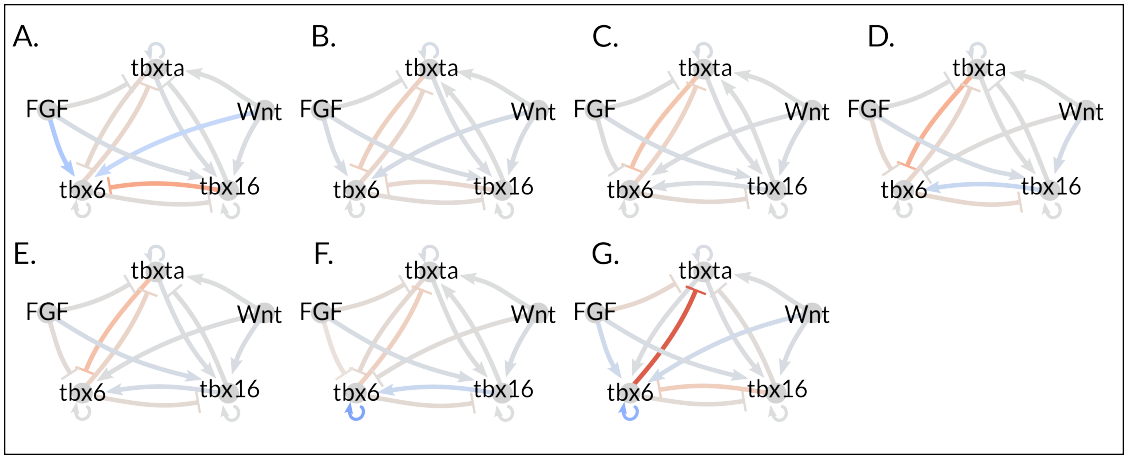


Figure 5.13: GRN topology for different clusters. A-F: Network topology for different clusters, 1-7, alphabetical order. Blue arrows represent positive (activation) interactions, and red bars represent negative (inhibitory) interactions. The darkness of colour represents the strength of interaction.

sets have a positive interaction between Wnt and *tbx16* (ie, Wnt activates *tbx16*) (Figure 5.12). *tbx16* is also known to activate *tbx6*: upregulating *tbx16* leads to an increase in *tbx6* expression, and *tbx16* mutants have a loss of *tbx6* expression (Fior et al., 2012; Bouldin et al., 2015). In our networks, 74% have *tbx16* upregulating *tbx6* and so are consistent with the literature; though the best fitting two clusters do predict this interaction as negative (Figure 5.12). *tbx16* is also known to repress *tbxta* indirectly, via Wnt (Bouldin et al., 2015). We find that this parameter is negative (correct) in 64% of our networks (Figure 5.12). FGF inhibits *tbxta*, as loss of FGF causes an expansion of *tbxta* (Goto et al., 2017). All of our parameter sets predict this interaction as negative, consistent with the literature (Figure 5.12). Finally, FGF is thought to activate *tbx16*, as dominant negative FGFR1 embryos display a loss of *tbx16* expression (Goto et al., 2017). Again, in all of our parameter sets FGF is activating *tbx16* (Figure 5.12). Altogether, there is a high level of consistency between our parameter values and the sign of genetic interactions known from the literature. This means that we can have a high level of confidence in the reverse-engineering procedure. Network topologies for cluster representatives are summarised in Figure 5.13.

The obtained parameter values can also be used to predict novel genetic

interactions that have not been investigated experimentally. For example, *tbx6* is always predicted to be inhibiting *tbxta*, which would be a convenient way to restrict *tbxta* expression to the posterior. *tbxta* also activates *tbx16* expression in 96% of networks; *tbx6* always represses *tbx16*; and Wnt activates *tbx6* in 85% of networks (Figure 5.12). If these interactions were identified in the embryo under further testing, it would further validate the reverse-engineering method.

5.2.7 Parameter sets recapitulate some genetic perturbations

Having shown that the sign of parameter values is often consistent with experimentally determined genetic interactions, I then investigated whether our parameter sets can recapitulate the results of experimental perturbations. I therefore attempted to reproduce *in silico* the effect of perturbing FGF and Wnt signalling.

Bouldin et al., 2015 over-express Wnt in the zebrafish tailbud by over-expressing beta-catenin under a heat-shock promoter. Four hours post heat-shock, the *tbx16* expression domain is expanded toward the anterior PSM. I replicated this experiment *in silico* using the MAP network by setting Wnt expression to 1.5 (about the maximum Wnt values in the AGETs) for the duration of the simulation. I found that at the final time-point of simulation *tbx16* is up-regulated in the anterior PSM, as in the experimental data (Figure 5.14A, B). However, *tbxta* is downregulated in the simulation, in contrast to the upregulation reported by Bouldin et al., 2015 (Figure 5.14A, B). This could be because our model formulation does not include the known feedback loop between Wnt and *tbxta*. Our other parameter sets also recapitulate this result, with the vast majority predicting an average increase in *tbx16* expression and decrease in *tbxta* expression (Figure 5.14A, B).

Goto et al., 2017 downregulate FGF expression by inducing a heat-shock dominant negative form of FGF at 8 ss, resulting in a posterior expansion of *tbxta* expression. I reproduced this experiment by setting FGF to 0.01AU (of a minimum value of 0AU) in every cell and simulating as before using the MAP network. In this simulation, *tbxta* is significantly up-regulated after the

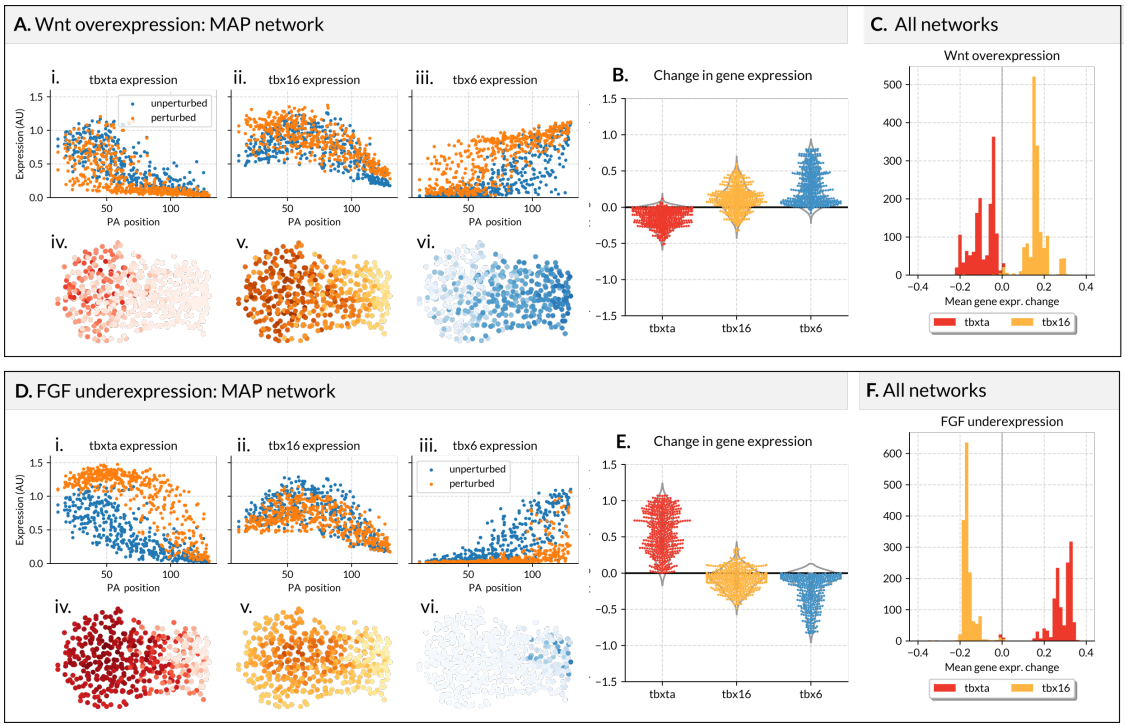


Figure 5.14: Over- and under- expression of FGF and Wnt is able to recapitulate patterns seen in the literature. A. Results of simulating Wnt over-expression using the MAP network. A Ai. *tbxta*, Aii. *tbx16*, and Aiii. *tbx6* expression in unperturbed (blue) and perturbed (orange) simulations using the MAP network. Each dot represents the expression level in a cell at the last time point of the simulation. Aiv. *tbxta* (red), Av. *tbx16* (yellow), and Avi. *tbx6* (blue) expression shown in cells within the PSM. Posterior left, dorsal up. B. Difference in gene expression between the unperturbed and perturbed simulations calculated for every cell at the last time point of the simulations. C. The mean gene expression difference in single cells between perturbed and unperturbed simulations, calculated for every network and displayed as a histogram for *tbxta* (red) and *tbx16* (yellow). F. Results of simulating FGF under-expression, depicted as for (A). E. Difference in gene expression between the unperturbed and perturbed simulations calculated for every cell at the last time point of the simulations. F. The mean gene expression difference in single cells between perturbed and unperturbed simulations, calculated for every network and displayed as a histogram for *tbxta* (red) and *tbx16* (yellow).

downregulation of FGF, in agreement with the experimental results. *tbx16* is slightly down-regulated, and *tbx6* is significantly down-regulated (Figure 5.14). When I repeated this analysis on all successful parameter sets, a majority also recapitulate an increase in *tbxta* expression.

Altogether, this shows that that our reverse-engineered networks are also capable of reproducing the results of experimental perturbations, namely overexpression of Wnt signalling, and loss of FGF. These data were not used for reverse-engineering, further reinforcing my confidence in the methodology developed here, and resulting parameter sets.

5.2.8 Parameter sets can reproduce upregulation of *tbx6 in silico*

Finally, I investigated whether our reverse-engineered networks are capable of autonomously upregulating *tbx16* expression when removed from the tailbud. Fulton, 2021, in Ben Steventon's lab, shows that when cells are dissected from the posterior zebrafish tailbud and cultured individually without any added FGF or Wnt signalling, a two distinct cell behaviours emerge: one population of cells upregulates *tbx6*, and the other does not (Fulton, 2021). I wished to know if our networks would produce this behaviour. To investigate this, I took cells from the posterior 20% of the PSM, fixed their signalling to that of the initial timepoint, and simulated gene expression *in silico* for the equivalent of six hours. The results of these simulations are displayed as categorical scatterplots, with timepoint (in hours) on the X axis, and the Y axis denoting the level of *tbx6* expression. If our results precisely match those of Fulton, 2021, we would expect a bimodal distribution of cells at the 6hr timepoint.

For the MAP network, one population of cells autonomously upregulates *tbx6* expression, while another maintains low *tbx6* expression for at least six hours (Figure 5.15A). The MAP network thus recapitulates the behaviour seen in the embryo. Four of the seven cluster representatives are also able to produce this behaviour: for clusters 1, 3, 4, and 6, after six simulated hours, there is a clear population of cells with low (<0.1) and high (>1.0) *tbx6* expression (Figure

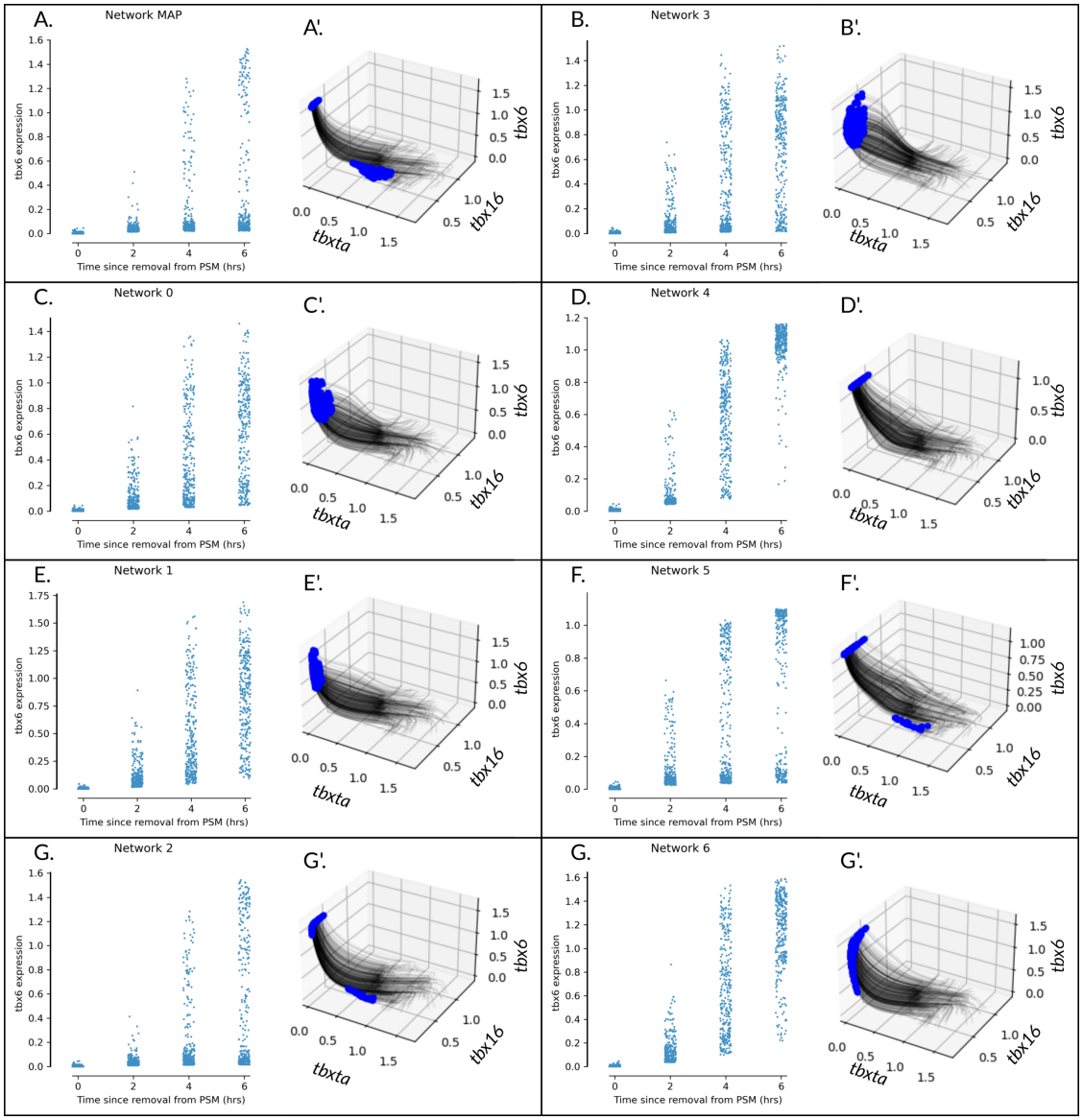


Figure 5.15: Some parameter sets can upregulate *tbx6 in silico*. A-H: *tbx6* expression after cells from the posterior 20% of the embryo are simulated for the equivalent of 0-6hours. Most parameter sets (A, B, C, E, F, G, and H) have cells at both low and high *tbx6*, as expected. A'-G': Simulated long term trajectories of cells from A-H. Grey lines represent the simulated trajectories, and blue dots show the final gene expression position.

5.15). The ability of the tbox network to achieve two distinct cellular behaviours was not used for reverse-engineering, meaning that this test provides strong validation for our reverse-engineering procedure.

5.3 Discussion

5.3.1 Parameter sets recapitulate key biological behaviours of the tbox GRN

My efforts to troubleshoot and validate the reverse-engineering aspect of Live Modelling produced parameter sets that could reproduce multiple different biological functions of the Tbox GRN: gene expression pattern, aspects of the network topology, response to signalling perturbation, and ability to produce two different populations of cells expressing *tbx6* in the absence of signalling variation.

The gene expression pattern is the easiest feature to reproduce: all clusters of parameter values obtained using 200 AGETs are able to simulate the embryonic patterning very closely, and even some networks reverse-engineered using smaller numbers of AGETs can faithfully recapitulate gene expression pattern onto the embryo.

I was also able to recapitulate known genetic interactions in the tbox network. Of our clustered, filtered parameter sets, all predict the known genetic interactions between Wnt and *tbxta*, Wnt and *tbx16*, FGF and *tbxta*, as well as FGF and *tbx16*. In addition, 74% of parameter sets correctly predict that Wnt activates *tbx16*. *tbx16* is known to activate *tbxta* via Wnt; 64% of parameter sets predict this interaction is positive. The sign of these genetic interactions was not used for reverse-engineering. Therefore, the fact that so many are recovered during reverse-engineering implies that we have an effective and robust reverse-engineering methodology.

I also tested whether these networks could reproduce the results of signalling perturbations. This was largely successful: the MAP network is able to recapitulate upregulation of *tbx16* following Wnt overexpression (Bouldin et al., 2015),

though in our simulations *tbxta* is downregulated, not upregulated as in the embryo. This discrepancy may be because our model does not recapitulate *tbxta* feedback onto Wnt (Goering et al., 2003; Martin & Kimelman, 2008). The parameter sets also predict the correct response to FGF under-expression: loss of FGF leads to *tbxta* upregulation in the embryo and in our models (Goto et al., 2017). Thus, parameter sets can again recapitulate experimental predictions.

Finally, I investigated whether the parameter sets could recapitulate another experimental finding: the ability of cells to upregulate or not upregulate *tbx6* when dissected from the tailbud. Four of our seven clusters reproduce this behaviour, as does our MAP network.

In summary, the networks I reverse-engineered are capable of recapitulating many experimental features of the zebrafish tbox GRN. This suggests that our model parameters, and the reverse-engineering methodology, are robust.

5.3.2 "Live modelling" allows one to investigate patterning in a tissue undergoing morphogenesis

"Live modelling" provides a toolkit to understand how patterning can occur in a tissue undergoing morphogenesis and cell movement, presenting a method to:

1. Infer gene expression trajectories from cell movement trajectories and static gene expression patterns
2. Reverse-engineer parameter sets from these trajectories
3. Simulate GRNs onto cell movement trajectories; and
4. Validate the resulting networks.

For systems where gene expression in an individual cell is well-approximated by that cell's position, such as the zebrafish tbox GRN, this methodology allows us to infer, simulate, and then study GRNs in systems where cells are moving during patterning. Ultimately this enables the study of GRNs and morphogenesis and how they interact.

This technique could easily be expanded to other questions and model organisms. It is already being adapted to understand interactions between tissue growth and GRNs in the mouse limb bud (Aviñó-Esteban & Sharpe, 2024). It could also be used, for example, to simulate segment formation in the beetle (*Tribolium castaneum*): segmentation in beetles is proposed, like vertebrates, to involve a segmentation-clock like GRN, however there is no known mechanism that synchronizes gene expression between cells (Peel et al., 2005; Clark & Peel, 2018; Clark et al., 2019). Live modelling could be used to establish whether such a mechanism is required for segmentation in *Tribolium*, or whether there is sufficient coherence in cell movements that synchronization is not required. It would also be very interesting to infer and simulate the GRN governing NMP cell fate choice, in both zebrafish and other vertebrates (Sambasivan & Steventon, 2021; Toh et al., 2022). There are therefore many potential applications for the Live Modelling framework.

Live Modelling could also be adapted for biological systems where gene expression patterns are not stable over time. This occurs in, for example, the cichlid tbox genes (see Chapter 3). This phenomenon would be relatively straightforward to reverse-engineer using the Live Modelling methodology; one would simply need to adapt the inference methodology to use different, stage matched, embryos at different timepoints, instead using the same pattern at each time point to generate AGETs.

Live Modelling relies on one crucial assumption: that the gene expression within an individual cell can be assumed based on that cell's position within the tissue. This assumption does not hold for all systems. For example, in some systems, there is significant spatial heterogeneity within a tissue. For example, in vertebrate NMPs, the relative levels of *tbxta*, *sox2*, and *tbx16* are heterogenous (Attardi et al., 2018; Toh et al., 2022; Saunders et al., 2025). So-called "salt and pepper" expression patterns occur in other biological systems (eg. Fischer et al., 2023; Hilbrant et al., 2012). In these systems exhibiting spatial heterogeneity, because there is no spatial order it would not be possible to generate AGETs

to use for reverse-engineering. However, the simulation component of Live Modelling could still be used to simulate GRNs generated by other means (eg. hand-tuning), and the resulting gene expression patterns analysed qualitatively. This approach could investigate whether stochasticity is required to maintain seemingly random gene expression patterns, or whether these patterns come about due to cell movement or some other mechanism.

Other possibilities for the Live Modelling framework include cross species simulation: where a reverse-engineered GRN from one species (e.g. zebrafish) is simulated on to cell tracks from another species (e.g. cichlids), to investigate whether any changes to one species' GRN is required to produce divergent patterning outcomes. Alternatively *in silico* cell tracks could be manipulated to study how cell movements, combined with a known GRN, impact pattern formation and maintenance.

Altogether, then, there are many potential uses for Live Modelling, beyond the present study.

5.3.3 Future work to model the tbox GRN in zebrafish

The core assumption of this method, that I was unable to fully test, is that our AGETs are able to recapitulate the diversity of gene expression trajectories within the developing embryo itself. It is possible that in the developing embryo, some cells exhibit aberrant/heterogenous expression patterns that are not captured by AGETs. For example, there are oscillations in the anterior boundary of *tbx6* expression (Ban et al., 2019; Venzin et al., 2023); as these are not detectable in static gene expression profiles, they are impossible to capture using the current Live Modelling methodology. Future work should address this by, where possible, comparing inferred AGETs with gene expression profiles obtained from tracking individual cells expressing genes of interest. However, such datasets are difficult to produce, are not yet possible for all genes in the zebrafish tbox GRN, and are likely to ever be feasible in non-model species (eg. cichlids).

Therefore, I extensively tested and validated the parameter sets generated here, as described above.

Ultimately, a full understanding of how cell movements and GRNs interact to achieve morphogenesis and patterning in the zebrafish will require a model simulating cell movements, signalling, and the GRN simultaneously, allowing a full and robust study of how these systems co-regulate each other. Live Modelling presents one step towards such a model. The methodology assumes that signalling regulates the tbox GRN (and not the other way around), and that the GRN does not impact morphogenesis. We know this is an oversimplification: *tbxta* and *tbx16* regulate FGF and Wnt; and *tbx16* expression is known to be required for directed cell movement in the PSM (Manning & Kimelman, 2015; Das et al., 2017).

It would be relatively straightforward to extend the current GRN to include feedback onto the Wnt and FGF signalling profiles. This would require addition of 16 extra parameters (the production, degradation, and basal production rate for FGF and Wnt, as well as regulation of FGF and Wnt by the three tbox genes), and treating the signalling profiles as an output of the model, not an input. These 16 extra parameters would dramatically increase the parameter space to search, creating difficulties for reverse-engineering (Foreman-Mackey et al., 2013). It may therefore be necessary to constrain parameters, for example (e.g. by fixing the production, degradation, and basal production rates for all genes), and/or to use a longer live imaging dataset and therefore longer AGETs for reverse-engineering.

Incorporating the well-established influence of *tbx16* on cell movement would be far more complicated (Manning & Kimelman, 2015; Das et al., 2017). First, a realistic model of cell movement in the PSM would need to be constructed. Such models do exist in two and three dimensions but do not reproduce important elements of morphogenesis such as the shape of the PSM in three dimensions or its elongation (Lawton et al., 2013; Uriu et al., 2021; Hammond et al., 2025). This model would then need to be altered to include impact of *tbx16* on cell motility, before a reverse-engineered GRN could be incorporated into the model.

5.4 Conclusions

The results of this chapter establish a robust methodology to, firstly, infer gene expression onto moving cell tracks from using fixed and stained samples, secondly reverse-engineer robust parameter sets, and thirdly simulate these parameter sets onto said moving cell tracks. These parameter sets have been carefully validated, and are able to recapitulate gene expression pattern, tbox network topology, and two different sets of biological perturbations. Altogether, this establishes Live Modelling as a viable method to investigate how morphogenesis and GRNs interact to achieve patterning. The parameter sets generated in this chapter will be further analysed in the next chapter, to investigate the role of signalling and cell movements in achieve patterning in the zebrafish PSM.

Chapter 6: The role of signalling and GRN dynamics in pattern formation in the developing zebrafish PSM

Contents

6.1 Introduction . . . 126
6.1.1 The zebrafish tbox GRN achieves pattern formation alongside cell movement . . . 126
6.1.2 Using dynamical systems theory to study pattern formation and differentiation . . . 128
6.2 Results . . . 130
6.2.1 Signalling modifies speed of differentiation . . . 130
6.2.2 Change in signalling as cells transit PSM do not impact pattern formation. . . . 136
6.2.3 The MAP network is bistable, allowing different long-term behaviour of cells . . . 142
6.2.4 Multiple potential dynamical mechanisms in the tbox GRN . . . 146
6.3 Discussion . . . 148
6.3.1 Two dynamical mechanisms achieving Tbox patterning . . . 149
6.3.2 The Tbox GRN: a cell-autonomous timer of differentiation? . . . 151
6.3.3 Degeneracy in the genotype-phenotype map . . . 152
6.3.4 Is positional information required for pattern maintenance in the tbox GRN? . . . 153
6.4 Conclusions . . . 154

6.1 Introduction

A major question in developmental biology is how gene expression patterns are formed and maintained in developing tissues. Pattern formation is usually studied in systems without extensive cell movement, such as the *Drosophila* blastoderm or the vertebrate neural tube (Jaeger, 2011; Fulton et al., 2022a; Marković et al., 2024). However, embryonic development also requires extensive cell movements and rearrangements, for example to establish the initial body axis in zebrafish, or form morphological somites (Kanki & Ho, 1997; Shah et al., 2019; Fulton et al., 2022a). A major technical challenge for developmental biology is therefore to develop a methodology to reverse-engineer gene regulatory networks achieving pattern formation alongside cell movement. In Chapter 5, I developed and validated such a methodology. I reverse-engineered robust parameter sets that could faithfully reproduce both pattern formation, known genetic interactions, and known functions of the tbox GRN, alongside cell movements in the zebrafish PSM. In this chapter, I will further study these parameter sets, analysing their function, and how they are able to reproduce gene expression patterning alongside cell movement.

6.1.1 The zebrafish tbox GRN achieves pattern formation alongside cell movement

The ability of the zebrafish tbox GRN to achieve pattern formation alongside cell movement makes it an interesting system to study the interaction of these processes. In this chapter, I will endeavour to explain two features of the tbox GRN: its ability to maintain coherent gene expression pattern within the PSM despite cells taking different amounts of time to transit the PSM, and autonomous upregulation of *tbx6* in cells removed from the tailbud (Fulton, 2021; Fulton et al., 2022b; Rohde et al., 2024).

Cells take different amounts of time to transit the zebrafish PSM depending on when they exit the tailbud (Fulton et al., 2022b). In addition, there are

extensive cell rearrangements in the tailbud, which in principle should introduce heterogeneity in the gene expression pattern (Kanki & Ho, 1997; Lawton et al., 2013; Thomson et al., 2021). Despite this variation, expression of *tbx16* and *tbx6* is strikingly coherent within the PSM (Fulton et al., 2022b). How gene expression patterning within the PSM is co-ordinated with cell movement is therefore a major question I will approach in this chapter.

Cells within the zebrafish PSM are also able to autonomously differentiate, in the absence of changes in signalling levels (Fulton, 2021; Rohde et al., 2024). That is, cells from the posterior PSM can autonomously upregulate *tbx6* and *mesp-ba* when dissected from the PSM and cultured individually *in vitro*, without supplementation of signalling molecules (Fulton, 2021; Rohde et al., 2024). This behaviour has been demonstrated independently by two different groups. Fulton, 2021 adapt the methodology of Webb et al., 2014, 2016 to investigate autonomous differentiation of cells in the zebrafish PSM. Cells are dissected from the posterior 20% of the tailbud, dissociated, and cultured individually for six hours, in the absence of additional signalling molecules (Fulton, 2021). Cells respond differently to this treatment: one population of cells upregulates *tbx6* and downregulates *tbx16* after six hours, while another cell population maintains low *tbx6* expression. Rohde et al., 2024 demonstrate a similar result: that cells from the posterior PSM, when cultured *in vitro*, will slow oscillations of the segmentation clock gene *Her1*, and eventually upregulate *mesp-ba*, a somite specification marker, that is activated by *tbx6* (Sawada et al., 2000; Cutty et al., 2012; Windner et al., 2015). This demonstrates that, without any positional information from changing signalling levels, cells in the zebrafish tbox PSM are able to autonomously differentiate into the somite fate (Fulton, 2021; Rohde et al., 2024). In this chapter I will investigate how cells in the PSM are able to achieve this.

6.1.2 Using dynamical systems theory to study pattern formation and differentiation

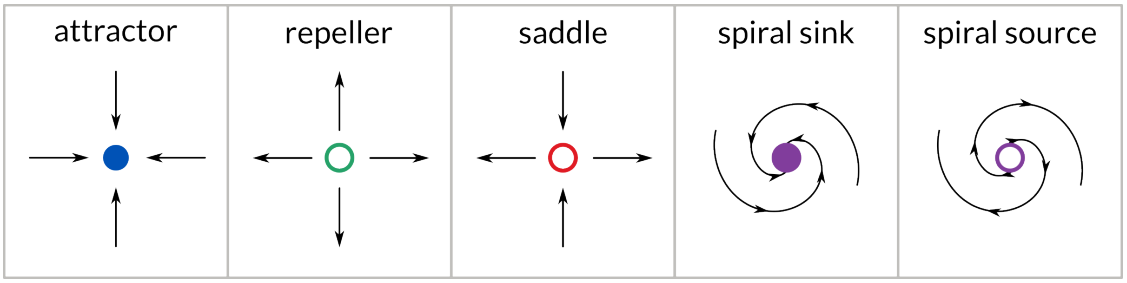


Figure 6.1: Types of fixed point in a dynamical system.

Dynamical systems theory (DST) provides a mathematical framework to study the function of GRNs during pattern formation and differentiation (Jaeger & Monk, 2014; Sáez et al., 2022a). Gene regulatory networks can be expressed as ordinary differential equations (ODEs), with parameters representing the strength of interactions between genes. These parameterized equations can then be studied by computing the phase portrait (Verd et al., 2014; Jaeger & Monk, 2014). Phase portraits consist of fixed points of gene expression, and trajectories towards these fixed points. Fixed points are positions in gene expression space where the rate of change is zero; meaning that if it is not perturbed, a cell with these gene expression values will not change its expression. There are different types of fixed points, summarized in Figure 6.1. Attractors draw trajectories towards them; meaning that a cell close to an attractor will move towards it. Repellers push trajectories away from them, and so are difficult to access in gene expression space. Saddle points draw trajectories towards them in one dimension, but not the other, and can separate pairs of attractors. Spiral sinks also attract trajectories, but non-linearly, and spiral repellers repel trajectories non-linearly (Strogatz, 2019). Finally, limit cycles are stable oscillatory gene expression trajectories that attract nearby trajectories and so also represent a different type of attractor.

Phase portraits can be interpreted in biological terms, with attractors or fixed points represent stable cell states or cell fates (Jaeger & Monk, 2014; Sáez et al., 2022a). Note that attractors can be created and destroyed during development, with changes in signalling values, just as certain cell states are only accessible

during particular developmental timepoints (Jaeger & Monk, 2014; Sáez et al., 2022a). This means that cell differentiation can be explained using dynamical systems theory. For example, Sáez et al., 2022b characterize cell fate choice during stem cell differentiation, identifying two different mechanisms by which cells achieve a binary cell fate choice. In this work, the transitions between different transcriptional states are modelled, not the GRN directly, abstracting the GRN itself away from the cell fate choice (Sáez et al., 2022b).

Dynamical systems theory has also been fruitful for understanding pattern formation in the gap GRN. For example, damped oscillations, driven by a spiral sink attractor in the posterior *Drosophila* blastoderm, are thought to drive the anterior shifts in gap gene expression in the *Drosophila* blastoderm (Verd et al., 2018). This therefore connects the dynamical function of a GRN with a biological phenotype (Verd et al., 2018).

These two examples illustrate the usefulness of the dynamical systems theory approach to understanding cell differentiation, and pattern formation in development. Cell differentiation can be explained by describing the attractors of a system, and how cells transition between those attractors (Sáez et al., 2022b, 2022a; Rand et al., 2021). Pattern formation can likewise be explained by describing trajectories towards biological attractors, and how these can vary with different signalling values (Verd et al., 2014; Jaeger & Monk, 2014; Sáez et al., 2022b).

In this chapter, I will investigate the parameter sets describing the tbox GRN that I reverse-engineered in Chapter 5. I will investigate, firstly, the role of signalling in controlling differentiation. Do the differing signalling levels that cells experience in the PSM predispose cells to differentiate? I will then investigate the impact of signalling on pattern formation: does perturbing signalling impact tbox pattern formation in the PSM? Finally, I will perform a phase portrait analysis of parameter sets, to better understand the behaviour of the tbox GRN.

6.2 Results

I have chosen two networks for further analysis: the best fitting (MAP) network, and Network 1. Both of these networks faithfully reproduce gene expression patterning onto the PSM after simulation (Figure 6.2). However, while the MAP network has the best scoring fit, it does not reproduce all genetic interactions: specifically, Bouldin et al., 2015 and Fior et al., 2012 show that *tbx16* mutants have lower levels of *tbx6* expression than in wildtype embryos, suggesting that *tbx16* activates *tbx6* expression. The MAP network predicts this as negative. I therefore also selected the best-fitting cluster which predicted this interaction as positive: Network 1 (Figure 6.2).

6.2.1 Signalling modifies speed of differentiation

I first investigated the impact of differing signalling profiles on cell differentiation. To do so, I compared cells moving anteriorly (e.g. their final position is more anterior than their initial position) with those moving posteriorly (their final position is more posterior than their initial position), from the most posterior 100μm of the PSM. These cell populations differ in signalling and gene expression profiles: cells moving anteriorly tend to have higher Wnt levels than those moving posteriorly, though both populations have similar FGF levels (Figure 6.3A,B). Cells also occupy different regions of gene expression space: while the relative levels of *tbxta* and *tbx16* are similar for both populations, there is a population of anteriorly-moving cells that already have increased *tbx6* levels (Figure 6.3C).

For both the MAP network and Cluster 1, I compared differentiation dynamics in cells moving anteriorly and posteriorly within the PSM. The gene expression dynamics and final gene expression levels differ between cells moving anteriorly and posteriorly. With regards to dynamics, far more cells moving anteriorly begin to upregulate *tbx6* and therefore differentiate (Figure 6.4). Specifically, for the MAP network, 60% of cells moving anteriorly have final *tbx6* expression greater than 0.1AU, compared to just 15% of cells moving posteriorly (Figure 6.4). For

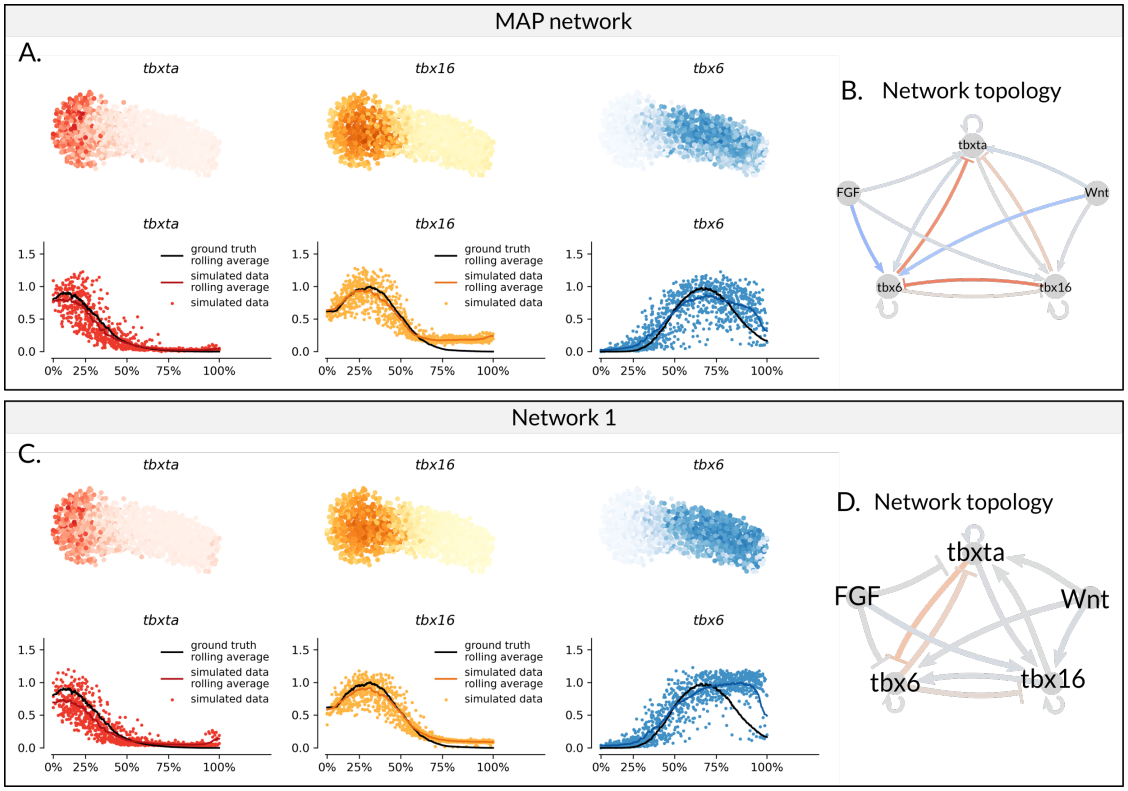


Figure 6.2: Gene expression pattern and network topology for best-fitting networks. A, C: Simulated gene expression pattern graphed onto the PSM (top line) or against AP position (bottom line), for the MAP network and Network 1. For the top panels, colour intensity represents the level of gene expression in individual cells. For the bottom panels, black lines indicate the rolling average gene expression for data used for reverse-engineering; coloured lines represent the rolling average of simulated gene expression, and individual points represent gene expression in individual cells. B, D: Network topology for the MAP network and Network 1. Blue arrows represent positive (activation) interactions, and red bars represent negative (inhibitory) interactions. The darkness of colour represents the strength of interaction.

Network 1, the same trend holds; 68% of cells moving anteriorly have begun to upregulate *tbx6*, compared to 19% of cells moving posteriorly (Figure 6.4). A threshold of 0.1AU of *tbx6* expression was chosen because, in the phase portraits of these networks, *tbx6* begins to appreciably upregulate at levels above 0.1AU (see below). Therefore *tbx6* differentiation appears to be irreversible at levels above 0.1AU. Overall final gene expression levels also differ between cells moving anteriorly and posteriorly: for both networks, *tbxta* is overall downregulated in cells moving anteriorly and *tbx6* is upregulated.

Note, however, that significant numbers of cells will maintain low *tbx6* in

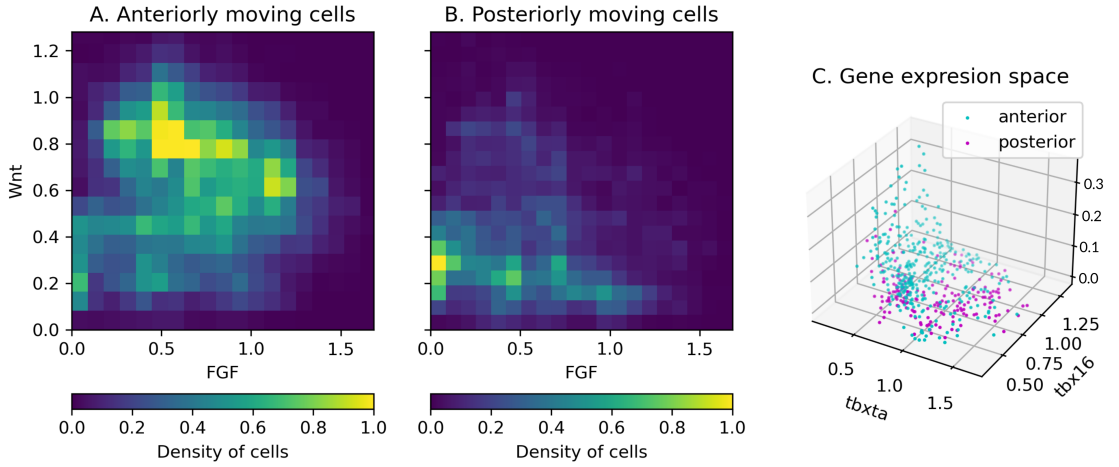


Figure 6.3: Cells moving anteriorly and posteriorly occupy different regions of signalling and gene expression space. A, B: two-dimensional histograms depicting the relative density of signalling values for cells moving anteriorly and posteriorly from the posterior 100μm of the PSM. The signalling profile from each cell, at each timepoint, was recorded and treated as an independent observation. Twenty bins were used for display. C: Relative initial expression of *tbxta*, *tbx16*, and *tbx6* for cells moving anteriorly (cyan) and posteriorly (magenta), in the posterior 100μm of the PSM.

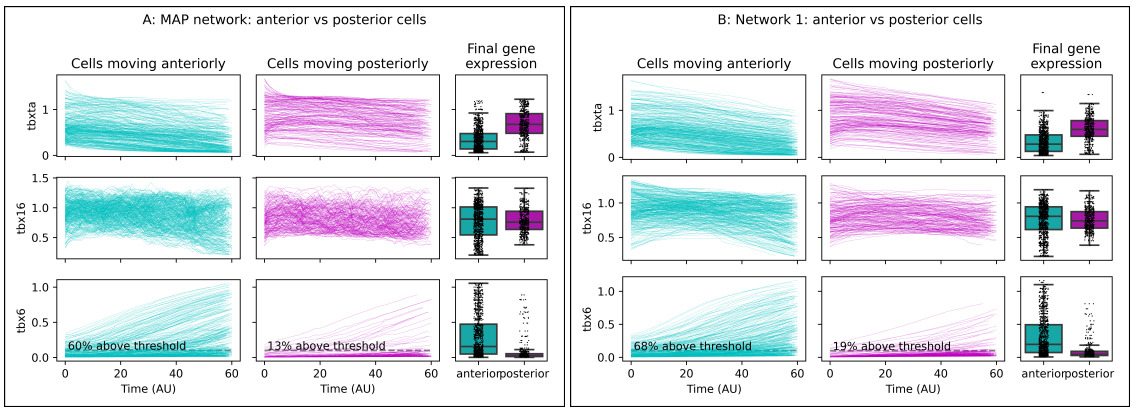


Figure 6.4: Cells moving anteriorly are move likely to differentiate than those moving posterior. A: simulation for MAP network. B: simulation for Network 1. Figure depicts gene expression trajectories for cells where their final timepoint is anterior to their initial timepoint ("Cells moving anteriorly"; left column), and cells where the final timepoint is posterior to their initial timepoint ("cells moving posteriorly"); middle column. Only cells beginning in the tailbud (posterior 100 μm) are presented. Right column presents a boxplot, and individual data points, at the final timepoint for cells that persisted for more than 50 timepoints. The boxpot depicts the four data data quartiles; whiskers represent the farthest data point lying within 1.5IQR of the median. Rows represent *tbxta*, *tbx16*, and *tbx6* expression in turn.

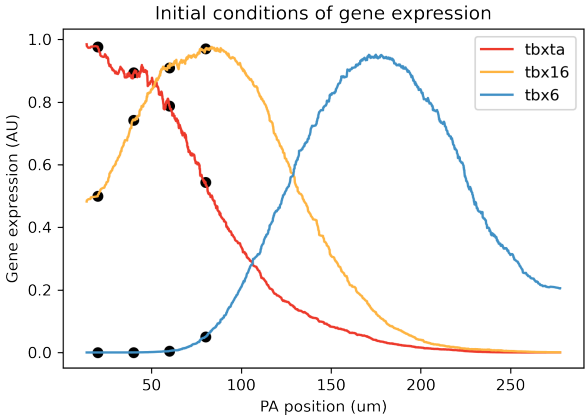


Figure 6.5: Initial gene expression used to control for gene expression heterogeneity. Coloured lines represent rolling average gene expression, averaged over 50 cells. Dots depict expression profiles used for simulations below. Red line: *tbxta*. Yellow line: *tbx16*. Blue line: *tbx6*.

both simulated cell populations, and some cells moving posteriorly will still upregulate *tbx6* (Figure 6.4). This means that while more anterior-moving signalling profiles may accelerate the speed of differentiation, cells are able to differentiate (upregulate *tbx6*) in a variety of different signalling contexts.

The initial conditions of gene expression could have a significant effect on the simulated gene expression dynamics. Next, therefore, I controlled for this variable, by providing the same initial gene expression to the signalling profiles from cells moving anteriorly or posteriorly. To identify relevant and representative initial conditions, I computed the rolling average of gene expression across the anterior-posterior axis. This represents an idealization of patterning across the AP axis and removes (biologically relevant) gene expression heterogeneity from the simulation (Figure 6.5). I then selected points from 20, 40, 60, and 80μm from the posterior tip of the tailbud, measured gene expression at these positions (Figure 6.5), and use these gene expression values as initial conditions to the simulations below.

I then repeated the simulations, this time using the same initial gene expression for each cell within the PSM, for the MAP network and Network 1. I used the four initial gene expression values from Figure 6.5 to initialize the simulations. Overall, for all of these initial conditions and for both networks tested, cells moving posteriorly are less likely to upregulate *tbx6* and begin differentiation.

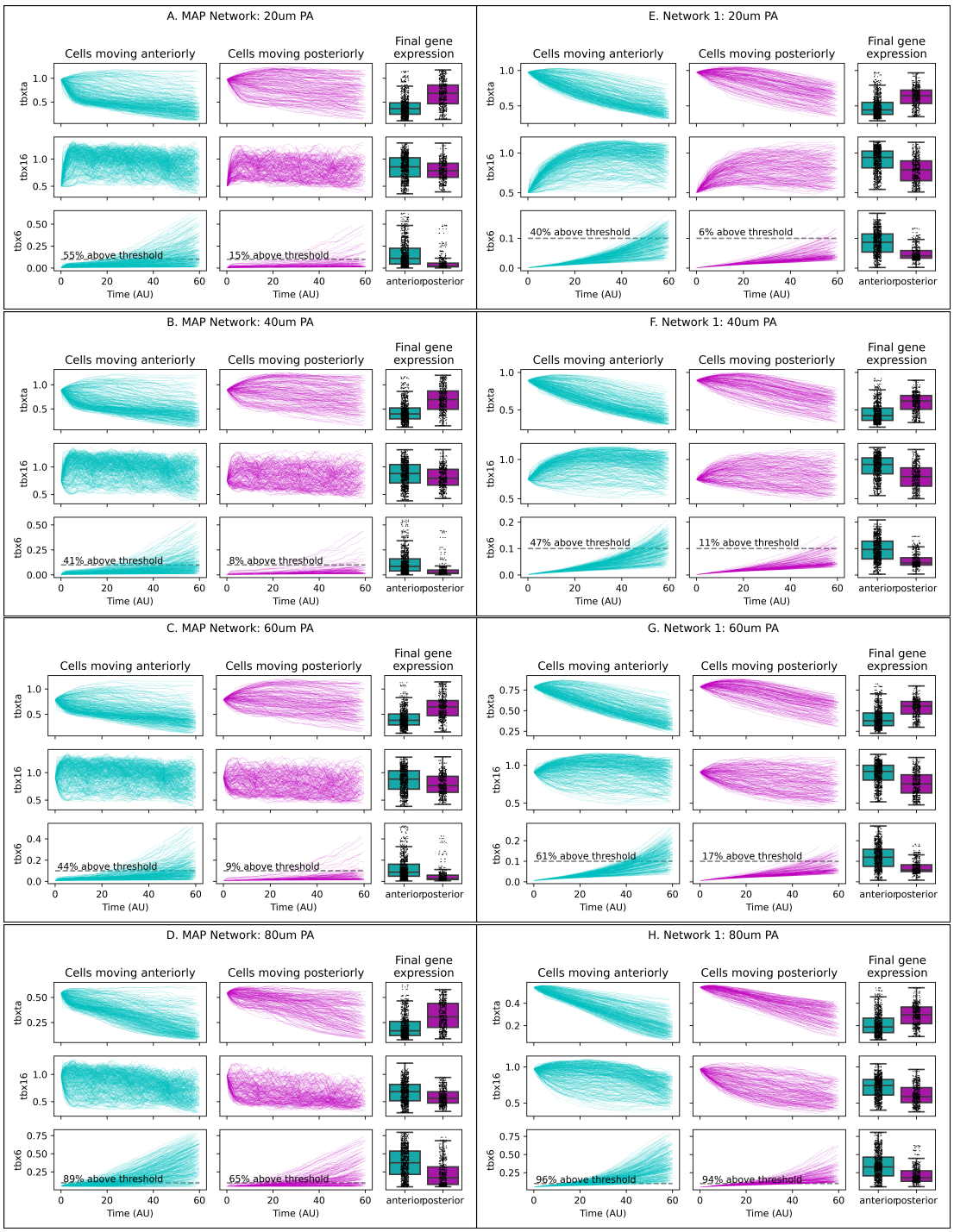


Figure 6.6: Cells moving anteriorly more likely to begin to differentiate, after correcting for initial conditions. A-D: Simulations using the MAP network, with initial conditions drawn from different AP positions (20, 40, 60, and 80μm from the posterior tip of the tailbud). E-H: Simulations using Network 1, with initial conditions drawn from different positions in the PSM as for A-D. Graphs are drawn as for Figure 6.4.

Table 6.1: Cells moving anteriorly are more likely to differentiate, even after controlling for initial conditions. The percentage of cells in a given condition, that increase *tbx6* above 0.1AU at the final timepoint of simulations, is given for each simulation. Note that because we are comparing different simulations, the total percentage in each row does not add to 100%.

	Percentage of cells upregulating *tbx6*			
	MAP network		Network 1	
Position (μm from tailbud)	Cells moving anterior	Cells moving posterior	Cells moving anterior	Cells moving posterior
20μm	55%	15%	40%	6%
40μm	41%	8%	47%	11%
60μm	44%	9%	61%	17%
80μm	89%	65%	96%	94%

Specifically, when both simulations are initialized with gene expression drawn from 20μm PA position, *tbxta* is downregulated in cells moving anteriorly compared to those moving posteriorly, and *tbx6* is upregulated, for both the MAP network and Network 1 (Figure 6.6A, E, Table 6.1). More cells also begin to differentiate: for the MAP network, 55% of cells moving anteriorly have begun to differentiate (defined as *tbx6* expression rising above 0.1AU by the final timepoint of the simulation), compared to only 15% of cells moving posteriorly (Figure 6.6A, E, Table 6.1). For network 1, 40% of cells moving anteriorly begin to differentiate, compared to 6% of cells moving posteriorly.

There are similar results for other initial conditions: for both parameter sets, cells moving anteriorly tend to downregulate *tbxta* and a number also begin to upregulate *tbx6* (Figure 6.6, Table 6.1). The number of cells upregulating *tbx6* does increase as initial conditions are drawn from more anterior positions within the PSM (Figure 6.6, Table 6.1). This shows that more anterior initial conditions make cells either more likely to differentiate, or able to differentiate faster. However, for all initial conditions, cells moving anteriorly are more likely to upregulate *tbx6* than those moving posteriorly, showing that the result above is not solely due to a bias in initial conditions.

Altogether, this shows that the signalling profiles cells experience as they move anteriorly make them more likely to begin to differentiate (begin to upreg-

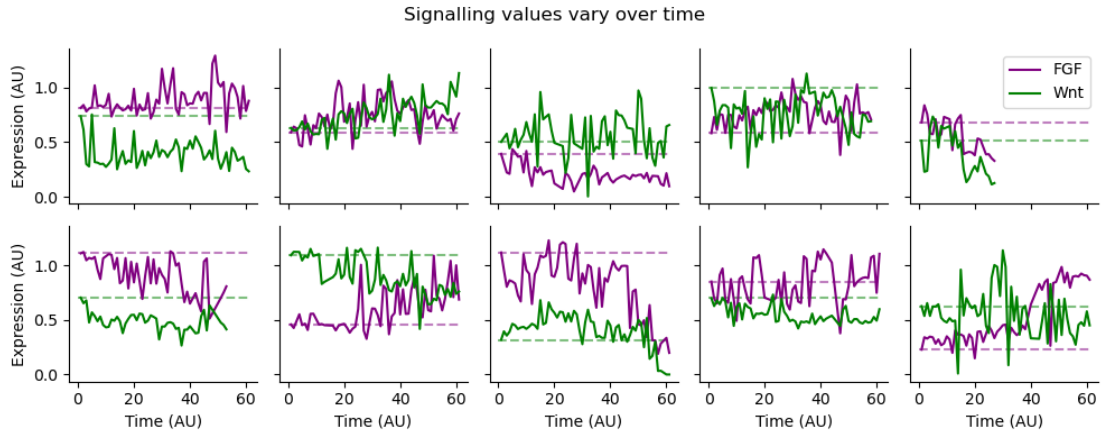


Figure 6.7: Cells experience different signalling values as they transit the PSM. Each panel depicts a different cell, drawn from the posterior 100μm of the PSM. Solid lines indicate approximated FGF and Wnt expression, and dashed lines indicate the replaced expression for simulations below.

ulate *tbx6*) than cells moving posteriorly. The initial gene expression of cells also makes them more likely to differentiate, with simulations with initial conditions from the anterior having more cells begin to differentiate. Therefore, both signalling, and a cells' initial gene expression, together determine whether a cell will differentiate.

6.2.2 Change in signalling as cells transit PSM do not impact pattern formation.

As cells transit the PSM, they experience changes in the level of FGF and Wnt signalling (Figure 6.7). Cells experience changes in the overall level of signalling: as cells transit the PSM, they tend to up- and then down- regulate FGF and Wnt signalling (Figure 6.7). There are also extensive fluctuations in signalling level from one timepoint to another (Figure 6.7). Such fluctuations are observed in the embryo: for example Erk levels oscillate within the zebrafish PSM (Simsek et al., 2023). I wanted to investigate how these changes in signalling values impact pattern formation. Note that global perturbations of signalling do significantly impact tbox gene expression pattern (Chapter 5); in this section I am interested in more subtle effects, namely how the changes to signalling profiles that cells experience as they transit the PSM contribute to pattern maintenance.

MAP network and Network 1

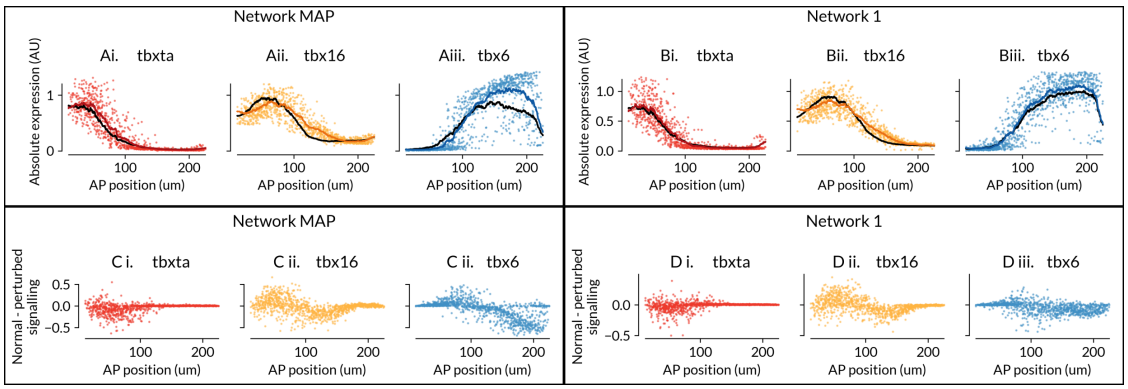


Figure 6.8: Gene expression patterning is robust to signalling fluctuations for the MAP network and Network 1. A: Gene expression at the final timepoint, when signalling is held constant for the duration of the simulation. The graph depicted represents gene expression at the final timepoint of simulation. Dots show expression in individual cells following perturbation; the black line represents the rolling average of wild type expression for the parameter set chosen (averaging over the closest 50 cells), and the dark coloured line represents rolling average of the perturbed simulation. B: Difference between wild type and perturbed gene expression at the single cell level.

I first devised an experiment to measure the impact of changes in signalling as cells transit the PSM. I run two simulations of pattern formation: one, with wild-type signalling, and one, where cells are unable to update their signalling profiles as they transit the PSM. That is, signalling for each cell is fixed to that of the initial timepoint (see Figure 6.7 for example gene expression trajectories). I call this the "static" simulation, and the simulation where cells do update their signalling profiles as they transit the PSM the "moving" simulation, after the impact on signalling trajectories.

This perturbation has a subtle effect on pattern formation. For *tbxta* expression, the rolling average of gene expression is almost identical (Figure 6.8). The simulated expression of *tbx16* is slightly upregulated in the anterior in the "static" simulation, though this is so subtle it is not obvious without quantitative comparison (Figure 6.8). Both the MAP network and Network 1 exhibit this effect (Figure 6.8). *tbx6* expression does differ between "static" and wild type simulations: in the "static" simulation, there is a sharper transition between the high and low *tbx6* expression states (Figure 6.8). Despite limited impact at the

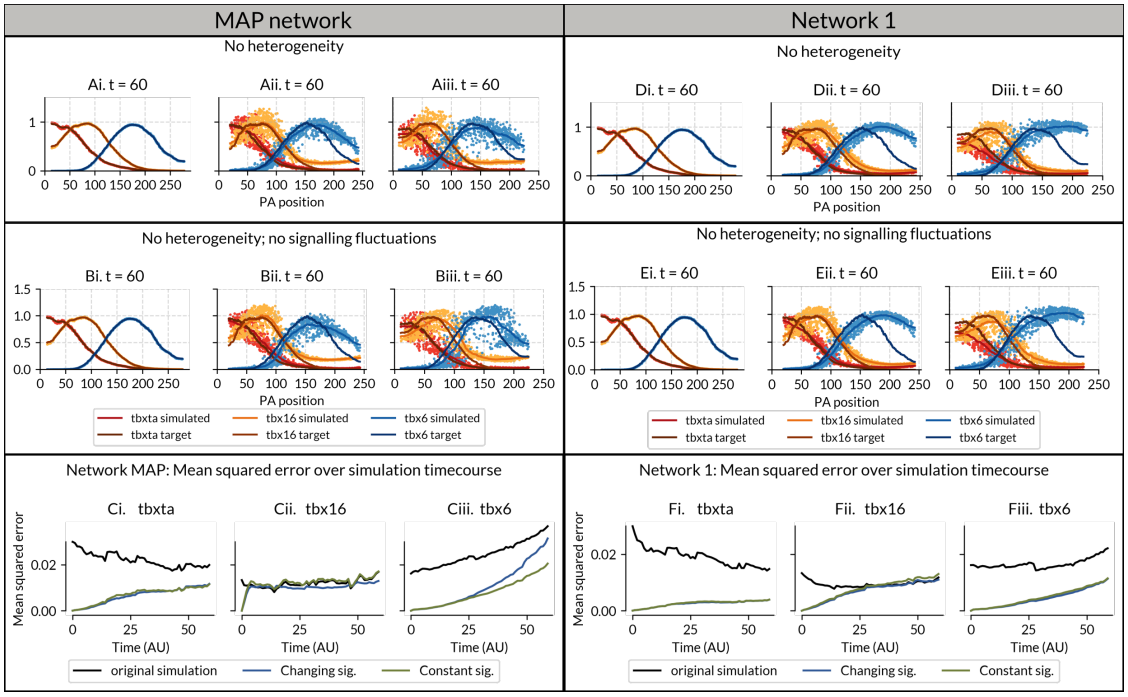


Figure 6.9: Gene expression heterogeneity is due to cell movement. A, A', A'': Simulated gene expression pattern after replacing initial gene expression with the rolling average. Black lines represent the "moving" simulation rolling average, light lines represent the perturbed rolling average, and dots represent individual cell's gene expression. B, B', B'': Simulated gene expression pattern after replacing initial gene expression with the rolling average **and** fixing signalling equal to the initial timepoint. Graph is formatted as in A. C, C', C'': Quantification of the mean squared distance to the rolling average, over time, for different simulations.

level of patterning, gene expression is impacted at the single-cell level, with a difference of up to 0.5AU (40-50% max gene expression levels) in individual cells (Figure 6.8). Again, both the MAP network and Network 1 exhibit these effects (Figure 6.8). Altogether, this shows that pattern formation along the AP axis is robust to fluctuations in signalling at the tissue level, but not the single-cell level.

Finally, there is extensive heterogeneity in gene expression at the single cell level. I wanted to know whether this comes from cell movement, or heterogeneity in signalling profiles. To test this, I set the initial gene expression in each cell equal to that of the rolling average of gene expression (Figure 6.9A), removing all gene expression heterogeneity from the initial conditions of the simulation. Running this simulation with normal ("moving") signalling results in heterogeneity in gene expression increasing during the simulation, for both Network 1 and the MAP

network (Figure 6.9A, C). Heterogeneity is quantified by measuring the mean squared distance from the gene expression in each cell to the rolling average gene expression for each time point. This metric increases over time, for both networks, representing the increase in gene expression heterogeneity (Figure 6.9C).

Such heterogeneity could be due to cell movement, or signalling. I removed signalling heterogeneity by preventing cells from updating their signalling as they transmitted the PSM, as in the simulations above. As before, gene expression heterogeneity also increases over time. For *tbxta* and *tbx16* heterogeneity is identical between both simulations (Figure 6.9B, C). This implies that, for the MAP network and for Network 1, the heterogeneity in gene expression along the AP axis is due to cell movement, not fluctuations in signalling.

All clusters

I then wanted to verify that this result was not unique to the MAP network, so repeated these experiments on all of the cluster representatives from Chapter 5.

Other clusters behave very similarly to the MAP network when subjected to the same tests. For all clusters, the "moving" and "static" *tbxta* rolling averages are virtually identical (Figure 6.10A-G). As for the MAP network, all clusters exhibit a slight upregulation of *tbx16* in the anterior of "static" simulations (Figure 6.10A-G). The magnitude of this effect does differ between parameter sets, being more subtle in Network 1 than other networks. The expression of *tbx6* in "static" simulations is more variable: in Networks 0 and 1 it is very similar to the "moving" simulation, but in other networks (2, 3, and 4) the "static" simulations have a sharper/more binary transition between low and high *tbx6* expression (Figure 6.10A-G). Difference in gene expression at the single cell level likewise resembles the MAP network: *tbxta*, *tbx16*, and *tbx6* expression differ between the "static" and "moving" simulations by around 0.5AU. Cells in the posterior tend to overexpress *tbx16* and cells further anterior tend to under-express this gene (Figure 6.10), consistent with the subtle change in gene expression pattern at the tissue level and similar to that seen in the MAP network. The difference in *tbx6* expression

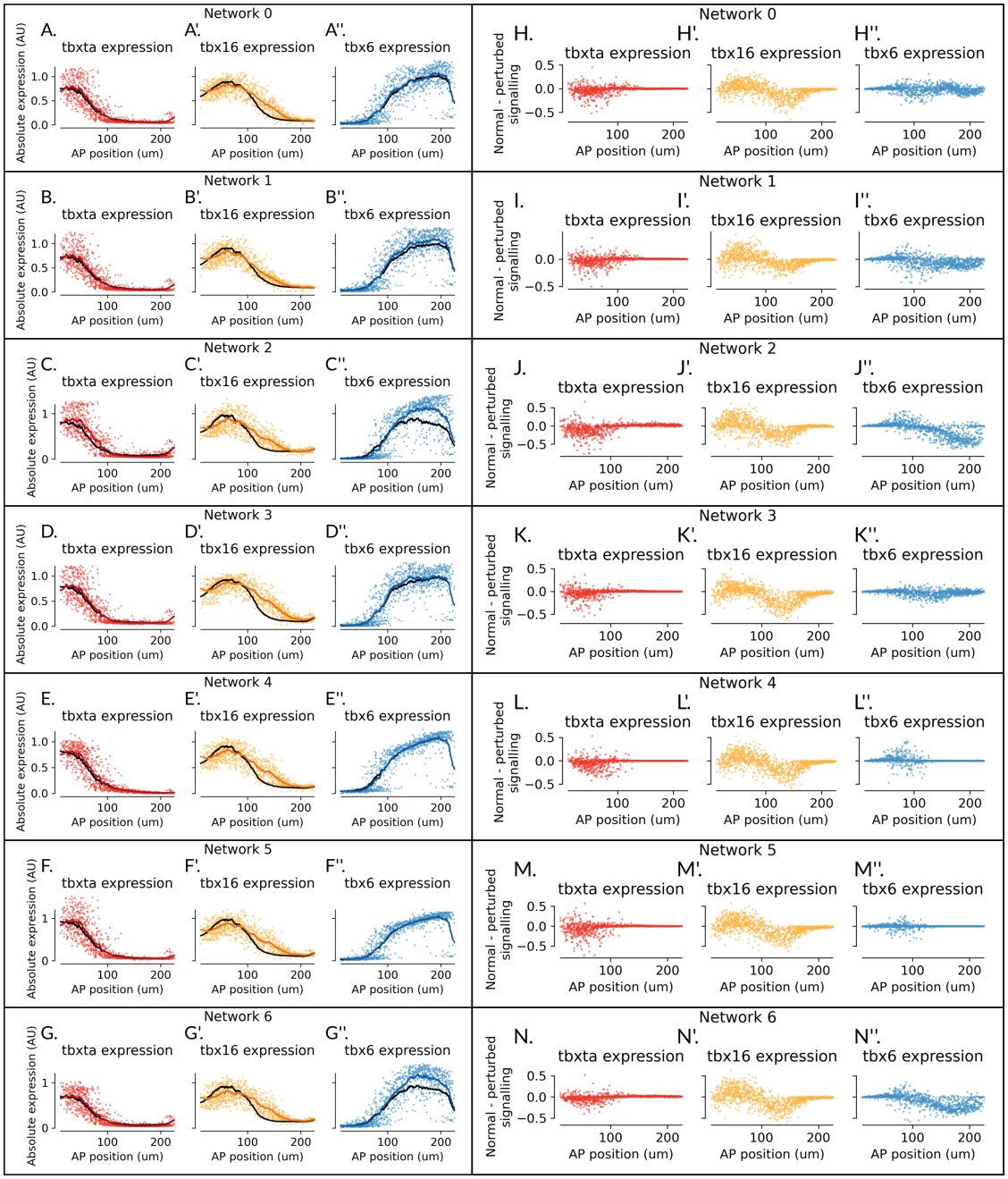


Figure 6.10: Gene expression pattern in perturbed simulations for all clusters. All graphs display gene expression at the final timepoint, when signalling is held constant for the duration of the simulation. Dots show expression in individual cells following perturbation; the black line represents the rolling average of "moving" expression for the parameter set chosen (averaging over the closest 50 cells), and the dark coloured line represents rolling average of the perturbed simulation. Different panels represent different parameter sets.

varies more between networks: in some (eg networks 4, 5) there is almost no difference between gene expression in static and moving simulations at the single cell level, and in other networks (eg. networks 6) expression can vary by up to 0.5AU (Figure 6.10). Altogether this shows that as with the MAP network, the other parameter sets are robust to fluctuations in signalling at the tissue level, but not the single cell level.

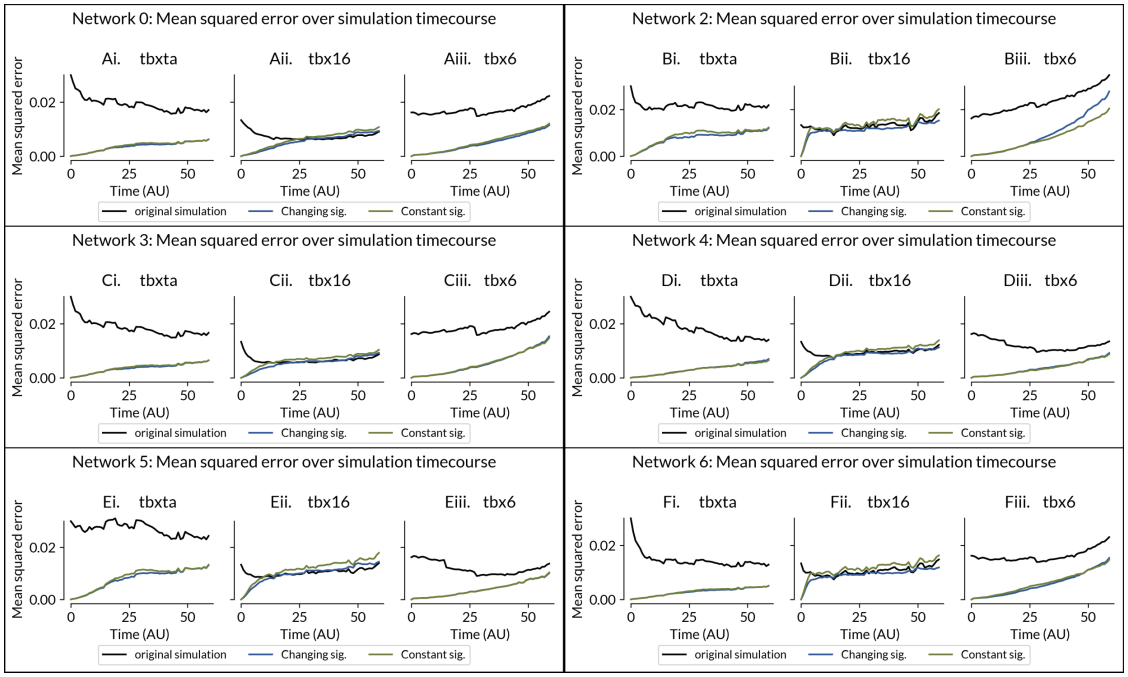


Figure 6.11: Heterogeneity in gene expression is mostly due to cell movement. Panels show quantification of the mean squared distance to the rolling average, over time, for different parameters sets.

Finally, I investigated whether heterogeneity in gene expression arises from cell movement or fluctuations in signalling. For all cluster representatives, I removed gene expression heterogeneity from the simulations as described above. As in the MAP network, heterogeneity increases over time for all clusters (Figure 6.11). Again as above, there is no difference in gene expression heterogeneity measured in the simulations with and without signalling fluctuations, meaning that gene expression heterogeneity is due to cell movement (Figure 6.11).

Altogether, this section shows that the PSM is surprisingly robust to the changes in signalling cells experience as they transit the PSM. It seems that

removing this changing signalling profiles has only a very subtle impact on the final expression pattern of the PSM, even though it does change gene expression at the single cell level. This effect is present in all parameter sets tested.

6.2.3 The MAP network is bistable, allowing different long-term behaviour of cells

In this section I will analyse the behaviour and function of the best fitting (MAP) network, reverse-engineered in Chapter 5. Dynamical systems theory provides powerful tools to achieve this, allowing one to identify the long-term behaviour of a system for given signalling values by computing the phase portrait. I chose signalling values to use for phase portrait analysis by computing the arithmetic average of signalling profiles across the AP axis of the simulation, and using values at evenly spaced intervals for simulation (see Figure 6.12 for signalling values).

Attractors were identified as described in the Methods section; I constrained the search space to between 0 and 5AU (ie five times the maximum gene expression in the PSM). I used seeds first, drawn from the initial gene expression at the first timepoint of the simulation, and second, from evenly spaced points between 0 and 5AU. These points were used as seeds to identify the root (fixed point) of the equation. If root-finding failed, I then simulated gene expression trajectories

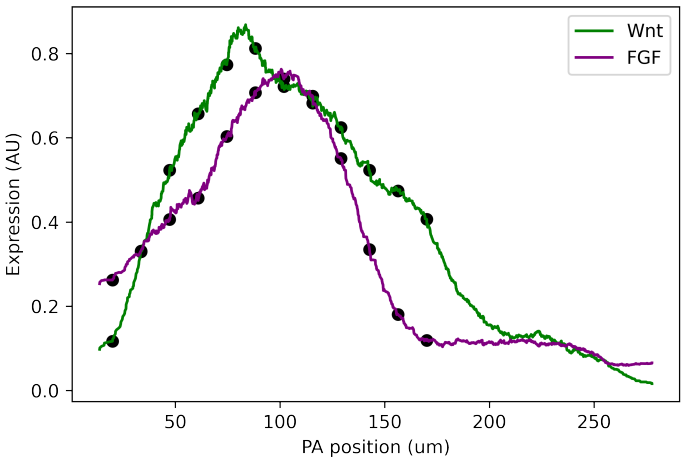


Figure 6.12: Signalling profiles chosen for phase portrait analysis. Green and purple lines: arithmetic average gene expression, averaged across 100 cells, for Wnt and FGF respectively. Black dots: individual expression levels used in the phase portraits.

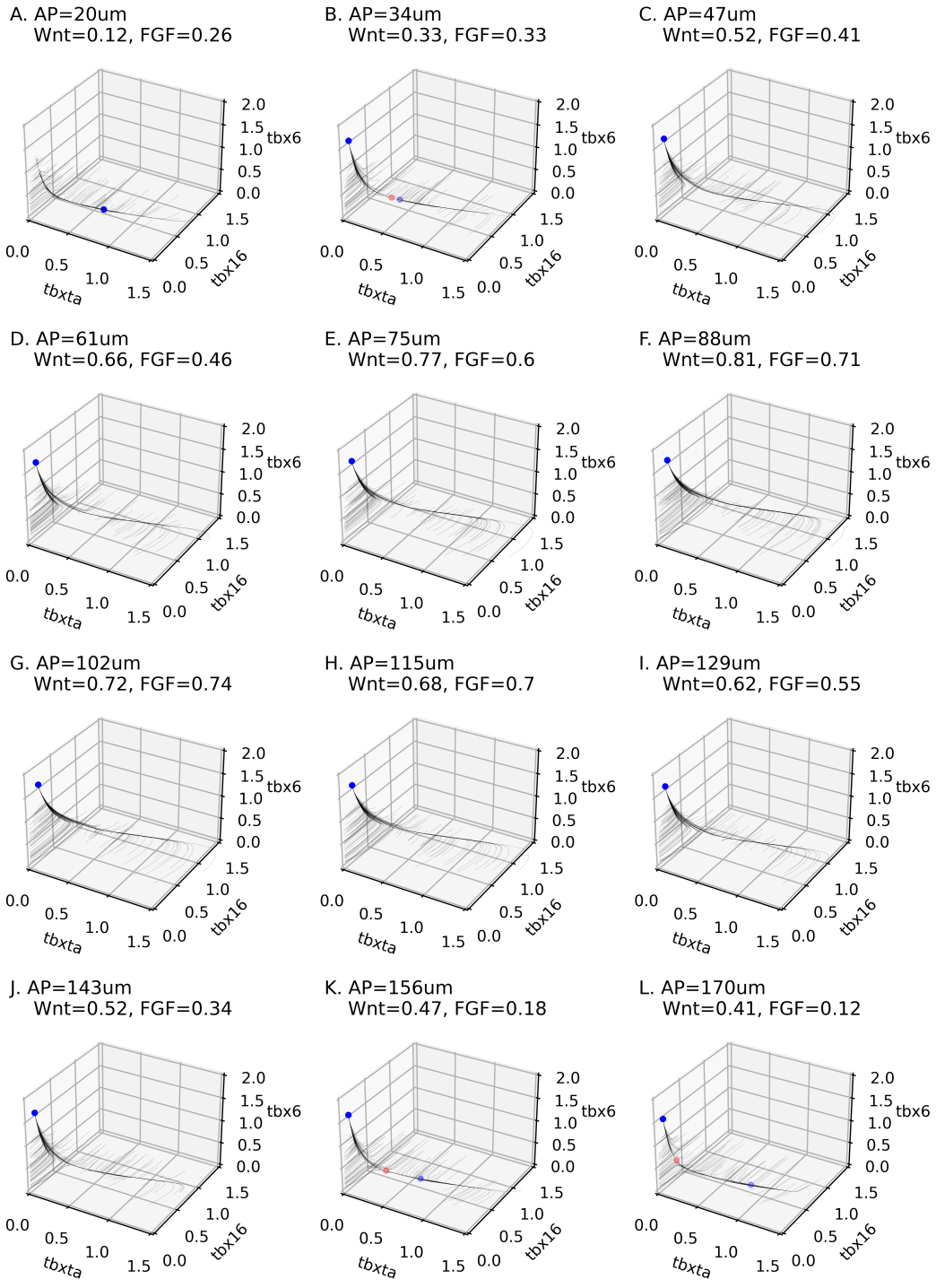


Figure 6.13: Phase portrait for the MAP network. Each panel represents the phase portrait for a given signalling value, depicted in Figure 6.12. Blue dots represent attractors and red dots represent saddle points. Black lines depict trajectories from initial gene expression positions for a randomly sampled 10% of cells.

for 20 times the total simulation duration and tried again; this always identified a fixed point (as defined by the rate of change being less than 10^{-5}AU for all genes). This method would miss any limit cycles, so I then inspected the simulated gene expression trajectories to show that these are absent.

The phase portrait for the MAP network differs at different signalling values, with regard to the number and type of attractor. At very low signalling values (Wnt=0.12; FGF=0.26) there is only one fixed point, an attractor, at *tbxta* = 0.62; *tbx16* = 0.51; *tbx6* = 0.04 (Figure 6.13A). At slightly higher signalling values (Wnt = 0.33, FGF = 0.33), another attractor at high *tbx6* and low *tbxta*/*tbx16* has appeared; this is separated from the low *tbx6* attractor by a saddle point at *tbxta* = 0.28; *tbx16* = 0.50; *tbx6* = 0.13 (Figure 6.13B). By FGF=0.52; Wnt=0.41, or 50μm from the posterior of the tailbud position, the low *tbx6* attractor has disappeared and there is only a high *tbx6* attractor that persists in about the same position for other high signalling values (Figure 6.13C). The low *tbx6* attractor does re-appear in the posterior, where there are lower signalling values (Figure 6.13K-L).

I also analysed the flow/dynamics of the phase portrait by graphing gene expression trajectories for a randomly chosen proportion (10%) of initial conditions. For all signalling profiles analysed, gene expression profiles all rapidly converge onto a single trajectory, a stable manifold (Strogatz, 2019). The position of this stable manifold differs between different signalling profiles: at low FGF and Wnt values, it lies at lower *tbx16* values (*tbx16*=about 1.0 in Figure 6.13A vs 1.6 in (Figure 6.13F).

The simplicity of the phase portrait for the MAP network means that it can be effectively summarised with a small number of graphics. The attractor positions for different combinations of signalling values do not overlap in gene expression space, and therefore can be graphed on the same plot (Figure 6.14A). To understand any movement of the attractor caused by differing signalling profiles, I colour-coded the fixed points by FGF and Wnt expression respectively (Figure 6.14B,C). This shows that higher levels of Wnt signalling move the low

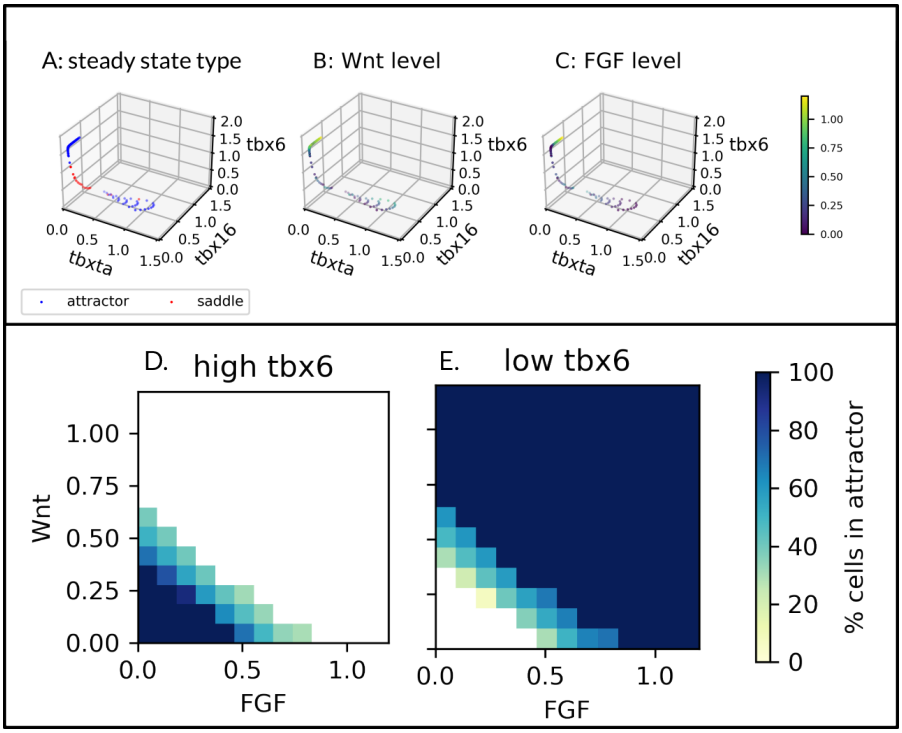


Figure 6.14: Condensed phase portrait visualization. For all panels, the phase portrait is computed for combinations of signalling between 0 and 1.3 and attractors graphed together. A. Fixed point position in gene expression space. Blue: attractors. Red: saddle points. B-C: Attractor position in gene expression space, colourcoded by Wnt and FGF level, respectively. D, E: Percentage of cells in the PSM present in a given fixed point's basin of attraction. Dark blue means all cells in the simulation are in a given basin of attraction; and white means zero cells.

tbx6 attractor towards higher *tbx16* levels and that increasing levels of FGF move this attractor towards lower *tbxta* levels (Figure 6.14B,C).

I also quantified the number of cells in the PSM inside the attractor basins, for differing signalling values (Figure 6.14D, E). For the MAP network, this analysis complements Figure 6.13. The low *tbx6* attractor is present for low FGF and Wnt values; both attractors are present for intermediate signalling values; and the high *tbx6* attractor is the only attractor in the majority of signalling space (Figure 6.14D, E).

Altogether, this shows that the MAP network has three fixed points: two attractors, at low and high *tbx6*; as well as a saddle point separating these attractors for signalling values where they coexist. The low *tbx6* attractor is only present for low signalling values.

6.2.4 Multiple potential dynamical mechanisms in the tbox GRN

I then investigated the other parameter sets generated in Chapter 5, to see whether any differing phase portraits can be used to achieve tbox pattern formation, and cell autonomous differentiation. For parameter sets representing each cluster, I generated graphs as in Figure 6.14, allowing analysis of the dynamical behaviour of different networks.

The seven clusters analysed differ in the topology of phase space, differing in the number, type, and position of attractors (Figure 6.15). Four of the eight networks studied have one attractor, at high *tbx6* and low *tbx16* and *tbxta* (Figure 6.15A, B, E, G). For these networks, this means that even cells at the most posterior positions of the zebrafish PSM have already committed to upregulating *tbx6* and differentiating. Despite the simplicity of these phase portraits, different signalling values do impact the position of attractors: for Clusters 0, 1 and 8, increasing Wnt levels moves the attractor to higher *tbx6* values, and increasing Wnt moves it to higher *tbx16* values (Figure 6.15Aii, iii, Bii,iii, Gii,iii). Note that Clusters 0 and 1 are still able to reproduce the dissociation experiment, despite only having one attractor (see Figure 5.15). Both FGF and Wnt signalling appear to similar effects on the Cluster 4 attractor, moving it to higher *tbx16* levels.

Other parameter sets have more complex phase spaces, with multiple attractors. Cluster 2 has six different fixed points. There are three attractors: one at high *tbx6* and low *tbxta/tbx16*, one at low *tbx6* and low *tbx16*, and another at high *tbxta* and low *tbx6* (Figure 6.15C). Cluster 2 also has a spiral sink at low *tbxta*, *tbx16*, and *tbx6*, as well as saddle points separating attractors (Figure 6.15C).

Clusters three and five also have multiple fixed points: one attractor at low *tbx6*, and another small attractor at low *tbx6*; these are separated by a saddle point (Figure 6.15D, F). The low *tbx6* attractors are only accessible when signalling is low, and are only accessed by some (25%) of cells in the PSM (Figure 6.16).

Altogether, then, our different clusters exhibit different dynamical behaviours, differing in number and position of attractors. This suggests that the embryo

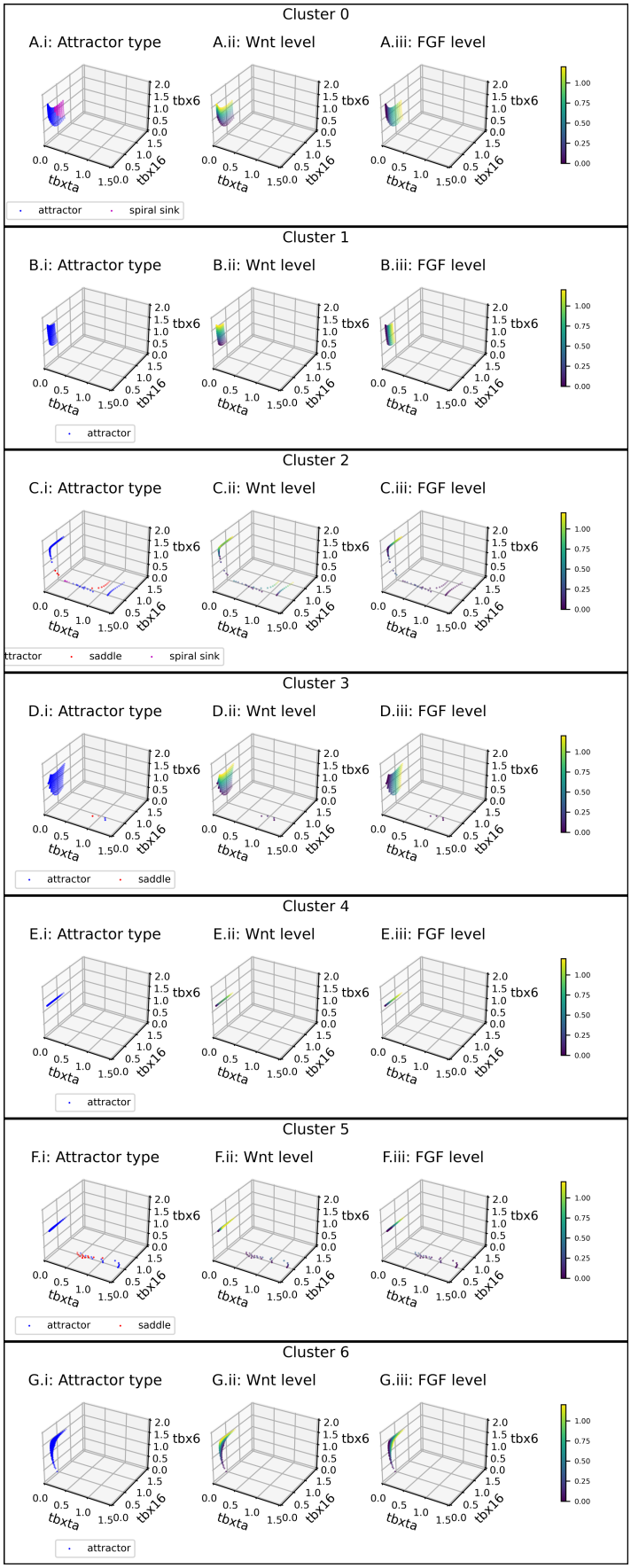


Figure 6.15: Different networks produce different phase portraits. A.i.-H.i.: Attractor position in gene expression space, colour-coded by attractor type. Blue: normal attractor; purple: spiral sink; red: saddle point. Attractors for all combinations of FGF and Wnt are shown on the same plot. A.ii.-H.ii: Attractor position in gene expression, colour-coded by Wnt signalling level producing that attractor. A.iii.-H.iii: Attractor position in gene expression space, colour-coded by FGF signalling level producing that attractor.

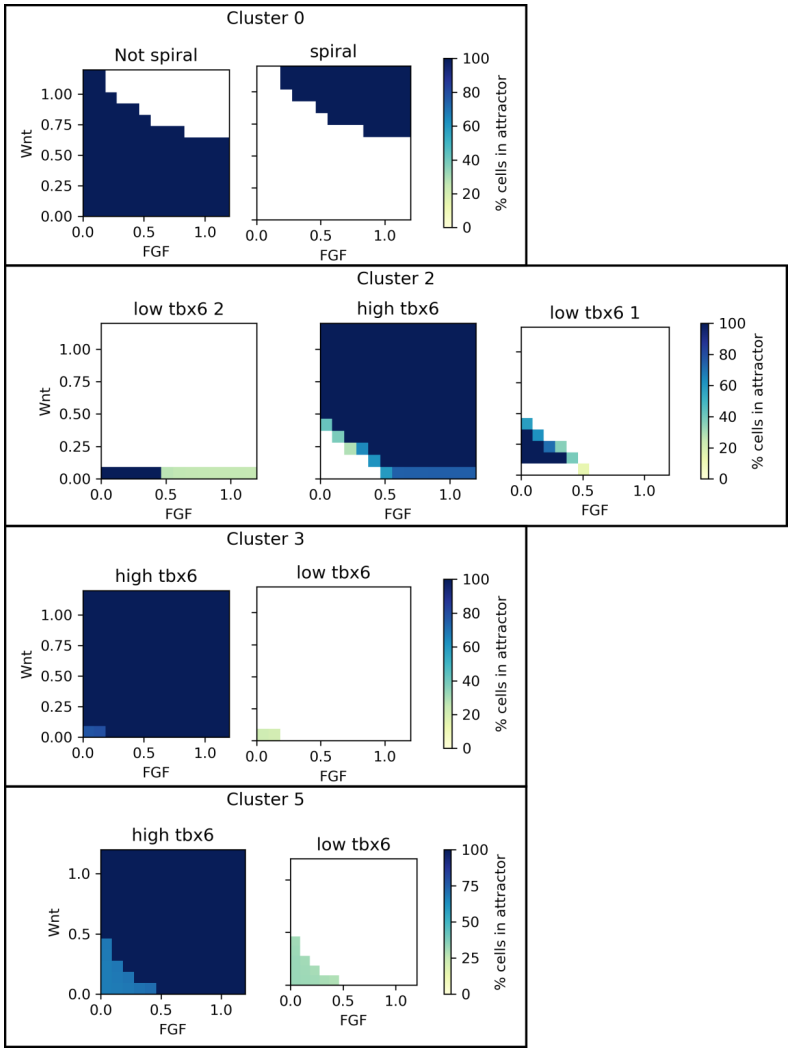


Figure 6.16: Percentage of cells in the PSM present in a given fixed point's basin of attraction. For differing signalling values, the number of cells present in a given attractor basin is quantified, as a measure of attractor basin size. Dark blue means all cells in the simulation are in a given basin of attraction; and white means zero cells.

could utilize one of several strategies to achieve pattern formation and the dissociation experiment.

6.3 Discussion

In this chapter, I analysed the behaviour of the networks generated using Live Modelling. I found that two different mechanisms can explain pattern maintenance in the tbox GRN: all parameter sets analysed have an attractor at high *tbx6* expression, which is responsible for *tbx6* upregulation in the PSM. Another set of parameters, including the MAP network, also have a second attractor at low *tbx6* and high *tbxta/tbx16* expression, representing a progenitor state of gene expression. This implies that the tbox GRN itself can act as a semi-

cell autonomous timer of differentiation. I also investigate whether positional information, as supplied by FGF and Wnt signalling, is required for correct pattern maintenance in the PSM, and show that this has a limited effect on pattern maintenance. These points will be expanded below.

6.3.1 Two dynamical mechanisms achieving Tbox patterning

Analysing the phase portraits of the different parameter sets revealed two mechanisms achieving the dissociation experiment and pattern formation. The first mechanism, used by the MAP network, has two attractors - one at high, and another at low, *tbx6*, separated by a saddle point. The low *tbx6* attractor is only present at low FGF and Wnt values; that is, in the very posterior of the PSM (Figure 6.13). This topology provides a tidy explanation for the dissociation experiment in Fulton, 2021: the two populations of cells exhibiting different differentiation dynamics are in different basins of attraction. Even cells in the tailbud with "undifferentiated" gene expression profiles are already committed to differentiate. Thus, when cells are dissected from the tailbud, they will simply continue differentiating according to their already-specified cell fates. It is also very easy to explain how some cells in the posterior can remain undifferentiated long-term using with this dynamical mechanism: they are simply in the low *tbx6* attractor and therefore unable to differentiate until they are exposed to different signalling values. Alternatively, intrinsic noise could cause a cell to change attractors.

The second mechanism consists of only one attractor, at high *tbx6*. Under this model, every cell in the tailbud is already committed to differentiate. The role of signalling is instead to control the rate of differentiation/the number of cells leaving the tailbud and upregulating *tbx6* in a given time interval. Parameter sets exhibiting this mechanism differ in whether they are able to maintain a population of cells with low *tbx6* during a simulated dissociation experiment: for some networks, all cells will upregulate *tbx6* during this perturbation, while others are able to maintain a low *tbx6* population for at least six hours.

Though we cannot determine from the *in silico* validations which mechanism is used by the embryo, experimental data does suggest that the one-attractor mechanism is more likely. Rohde et al., 2024 show that cells removed from the tailbud and cultured *in vitro* will all eventually differentiate, expressing *mesp-ba*. As *mesp-ba* is activated by *tbx6*, it is likely that (as discussed above) this differentiation is due to the upregulation of *tbx6*. As all cells eventually exhibit the same behaviour, this experiment suggests that there is only one cell fate (or attractor) in the PSM.

In mouse and chick, NMP cells represent a long-term progenitor cell population in the PSM, being able to adopt either a neural or mesodermal cell fate (Tzouanacou et al., 2009; Sambasivan & Steventon, 2021; Romanos et al., 2021). While the modelling framework in this chapter does not explicitly incorporate NMPs or NMP cell fate choice, if NMPs do exist in the zebrafish tailbud, we might expect a stable progenitor state in the PSM, representing cells that would adopt the neural cell fate if this was possible within the modelling framework. In zebrafish, a population of cells in the posterior PSM do co-express *tbxta* and *sox2*, and may therefore represent bi-potential progenitor cells (Attardi et al., 2018; Toh et al., 2022; Sambasivan & Steventon, 2021). However, existing experimental evidence implies that this population of cells is not bi-potent at the tailbud stage: Attardi et al., 2018 show that after tail eversion, NMP cells are mono-fated, and Saunders et al., 2025 show that there is no increased adoption of NMP cells into the neural fate after ablation of the neural tube. Therefore, at the tailbud stage, *sox2* and *tbxta* co-expressing cells in zebrafish do not appear to be true, bipotential, NMP cells. This therefore provides further evidence for the one-attractor proposed mechanism of cell differentiation in the zebrafish PSM.

In summary, our networks yield two different mechanisms that can achieve the dissociation experiment. Current experimental evidence suggests that the one-attractor mechanism is more likely.

6.3.2 The Tbox GRN: a cell-autonomous timer of differentiation?

My results suggest that the Tbox GRN acts as a cell-autonomous timer of differentiation in the zebrafish tailbud. There is a complex relationship between signalling and differentiation in our models. Cells moving anteriorly within the PSM are more likely to begin to upregulate *tbx6* and differentiate, even when the initial levels of gene expression are controlled. However, regardless of signalling profiles, some cells in the posterior PSM are able to autonomously upregulate *tbx6*; and cells with more anterior initial conditions are more likely to upregulate *tbx6*. This suggests that while signalling profiles alter the timing of differentiation, there is a substantial component of cell-autonomous differentiation. Phase portrait analysis supports this idea: for all networks analysed, there is an attracting fixed point at high *tbx6* values; and many cells, even those with 'progenitor-like' gene expression profiles, already lie within the basin of attraction for the high *tbx6* attractor.

There is experimental evidence for cell-autonomous differentiation in the tbox GRN. Rohde et al., 2024 study cell-intrinsic behaviour in the zebrafish segmentation clock and show that cells dissected from the very posterior PSM autonomously slow their oscillations, and upregulate the somite boundary gene *mesp-ba* when cultured *in vitro*. Therefore cell intrinsic factors are sufficient to produce cell differentiation, though there was increased noise in the system for cells cultured *in vitro*. Fulton, 2021 also demonstrates a very similar result for the zebrafish tbox GRN; namely that cells dissected from the posterior can autonomously upregulate *tbx6 in vitro* and downregulate *tbx16*; though in this case, importantly, one population of cells maintains low *tbx6* for the duration of culture (six hours). This discrepancy may simply be due to increased duration of incubation in Rohde et al., 2024, with cells incubated *in vitro* for longer, and some cells taking 10 hours to upregulate *mesp-ba in vitro*. Alternatively, it may show that there is an extrinsic signal required to upregulate a small number of cells.

Contrary to our results, Rohde et al., 2024 show that increasing FGF8 levels slows differentiation. When I decreased overall levels of FGF signalling I saw a decrease in *tbx6* level, contrary to this result. However this may be because we are using different readouts of FGF signalling (FGF8 in Rohde et al., 2024 versus dpERK in our data).

Rohde et al., 2024 suggest that an as-yet undiscovered gene may act as a cell-autonomous timer of the segmentation clock. My data suggests that, on the contrary, the Tbox GRN itself is acting as the timer. The Tbox GRN pushes cells through an invariant *tbxta-tbx16-tbx6* cycle; the rate at which they do so is modulated by signalling; but their fate is fixed. *tbx6*, in turn, upregulates *mesp-ba* and could ultimately control differentiation (Sawada et al., 2000; Nikaido et al., 2002; Cutty et al., 2012; Windner et al., 2015). Therefore, the zebrafish tbox GRN may act as a cell-autonomous timer of cell differentiation in the zebrafish PSM.

6.3.3 Degeneracy in the genotype-phenotype map

The fact that two dynamical mechanisms can achieve the same outcome suggests an additional level of degeneracy in the genotype-phenotype map. It is well-accepted that the relationship between genotypes and corresponding phenotypes is complex and degenerate; it is naïve to claim that any given gene "codes for" a particular phenotype (Alberch, 1991; Pigliucci, 2010; DiFrisco & Jaeger, 2020). Instead, complex interactions between different developmental processes, such as morphogenesis, gene regulation, and physical properties of developing tissues co-operate to produce the adult form (Alberch, 1991; Pigliucci, 2010; DiFrisco & Jaeger, 2020). DiFrisco and Jaeger, 2020 present a version of the genotype-phenotype map where dynamical regimes of GRNs mediate between the "genotype" — or GRN topology — and the "phenotype" (in that paper, patterning of the developing embryo). They propose that the relationship between GRN topology and dynamical mechanism is many-to-one, or degenerate, because one GRN topology can be associated with multiple parameter sets and therefore multiple dynamical mechanisms (DiFrisco & Jaeger, 2020).

The analysis in this chapter suggests a further level of degeneracy. Multiple different GRN topologies, and dynamical mechanisms, are able to achieve the same phenotypes. In this case I am defining phenotype as the patterning of the PSM, and behaviour of individual cells when isolated from the PSM. Two different dynamical mechanisms can achieve both of these behaviours; one with a single attractor, and another with multiple attractors (Figure 6.15). Therefore, not only can the same genotype (GRN topology) produce multiple phenotypes (dynamical behaviours of the GRN, and subsequent influences on pattern formation within the embryo), but identical phenotypes can be produced from different dynamical mechanisms.

6.3.4 Is positional information required for pattern maintenance in the tbox GRN?

I also investigated whether positional information, as supplied by the level of signalling experienced by the cell, is required for correct patterning. The overall level of signalling is required for correct pattern maintenance in the PSM: perturbing FGF and Wnt signalling in Chapter 5 results in a significant alteration of patterning compared to unperturbed simulations (Figure 5.14). Signalling also appears to modulate the rate of cell differentiation *in silico*. Cells that move anterior in the PSM during the simulation have different signalling profiles than those moving posteriorly, and additionally, are more likely to upregulate *tbx6* than those moving posteriorly. Therefore signalling appears to predispose cells moving anteriorly to differentiate/upregulate *tbx6*.

As cells transit the PSM, they are exposed to different levels of FGF and Wnt signalling. If these signalling levels are required to supply positional information to the tbox GRN, and therefore regulate pattern maintenance in the PSM, we would expect that preventing cells from updating their signalling as they transit the PSM to significantly alter the pattern of gene expression in the PSM. However, we do not observe this during tbox patterning of the PSM. I showed this by "freezing" the signalling profiles of individual cells to their initial point as they

move through the PSM. This perturbation has subtle effects at the tissue level. *tbxta* pattern was virtually unchanged, there is a slight expansion of *tbx16* pattern, and *tbx6* expression is sharper. Despite this robustness at the level of gene expression pattern, at the single cell level, gene expression does differ after perturbation, by up to 0.5AU for each gene. This therefore shows that signalling profiles do impact gene expression levels in individual cells, but not the overall gene expression pattern within the PSM, Therefore, at least over the timescales of this simulation and for our reverse-engineered parameter sets, positional information as supplied by FGF and Wnt signalling in the PSM is not required for correct tbox pattern maintenance. Similarly, I also show that heterogeneity in gene expression comes from cell movement, not heterogeneity in signalling.

Overall, then, in these models, pattern formation is highly robust at the tissue level — but not the single cell level — to fluctuations in the signalling profile of cells, and heterogeneity in gene expression arises from fluctuations in cell movement, not signalling. Though this result is surprising, there is experimental data supporting the idea. Fulton, 2021 prevents cell movement in the zebrafish PSM by treating embryos with blebbistatin. This has no effect on tbox gene expression within the PSM (Fulton, 2021). This shows that pattern formation within the zebrafish PSM, as in our computational model, is surprisingly robust to cell movement.

6.4 Conclusions

My results show that, surprisingly, the zebrafish tbox GRN appears to achieve robust and stable pattern formation without extensive signalling-mediated positional information. Instead, cells are able to upregulate *tbx6* and differentiate cell-autonomously. Additionally, I found that different dynamical mechanisms are able to achieve the same outcomes (tbox gene expression patterning, and autonomous upregulation of *tbx6*). This underscores the degeneracy of the genotype to phenotype map.

Chapter 7: General discussion

The overarching aim of this thesis is to investigate how morphogenesis and gene regulation interact to achieve phenotypic diversity.

I first investigated how Lake Malawi cichlids are able to evolve divergent phenotypes from such a similar genetic substrate. I found that *R. chillingali* appear to have evolved increased vertebral and somitic counts, not by evolving the process of segmentation itself, but instead by evolving the initial conditions of the process. At the onset of somitogenesis, the *R. chillingali* PSM is longer, greater in volume, and has more cells than the *A. calliptera* PSM. In addition, the only differences in PSM anatomy are detected at the early stages of segmentation, and expression of key genes is in the PSM is conserved between species. All this suggests that *R. chillingali* have evolved an increased somite number by evolving the initial conditions of segmentation, an example of a species evolving a phenotypic novelty by altering the initial conditions of development, not the process of development itself.

In Chapter 5, I developed and validated a method allowing GRNs achieving pattern formation in the zebrafish PSM to be reverse-engineered alongside cell movements. These parameter sets were able to recapitulate key functions of the tbox GRN: gene expression patterning within the PSM, network topology, response to over- and under- expression of signalling pathways, and the ability to achieve two different differentiation dynamics in the absence of variation in signalling. As none of these traits were used for reverse-engineering, this provides strong evidence that this reverse-engineering methodology is able

to effectively produce biologically plausible parameter sets which could be investigated further experimentally.

Ultimately, Live Modelling allows us to investigate a major question in developmental biology: how do morphogenetic and gene regulatory networks interact to achieve development? In Chapter 6 I tackle this problem. I show that surprisingly, patterning of the PSM is possible without the positional information provided by signalling, and that the tbox GRN is behaving as a cell-autonomous timer of cell differentiation. That is, fixing a cells' signalling levels as it transits the PSM leads to subtle effects on patterning: upregulation of *tbx16* in the anterior PSM, and sharper transitions in *tbx6* expression. However, the different signalling profiles experienced by cells moving anteriorly predispose them to differentiate (upregulate *tbx6*). This implies that, though signalling does modulate the speed of cell differentiation, this effect is subtle, and the effect of the changes in signalling as cells transit the PSM on pattern formation is minimal. Thus, at least for these parameter sets and over the timescales studied, the positional information provided by these signals is not required for patterning of the PSM.

7.1 Live modelling allows one to reverse-engineer parameter sets reproducing gene expression patterning alongside morphogenesis

A significant question in developmental biology is, how do morphogenesis and gene regulatory networks co-operate to achieve pattern formation during embryonic development. In Chapter 5, I develop and validate Live Modelling, a methodology with which to investigate the interaction of morphogenesis and gene expression patterning. This allows one to firstly, approximate gene expression trajectories in individual cells by combining fixed and live-imaged data sets, secondly, reverse-engineer parameter sets using these approximated gene expression trajectories, thirdly, simulate GRNs onto moving cell tracks, and finally, validate the resulting networks.

This methodology could be used in the future to infer and simulate GRNs in other biological systems where cell movement and gene expression patterning co-exist, such as establishment of the body axis during gastrulation. In addition, its various modules can be adapted and used in other systems. Ultimately this presents a step forwards in understanding how patterning and cell movements interact to achieve patterning.

The next step to expand this methodology is to incorporate the impact of gene regulation on cell movements, in particular the known impact of *tbx16* on cell motility (Manning & Kimelman, 2015). This will require a realistic model of cell movement in the PSM to be developed, incorporating the role of *tbx16* in cell motility. After this is achieved, including a gene regulatory network describing gene expression patterning in the PSM would allow the role of gene regulation on morphogenesis to be further investigated within and across species. This will allow us to explore whether morphogenetic or genetic parameters are more influential in achieving phenotypic evolution.

7.2 The Tbox GRN is a cell-autonomous timer of differentiation

I then analysed the parameter sets generated in Chapter 5, to investigate the effect of cell movement on gene expression patterning of the PSM, and differentiation of cells in the PSM. The parameter sets reverse-engineered in Chapter 5 are surprisingly robust to the changes in signalling levels that cells experience as they transit the PSM. This suggests that positional information from signalling is not strictly necessary for pattern maintenance in the zebrafish PSM, at least over the timescales of the simulation. The Tbox GRN is not completely indifferent to signalling. Under- or over- expressing FGF and Wnt signalling significantly interferes with correct pattern in the PSM, meaning that the correct level of FGF and Wnt is required for correct patterning. In addition, interfering with the signalling cells experience as they transit the PSM does alter gene expression at the single cell level. Therefore, signalling does appear to impact gene expression and

differentiation at the single cell level, but positional information, as supplied by signalling, is not required for maintenance of correct pattern in the PSM overall.

My work shows that tbox patterning in the PSM can be achieved by either a one- or two- attractor mechanism. Parameter sets reverse-engineered in Chapter 5 differ in the number and arrangement of attractors, with many having just one attractor, at high *tbx6* and low *tbxta* and *tb16* expression. This implies that the autonomous upregulation of *tbx6* seen by Fulton, 2021 and upregulation of *mesp-ba* seen by Rohde et al., 2024 may occur due to the cell-intrinsic dynamics of the tbox GRN. This finding is particularly significant given the proposal of Rohde and colleagues (Rohde et al., 2024) that an as-yet undiscovered gene is responsible for the autonomous slowing of the segmentation clock and upregulation of *mesp-ba*. Here I suggest that the Tbox GRN itself is acting as a cell autonomous timer, regulating cell differentiation. This idea should be investigated further.

The rate of differentiation is regulated by the level of signalling experienced by the cell. Cells which are moving anteriorly within the embryo are more likely to begin to upregulate *tbx6* and therefore differentiate, in contrast to those moving posteriorly. This effect persists even when the initial gene expression within the cell is controlled, showing that it can be attributed to signalling. However, signalling levels are not the only factor influencing the rate of differentiation: cells initialized with gene expression drawn from more anteriorly in the embryo are more likely to begin to differentiate. This suggests that signalling levels, and initial gene expression in the cell, co-operate to regulate cell differentiation.

The fact that different dynamical mechanisms can achieve the same outcomes — correct pattern maintenance in the PSM and autonomous up-regulation of *tbx6* — suggests that another layer of degeneracy in the genotype to phenotype map. It is well-established that the relationship between the genotype and the phenotype is complex and mediated by the developmental process and GRNs (Alberch, 1991; Pigliucci, 2010; DiFrisco & Jaeger, 2020). One GRN can produce different phase portraits, depending on its context and how it is parameterized. In turn, different dynamical mechanisms (phase portraits) can achieve different

phenotypic outcomes (DiFrisco & Jaeger, 2020). Here, I present an example of a GRN where the same phenotypic outcome — pattern formation, and cell differentiation dynamics — can be achieved by different dynamical mechanisms, thus presenting a further level of degeneracy to the genotype-to-phenotype map.

7.3 A detailed description of *A. calliptera* development during axial elongation

In Chapter 3, I produce a detailed description of patterning and morphology of the *A. calliptera* PSM during segmentation and axial elongation. *A. calliptera* closely resemble zebrafish in the development, anatomy, and gene expression patterning of the PSM during axial elongation, with a number of exceptions. Firstly, as also shown by Marconi et al., 2023, there is a heterochonic shift in the timing of tail eversion between zebrafish and *A. calliptera*, which may have contributed to morphological evolution between these species. Secondly, there is evidence for evolution in the GRNs controlling patterning of the PSM in *A. calliptera*: unlike in zebrafish, expression of the tbox genes retracts over time, and the relative expression of *mesp-aa* and *tbx6* differ. Finally, there is preliminary evidence that the mechanism of elongation may differ over time in *A. calliptera*: at early segmentation, there is a density gradient across the PSM, consistent with a diffusion gradient of cell motility achieving elongation, but this is absent at later stages. The detailed characterization of *A. calliptera* development during axial elongation in Chapter 3 will also form a useful resource for future evo-devo research in teleost fishes.

7.4 Key differences between cichlid species established at the onset of segmentation

In Chapter 4, I performed a detailed comparison of *A. calliptera* and *R. chillingali* development during axial elongation. I show that the increased somite count in *R. chillingali* can be explained by the increased size (length, volume, and cell

number) of the *R. chillingali* PSM at the onset of segmentation. This means that *R. chillingali* have likely evolved increased somite numbers by evolving the initial conditions of segmentation, not the process of segmentation. This proposes a novel way in which phenotypes can evolve, particularly over short timescales and with minimal genomic evolution.

This mechanism, evolution via initial conditions, may also operate in other species. For example, during arthropod segmentation, the overall dynamics of segmentation differ between different species: some arthropods, like the beetle *Tribolium* exhibit a method of segmentation resembling the vertebrate segmentation clock, while other species such as *Drosophila* appear to pre-pattern segments via gap genes (Peel et al., 2005; Clark et al., 2019). These two, apparently very different, mechanisms of development can be reconciled using the same gene regulatory network, but altering how it is deployed within the embryo (Clark, 2017; Clark & Peel, 2018). Therefore, evolving phenotypic diversity by evolving the initial or boundary conditions of a process may be a widespread phenomenon.

It is also possible that an innovative mechanism of epiboly predisposes haplochromine cichlids to morphological innovations via modifying the initial conditions of segmentation, and this may have contributed to the adaptive radiations of halpochromines in Lake Malawi and elsewhere.

7.5 Overall conclusions

My overarching goal in this thesis was to investigate how morphogenesis and GRNs interact to achieve development and the evolution of phenotypic diversity. My results suggest that, firstly, it is possible to evolve phenotypic diversity without evolving the developmental process itself: *R. chillingali* appear to have evolved an increased number of somites by increasing the size of the presomitic mesoderm at the onset of segmentation. This could be a general mechanism to achieve phenotypic diversity. I additionally develop and validate a method to reverse-engineer GRNs in a morphogenetic context. This allows the interaction

of cell movement and patterning to be explicitly studied. Indeed, GRNs reverse-engineered using this methodology are able to achieve patterning and differentiation with only a subtle effect from signalling. This therefore investigates the relationship between GRNs and morphogenesis from two angles: how phenotypic diversity can be evolved in the absence of genetic changes, as in Lake Malawi cichlids, and how morphogenetic processes (cell movements) interact with GRNs to regulate development.

References

Abe, G. et al. (2014). "The Origin of the Bifurcated Axial Skeletal System in the Twin-Tail Goldfish". *Nature Communications* 5.1, p. 3360. DOI: 10.1038/ncomms4360.

Abzhanov, A. et al. (2004). "Bmp4 and Morphological Variation of Beaks in Darwin's Finches". *Science* 305.5689, pp. 1462–1465. DOI: 10.1126/science.1098095.

Al-Mosleh, S. et al. (2021). "Geometry and Dynamics Link Form, Function, and Evolution of Finch Beaks". *Proceedings of the National Academy of Sciences* 118.46, e2105957118. DOI: 10.1073/pnas.2105957118.

Alberch, P. (1991). "From Genes to Phenotype: Dynamical Systems and Evolvability". *Genetica* 84.1, pp. 5–11. DOI: 10.1007/BF00123979.

Armstrong, P. B. and J. S. Child (1965). "Stages in the Normal Development of Fundulus Heteroclitus". *The Biological Bulletin* 128.2, pp. 143–168. DOI: 10.2307/1539545.

Attardi, A. et al. (2018). "Neuromesodermal Progenitors Are a Conserved Source of Spinal Cord with Divergent Growth Dynamics". *Development* 145.21, dev166728. DOI: 10.1242/dev.166728.

Aviñó-Esteban, L. and J. Sharpe (2024). "Spatio-Temporal Reconstruction of Gene Expression Patterns in Developing Mice". *bioRxiv*, p. 2024.08.13.607727. DOI: 10.1101/2024.08.13.607727.

Ban, H. et al. (2019). "Transcriptional Autoregulation of Zebrafish *Tbx6* Is Required for Somite Segmentation". *Development* 146.18, dev177063. DOI: 10.1242/dev.177063.

Banavar, S. P. et al. (2021). "Mechanical Control of Tissue Shape and Morphogenetic Flows during Vertebrate Body Axis Elongation". *Scientific Reports* 11.1, p. 8591. DOI: 10.1038/s41598-021-87672-3.

Barrett, T. et al. (2024). *Data.Table: Extension of 'data.Frame'*. manual.

Baum, B., J. Settleman, and M. P. Quinlan (2008). "Transitions between Epithelial and Mesenchymal States in Development and Disease". *Seminars in Cell & Developmental Biology*. Cell Shape and Tissue Morphogenesis 19.3, pp. 294–308. DOI: 10.1016/j.semcdb.2008.02.001.

Bénazéraf, B. (2019). "Dynamics and Mechanisms of Posterior Axis Elongation in the Vertebrate Embryo". *Cellular and Molecular Life Sciences* 76.1, pp. 89–98. DOI: 10.1007/s00018-018-2927-4.

Bénazéraf, B. et al. (2010). "A Random Cell Motility Gradient Downstream of FGF Controls Elongation of an Amniote Embryo". *Nature* 466.7303, pp. 248–252. DOI: 10.1038/nature09151. pmid: 20613841.

Bénazéraf, B. et al. (2017). "Multi-Scale Quantification of Tissue Behavior during Amniote Embryo Axis Elongation". *Development* 144.23, pp. 4462–4472. DOI: 10.1242/dev.150557.

Bindu, L. and K. G. Padmakumar (2016). "Developmental Aspects of the Endmic Cichlid Fish *Etroplus Suratensis* (Bloch, 1790) in Captivity". *Journal of Aquatic Biology & Fisheries* 4, pp. 45–55.

Bouldin, C. M. et al. (2015). "Wnt Signaling and Tbx16 Form a Bistable Switch to Commit Bipotential Progenitors to Mesoderm". *Development* 142.14, pp. 2499–2507. DOI: 10.1242/dev.124024.

Bruce, A. E. E. and R. Winklbauer (2020). "Brachyury in the Gastrula of Basal Vertebrates". *Mechanisms of Development* 163, p. 103625. DOI: 10.1016/j.mod.2020.103625.

Bucklow, C. V. et al. (2024). "A Whole-Body Micro-CT Scan Library That Captures the Skeletal Diversity of Lake Malawi Cichlid Fishes". *Scientific Data* 11.1, p. 984. DOI: 10.1038/s41597-024-03687-1.

Burel, J.-M. et al. (2015). "Publishing and Sharing Multi-Dimensional Image Data with OMERO". *Mammalian Genome* 26.9, pp. 441–447. DOI: 10.1007/s00335-015-9587-6.

Busby, L. and B. Steventon (2021). "Tissue Tectonics and the Multi-Scale Regulation of Developmental Timing". *Interface Focus* 11.3, p. 20200057. DOI: 10.1098/rsfs.2020.0057.

Buslov, A. V., N. P. Sergeeva, and O. I. Il'in (2010). "Embryonic Development of the Pacific Cod *Gadus Macrocephalus* (Gadidae)". *Russian Journal of Marine Biology* 36.7, pp. 526–538. DOI: 10.1134/S1063074010070060.

Cerqueira, V. R. (2005). "Egg Development of *Paralichthys Orbignyanus* (Valenciennes, 1839)". *Brazilian Archives of Biology and Technology* 48, pp. 459–465. DOI: 10.1590/S1516-89132005000300016.

Choi, H. M. T. et al. (2016). "Mapping a Multiplexed Zoo of mRNA Expression". *Development* 143.19, pp. 3632–3637. DOI: 10.1242/dev.140137.

Choi, H. M. T. et al. (2018). "Third-Generation in Situ Hybridization Chain Reaction: Multiplexed, Quantitative, Sensitive, Versatile, Robust". *Development* 145.12, dev165753. DOI: 10.1242/dev.165753.

Clark, B. et al. (2022). "*Oca2* Targeting Using CRISPR/Cas9 in the Malawi Cichlid *Astatotilapia Calliptera*". *Royal Society Open Science* 9.4, p. 220077. DOI: 10.1098/rsos.220077.

Clark, B. et al. (2023). "Genome Editing in East African Cichlids and Tilapias: State-of-the-Art and Future Directions". *Open Biology* 13.11, p. 230257. DOI: 10.1098/rsob.230257.

Clark, E. (2017). "Dynamic Patterning by the *Drosophila* Pair-Rule Network Reconciles Long-Germ and Short-Germ Segmentation". *PLOS Biology* 15.9, e2002439. DOI: 10.1371/journal.pbio.2002439.

Clark, E. and M. Akam (2016). "Odd-Paired Controls Frequency Doubling in Drosophila Segmentation by Altering the Pair-Rule Gene Regulatory Network". *eLife* 5, e18215. DOI: 10.7554/eLife.18215. pmid: 27525481.

Clark, E. and A. D. Peel (2018). "Evidence for the Temporal Regulation of Insect Segmentation by a Conserved Sequence of Transcription Factors". *Development (Cambridge, England)* 145.10, dev155580, dev.155580. DOI: 10.1242/dev.155580. pmid: 29724758.

Clark, E., A. D. Peel, and M. Akam (2019). "Arthropod Segmentation". *Development (Cambridge, England)* 146.18, dev170480. DOI: 10.1242/dev.170480. pmid: 31554626.

Cooke, J. and E. C. Zeeman (1976). "A Clock and Wavefront Model for Control of the Number of Repeated Structures during Animal Morphogenesis". *Journal of Theoretical Biology* 58.2, pp. 455–476. DOI: 10.1016/s0022-5193(76)80131-2. pmid: 940335.

Criswell, K. E. and J. A. Gillis (2020). "Resegmentation Is an Ancestral Feature of the Gnathostome Vertebral Skeleton". *eLife* 9. Ed. by P. C. Yelick, C. J. Rosen, and R. Keynes, e51696. DOI: 10.7554/eLife.51696.

Crombach, A. et al. (2012). "Efficient Reverse-Engineering of a Developmental Gene Regulatory Network". *PLoS Computational Biology* 8.7. DOI: 10.1371/journal.pcbi.1002589. pmid: 22807664.

Crombach, A. et al. (2016). "Gap Gene Regulatory Dynamics Evolve along a Genotype Network". *Molecular Biology and Evolution* 33.5, pp. 1293–1307. DOI: 10.1093/molbev/msw013.

Cutty, S. J. et al. (2012). "Identification and Expression Analysis of Two Novel Members of the Mesp Family in Zebrafish". *The International Journal of Developmental Biology* 56.4, pp. 285–294. DOI: 10.1387/ijdb.113447sc.

Daggett, D. F. et al. (2007). "Control of Morphogenetic Cell Movements in the Early Zebrafish Myotome". *Developmental Biology* 309.2, pp. 169–179. DOI: 10.1016/j.ydbio.2007.06.008.

Das, D. et al. (2017). "Patterned Disordered Cell Motion Ensures Vertebral Column Symmetry". *Developmental Cell* 42.2, 170–180.e5. DOI: 10.1016/j.devcel.2017.06.020.

De Jong, I. M., F. Witte, and M. K. Richardson (2009). "Developmental Stages until Hatching of the Lake Victoria Cichlid *Haplochromis Piceatus* (Teleostei: Cichlidae)". *Journal of Morphology* 270.5, pp. 519–535. DOI: 10.1002/jmor.10716.

DiFrisco, J. and J. Jaeger (2020). "Genetic Causation in Complex Regulatory Systems: An Integrative Dynamic Perspective". *BioEssays* 42.6, p. 1900226. DOI: 10.1002/bies.201900226.

Dubrulle, J. and O. Pourquié (2004). "Fgf8 mRNA Decay Establishes a Gradient That Couples Axial Elongation to Patterning in the Vertebrate Embryo". *Nature* 427.6973, pp. 419–422. DOI: 10.1038/nature02216.

Durbin, L. et al. (2000). "Anteroposterior Patterning Is Required within Segments for Somite Boundary Formation in Developing Zebrafish". *Development* 127.8, pp. 1703–1713. DOI: 10.1242/dev.127.8.1703.

Elmasri, H. et al. (2004a). "*Her7* and *Hey1*, but Not *Lunatic Fringe* Show Dynamic Expression during Somitogenesis in Medaka (*Oryzias Latipes*)". *Gene Expression Patterns* 4.5, pp. 553–559. DOI: 10.1016/j.modgep.2004.02.003.

Elmasri, H. et al. (2004b). "Mutations Affecting Somite Formation in the Medaka (*Oryzias Latipes*)". *Mechanisms of Development*. Medaka 121.7, pp. 659–671. DOI: 10.1016/j.mod.2004.04.003.

Essner, J. J. et al. (2005). "Kupffer's Vesicle Is a Ciliated Organ of Asymmetry in the Zebrafish Embryo That Initiates Left-Right Development of the Brain, Heart and Gut". *Development (Cambridge, England)* 132.6, pp. 1247–1260. DOI: 10.1242/dev.01663. pmid: 15716348.

Fior, R. et al. (2012). "The Differentiation and Movement of Presomitic Mesoderm Progenitor Cells Are Controlled by Mesogenin 1". *Development* 139.24, pp. 4656–4665. DOI: 10.1242/dev.078923.

Fischer, S. C. et al. (2023). "The Salt-and-Pepper Pattern in Mouse Blastocysts Is Compatible with Signaling beyond the Nearest Neighbors". *iScience* 26.11, p. 108106. DOI: 10.1016/j.isci.2023.108106. pmid: 37915595.

Fleig, R. (1993). "Embryogenesis in Mouth-Breeding Cichlids (Osteichthyes, Teleostei): Structure and Fate of the Enveloping Layer". *Roux's archives of developmental biology* 203.3, pp. 124–130. DOI: 10.1007/BF00365051.

Foreman-Mackey, D. et al. (2013). "Emcee: The MCMC Hammer". *Publications of the Astronomical Society of the Pacific* 125.925, pp. 306–312. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph, physics:physics, stat].

Frith, T. J. R., J. Briscoe, and G. L. M. Boezio (2024). "Chapter Five - From Signalling to Form: The Coordination of Neural Tube Patterning". In: *Current Topics in Developmental Biology*. Ed. by M. Mallo. Vol. 159. Vertebrate Pattern Formation. Academic Press, pp. 168–231. DOI: 10.1016/bs.ctdb.2023.11.004.

Fujimura, K. and N. Okada (2007). "Development of the Embryo, Larva and Early Juvenile of Nile Tilapia *Oreochromis Niloticus* (Pisces: Cichlidae). Developmental Staging System". *Development, Growth & Differentiation* 49.4, pp. 301–324. DOI: 10.1111/j.1440-169X.2007.00926.x.

Fulton, T. (2021). "Pattern Formation During Complex Morphogenesis". PhD thesis. University of Cambridge.

Fulton, T., B. Verd, and B. Steventon (2022a). "The Unappreciated Generative Role of Cell Movements in Pattern Formation". *Royal Society Open Science* 9.4, p. 211293. DOI: 10.1098/rsos.211293.

Fulton, T. et al. (2022b). "Cell Rearrangement Generates Pattern Emergence as a Function of Temporal Morphogen Exposure". *bioRxiv*, p. 2021.02.05.429898. DOI: 10.1101/2021.02.05.429898.

Goering, L. M. et al. (2003). "An Interacting Network of T-box Genes Directs Gene Expression and Fate in the Zebrafish Mesoderm". *Proceedings of the National Academy of Sciences of the United States of America* 100.16, pp. 9410–9415. DOI: 10.1073/pnas.1633548100. pmid: 12883008.

Gomez, C. et al. (2008). "Control of Segment Number in Vertebrate Embryos". *Nature* 454.7202 (7202), pp. 335–339. DOI: 10.1038/nature07020.

Goodman, J. and J. Weare (2010). "Ensemble Samplers with Affine Invariance". *Communications in Applied Mathematics and Computational Science* 5.1, pp. 65–80. DOI: 10.2140/camcos.2010.5.65.

Gorodilov, Y. N. et al. (2008). "Description of Embryogenesis of Atlantic Cod *Gadus Morhua*". *The Open Marine Biology Journal* 2.1, pp. 43–53. DOI: 10.2174/1874450800802010043.

Goto, H. et al. (2017). "FGF and Canonical Wnt Signaling Cooperate to Induce Paraxial Mesoderm from Tailbud Neuromesodermal Progenitors through Regulation of a Two-Step Epithelial to Mesenchymal Transition". *Development* 144.8, pp. 1412–1424. DOI: 10.1242/dev.143578.

Green, J. B. A. and J. Sharpe (2015). "Positional Information and Reaction-Diffusion: Two Big Ideas in Developmental Biology Combine". *Development* 142.7, pp. 1203–1211. DOI: 10.1242/dev.114991.

Griffin, K. J. P. et al. (1998). "Molecular Identification of Spadetail: Regulation of Zebrafish Trunk and Tail Mesoderm Formation by T-box Genes". *Development* 125.17, pp. 3379–3388. DOI: 10.1242/dev.125.17.3379.

Gültekin, H. and E. Yilmaz (2023). "The Embryonic and Larval Development of the Fire Mouth Cichlid (*Thorichthys Meeki* Brind, 1918)". *Aquatic Animal Reports* 1.2 (2), pp. 65–81. DOI: 10.5281/zenodo.8229232.

Haase, R. et al. (2020). "CLIJ: GPU-accelerated Image Processing for Everyone". *Nature Methods* 17.1 (1), pp. 5–6. DOI: 10.1038/s41592-019-0650-1.

Hagolani, P. F. et al. (2021). "On the Evolution and Development of Morphological Complexity: A View from Gene Regulatory Networks". *PLOS Computational Biology* 17.2, e1008570. DOI: 10.1371/journal.pcbi.1008570.

Hall, T. E., P. Smith, and I. A. Johnston (2004). "Stages of Embryonic Development in the Atlantic Cod *Gadus Morhua*". *Journal of Morphology* 259.3, pp. 255–270. DOI: 10.1002/jmor.10222.

Hamburger, V. and H. L. Hamilton (1951). "A Series of Normal Stages in the Development of the Chick Embryo". *Journal of Morphology* 88.1, pp. 49–92. DOI: 10.1002/jmor.1050880104.

Hammond, J. E., R. E. Baker, and B. Verd (2025). "Modularity of the Segmentation Clock and Morphogenesis". *eLife* 14. Ed. by C. Desplan, RP106316. DOI: 10.7554/eLife.106316.

Harris, C. R. et al. (2020). "Array Programming with NumPy". *Nature* 585.7825, pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

Harvey, S. A. et al. (2010). "No Tail Integrates Two Modes of Mesoderm Induction". *Development* 137.7, pp. 1127–1135. DOI: 10.1242/dev.046318.

Hilbrant, M., W. G. M. Damen, and A. P. McGregor (2012). "Evolutionary Crossroads in Developmental Biology: The Spider *Parasteatoda Tepidariorum*". *Development* 139.15, pp. 2655–2662. DOI: 10.1242/dev.078204.

Hinaux, H. et al. (2011). "A Developmental Staging Table for *Astyanax Mexicanus* Surface Fish and Pachón Cavefish". *Zebrafish* 8.4, pp. 155–165. DOI: 10.1089/zeb.2011.0713.

Holley, S. A. (2007). "The Genetics and Embryology of Zebrafish Metamerism". *Developmental Dynamics* 236.6, pp. 1422–1449. DOI: 10.1002/dvdy.21162.

Ikeda, T. et al. (2022). "Reemployment of Kupffer's Vesicle Cells into Axial and Paraxial Mesoderm via Transdifferentiation". *Development, Growth & Differentiation* 64.3, pp. 163–177. DOI: 10.1111/dgd.12774.

Iwamatsu, T. (2004). "Stages of Normal Development in the Medaka *Oryzias Latipes*". *Mechanisms of Development*. Medaka 121.7, pp. 605–618. DOI: 10.1016/j.mod.2004.03.012.

Jaeger, J. (2011). "The Gap Gene Network". *Cellular and molecular life sciences: CMLS* 68.2, pp. 243–274. DOI: 10.1007/s00018-010-0536-y. pmid: 20927566.

Jaeger, J. and N. Monk (2014). "Bioattractors: Dynamical Systems Theory and the Evolution of Regulatory Processes". *The Journal of Physiology* 592.11, pp. 2267–2281. DOI: 10.1113/jphysiol.2014.272385.

Jaeger, J. et al. (2004). "Dynamical Analysis of Regulatory Interactions in the Gap Gene System of Drosophila Melanogaster". *Genetics* 167.4, pp. 1721–1737. DOI: 10.1534/genetics.104.027334. pmid: 15342511.

Jouve, C., T. Iimura, and O. Pourquie (2002). "Onset of the Segmentation Clock in the Chick Embryo: Evidence for Oscillations in the Somite Precursors in the Primitive Streak". *Development* 129.5, pp. 1107–1117. DOI: 10.1242/dev.129.5.1107.

Jung, H. S. et al. (2015). "Embryonic Development of Haploid Olive Flounder *Paralichthys olivaceus*". *Korean Journal of Fisheries and Aquatic Sciences* 48.5, pp. 696–703. DOI: 10.5657/KFAS.2015.0696.

Jung, H. S. et al. (2018). "Embryonic Development of Convict Cichlid *Amatitlania nigrofasciata*". *Korean Journal of Fisheries and Aquatic Sciences* 51.4, pp. 420–425. DOI: 10.5657/KFAS.2018.0420.

Juntti, S. A., C. K. Hu, and R. D. Fernald (2013). "Tol2-Mediated Generation of a Transgenic Haplochromine Cichlid, *Astatotilapia Burtoni*". *PLOS ONE* 8.10, e77647. DOI: 10.1371/journal.pone.0077647.

Kanki, J. P. and R. K. Ho (1997). "The Development of the Posterior Body in Zebrafish". *Development* 124.4, pp. 881–893. DOI: 10.1242/dev.124.4.881.

Kassambara, A. (2023). *Ggpubr: 'ggplot2' Based Publication Ready Plots*. manual.

Kassambara, A. and F. Mundt (2020). *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. manual.

Kawamura, A. et al. (2005). "Groucho-Associated Transcriptional Repressor Ripply1 Is Required for Proper Transition from the Presomitic Mesoderm to Somites". *Developmental Cell* 9.6, pp. 735–744. DOI: 10.1016/j.devcel.2005.09.021. pmid: 16326386.

Kicheva, A. and J. Briscoe (2015). "Developmental Pattern Formation in Phases". *Trends in Cell Biology* 25.10, pp. 579–591. DOI: 10.1016/j.tcb.2015.07.006. pmid: 26410404.

Kimmel, C. B. et al. (1989). "A Mutation That Changes Cell Movement and Cell Fate in the Zebrafish Embryo". *Nature* 337.6205, pp. 358–362. DOI: 10.1038/337358a0.

Kimmel, C. B. et al. (1995). "Stages of Embryonic Development of the Zebrafish". *Developmental Dynamics* 203.3, pp. 253–310. DOI: 10.1002/aja.1002030302.

Kratochwil, C. F., M. M. Sefton, and A. Meyer (2015). "Embryonic and Larval Development in the Midas Cichlid Fish Species Flock (Amphilophus Spp.): A New Evo-Devo Model for the Investigation of Adaptive Novelties and Species Differences". *BMC Developmental Biology* 15.1, p. 12. DOI: 10.1186/s12861-015-0061-1.

Krol, A. J. et al. (2011). "Evolutionary Plasticity of Segmentation Clock Networks". *Development* 138.13, pp. 2783–2792. DOI: 10.1242/dev.063834.

Kumar, S. et al. (2022). "TimeTree 5: An Expanded Resource for Species Divergence Times". *Molecular Biology and Evolution* 39.8, msac174. DOI: 10.1093/molbev/msac174.

Laglbauer, B. J. et al. (2017). "Reproductive Synchrony in a Temperate Damselfish, *Chromis Limbata*". *acta ethologica* 20.3, pp. 297–311. DOI: 10.1007/s10211-017-0269-0.

Lawton, A. K. et al. (2013). "Regulated Tissue Fluidity Steers Zebrafish Body Elongation". *Development* 140.3, pp. 573–582. DOI: 10.1242/dev.090381.

Li, I.-J. et al. (2019). "Embryonic and Postembryonic Development of the Ornamental Twin-Tail Goldfish". *Developmental Dynamics* 248.4, pp. 251–283. DOI: 10.1002/dvdy.15.

Maia-Fernandes, A. C. et al. (2024). "Spatio-Temporal Dynamics of Early Somite Segmentation in the Chicken Embryo". *PloS One* 19.4, e0297853. DOI: 10.1371/journal.pone.0297853. pmid: 38635504.

Malinsky, M. et al. (2015). "Genomic Islands of Speciation Separate Cichlid Ecomorphs in an East African Crater Lake". *Science* 350.6267, pp. 1493–1498. DOI: 10.1126/science.aac9927.

Malinsky, M. et al. (2018). "Whole-Genome Sequences of Malawi Cichlids Reveal Multiple Radiations Interconnected by Gene Flow". *Nature Ecology & Evolution* 2.12 (12), pp. 1940–1955. DOI: 10.1038/s41559-018-0717-x.

Manning, A. J. and D. Kimelman (2015). "Tbx16 and Msgn1 Are Required to Establish Directional Cell Migration of Zebrafish Mesodermal Progenitors". *Developmental Biology* 406.2, pp. 172–185. DOI: 10.1016/j.ydbio.2015.09.001.

Manu et al. (2009a). "Canalization of Gene Expression and Domain Shifts in the Drosophila Blastoderm by Dynamical Attractors". *PLOS Computational Biology* 5.3, e1000303. DOI: 10.1371/journal.pcbi.1000303.

— (2009b). "Canalization of Gene Expression in the Drosophila Blastoderm by Gap Gene Cross Regulation". *PLOS Biology* 7.3, e1000049. DOI: 10.1371/journal.pbio.1000049.

Mara, A. et al. (2007). "Priming, Initiation and Synchronization of the Segmentation Clock by *deltaD* and *deltaC*". *Nature Cell Biology* 9.5, pp. 523–530. DOI: 10.1038/ncb1578.

Marconi, A. et al. (2023). "Morphological and Temporal Variation in Early Embryogenesis Contributes to Species Divergence in Malawi Cichlid Fishes". *Evolution & Development* 25.2, pp. 170–193. DOI: 10.1111/ede.12429.

Marconi, A. et al. (2024). "Genetic and Developmental Divergence in the Neural Crest Program between Cichlid Fish Species". *Molecular Biology and Evolution* 41.11, msae217. DOI: 10.1093/molbev/msae217.

Marin-Riera, M. et al. (2016). "Computational Modeling of Development by Epithelia, Mesenchyme and Their Interactions: A Unified Model". *Bioinformatics* 32.2, pp. 219–225. DOI: 10.1093/bioinformatics/btv527.

Marković, A., J. Briscoe, and K. M. Page (2024). "Dynamics of Positional Information in the Vertebrate Neural Tube". *Journal of The Royal Society Interface* 21.221, p. 20240414. DOI: 10.1098/rsif.2024.0414.

Martin, B. L. and D. Kimelman (2008). "Regulation of Canonical Wnt Signaling by Brachyury Is Essential for Posterior Mesoderm Formation". *Developmental Cell* 15.1, pp. 121–133. DOI: 10.1016/j.devcel.2008.04.013. pmid: 18606146.

— (2010). "Brachyury Establishes the Embryonic Mesodermal Progenitor Niche". *Genes & Development* 24.24, pp. 2778–2783. DOI: 10.1101/gad.1962910. pmid: 21159819.

McCormick, M. I. and I. V. Nechaev (2002). "Influence of Cortisol on Developmental Rhythms during Embryogenesis in a Tropical Damselfish". *Journal of Experimental Zoology* 293.5, pp. 456–466. DOI: 10.1002/jez.10138.

McDowall, R. M. (2008). "Jordan's and Other Ecogeographical Rules, and the Vertebral Number in Fishes". *Journal of Biogeography* 35.3, pp. 501–508. DOI: 10.1111/j.1365-2699.2007.01823.x.

McLaren, S. B. P. and B. J. Steventon (2021). "Anterior Expansion and Posterior Addition to the Notochord Mechanically Coordinate Zebrafish Embryo Axis Elongation". *Development* 148.18, dev199459. DOI: 10.1242/dev.199459.

McQueen, C. and M. Towers (2020). "Establishing the Pattern of the Vertebrate Limb". *Development* 147.17, dev177956. DOI: 10.1242/dev.177956.

Meijide, F. J. and G. A. Guerrero (2000). "Embryonic and Larval Development of a Substrate-Brooding Cichlid *Cichlasoma Dimerus* (Heckel, 1840) under Laboratory Conditions". *Journal of Zoology* 252.4, pp. 481–493. DOI: 10.1111/j.1469-7998.2000.tb01231.x.

Melby, A. E., D. Kimelman, and C. B. Kimmel (1997). "Spatial Regulation of Floating Head Expression in the Developing Notochord". *Developmental Dynamics* 209.2, pp. 156–165. DOI: 10.1002/(SICI)1097-0177(199706)209:2<156::AID-AJA2>3.0.CO;2-H.

Meurer, A. et al. (2017). "SymPy: Symbolic Computing in Python". *PeerJ Computer Science* 3, e103. DOI: 10.7717/peerj-cs.103.

Mongera, A. et al. (2018). "A Fluid-to-Solid Jamming Transition Underlies Vertebrate Body Axis Elongation". *Nature* 561.7723 (7723), pp. 401–405. DOI: 10.1038/s41586-018-0479-2.

Mongera, A. et al. (2023). "Mechanics of the Cellular Microenvironment as Probed by Cells in Vivo during Zebrafish Presomitic Mesoderm Differentiation". *Nature Materials* 22.1, pp. 135–143. DOI: 10.1038/s41563-022-01433-9.

Morin-Kensicki, E. M., E. Melancon, and J. S. Eisen (2002). "Segmental Relationship between Somites and Vertebral Column in Zebrafish". *Development* 129.16, pp. 3851–3860. DOI: 10.1242/dev.129.16.3851.

Morrow, Z. T. et al. (2017). "Tbx6l and Tbx16 Are Redundantly Required for Posterior Paraxial Mesoderm Formation during Zebrafish Embryogenesis". *Developmental Dynamics* 246.10, pp. 759–769. DOI: 10.1002/dvdy.24547.

Mourabit, S. et al. (2011). "Embryonic Development of the Self-Fertilizing Mangrove Killifish *Kryptolebias Marmoratus*". *Developmental Dynamics* 240.7, pp. 1694–1704. DOI: 10.1002/dvdy.22668.

Mousavi, S. E. and J. G. Patil (2022). "Stages of Embryonic Development in the Live-Bearing Fish, *Gambusia Holbrooki*". *Developmental Dynamics* 251.2, pp. 287–320. DOI: 10.1002/dvdy.388.

Mueller, R. L., C. Huang, and R. K. Ho (2010). "Spatio-Temporal Regulation of Wnt and Retinoic Acid Signaling by Tbx16/Spadetail during Zebrafish Mesoderm Differentiation". *BMC Genomics* 11.1, p. 492. DOI: 10.1186/1471-2164-11-492.

Naganathan, S. R. and A. C. Oates (2020). "Patterning and Mechanics of Somite Boundaries in Zebrafish Embryos". *Seminars in Cell & Developmental Biology*. 1. Cyclins Edited by Josep Clotet 107, pp. 170–178. DOI: 10.1016/j.semcdb.2020.04.014.

Narayanan, R., M. A. Mendieta-Serrano, and T. E. Saunders (2021). "The Role of Cellular Active Stresses in Shaping the Zebrafish Body Axis". *Current Opinion in Cell Biology*. Differentiation and Development 73, pp. 69–77. DOI: 10.1016/j.ceb.2021.06.002.

Nikaido, M. et al. (2002). "*Tbx24*, Encoding a T-box Protein, Is Mutated in the Zebrafish Somite-Segmentation Mutant *Fused Somites*". *Nature Genetics* 31.2, pp. 195–199. DOI: 10.1038/ng899.

Ninhaus-Silveira, A., F. Foresti, and A. de Azevedo (2006). "Structural and Ultrastructural Analysis of Embryonic Development of *Prochilodus Lineatus* (Valenciennes, 1836) (Characiforme; Prochilodontidae)". *Zygote* 14.3, pp. 217–229. DOI: 10.1017/S096719940600373X.

Oates, A. C., L. G. Morelli, and S. Ares (2012). "Patterning Embryos with Oscillations: Structure, Function and Dynamics of the Vertebrate Segmentation Clock".

Development (Cambridge, England) 139.4, pp. 625–639. DOI: 10.1242/dev.063735. pmid: 22274695.

Ortega, A. et al. (2024). "Effects of Temperature on Embryonic Development of Atlantic Bluefin Tuna (*Thunnus Thynnus*, L 1758) and Atlantic Bonito (*Sarda Sarda*, Bloch 1793)". *Fisheries Research* 277, p. 107066. DOI: 10.1016/j.fishres.2024.107066.

Osborn, D. P. S. et al. (2020). "Fgf-Driven Tbx Protein Activities Directly Induce Myf5 and Myod to Initiate Zebrafish Myogenesis". *Development* 147.8, dev184689. DOI: 10.1242/dev.184689.

Palmeirim, I. et al. (1997). "Avian Hairy Gene Expression Identifies a Molecular Clock Linked to Vertebrate Segmentation and Somitogenesis". *Cell* 91.5, pp. 639–648. DOI: 10.1016/S0092-8674(00)80451-1. pmid: 9393857.

Park, J. M. et al. (2017). "Egg Development and Larvae and Juveniles Morphology of Carp, *Cyprinus Carpio* in Korean [Sic]". *Development & Reproduction* 21.3, pp. 287–295. DOI: 10.12717/DR.2017.21.3.287. pmid: 29082344.

Pedregosa, F. et al. (2011). "Scikit-Learn: Machine Learning in Python". *Journal of Machine Learning Research* 12, pp. 2825–2830.

Peel, A. D., A. D. Chipman, and M. Akam (2005). "Arthropod Segmentation: Beyond the Drosophila Paradigm". *Nature Reviews. Genetics* 6.12, pp. 905–916. DOI: 10.1038/nrg1724. pmid: 16341071.

Piatkowska, A. M., S. E. Evans, and C. D. Stern (2021). "Cellular Aspects of Somite Formation in Vertebrates". *Cells & Development*. Quantitative Cell and Developmental Biology 168, p. 203732. DOI: 10.1016/j.cdev.2021.203732.

Pigliucci, M. (2010). "Genotype–Phenotype Mapping and the End of the 'Genes as Blueprint' Metaphor". *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1540, pp. 557–566. DOI: 10.1098/rstb.2009.0241.

Podrabsky, J. E. et al. (2017). "Embryonic Development of the Annual Killifish *Austrofundulus Limnaeus*: An Emerging Model for Ecological and Evolutionary Developmental Biology Research and Instruction". *Developmental Dynamics* 246.11, pp. 779–801. DOI: 10.1002/dvdy.24513.

Postlethwait, J. H. et al. (2016). "Embryogenesis and Early Skeletogenesis in the Antarctic Bullhead Notothen, *Notothenia Coriiceps*". *Developmental Dynamics* 245.11, pp. 1066–1080. DOI: 10.1002/dvdy.24437.

Powder, K. E. et al. (2015). "Constraint and Diversification of Developmental Trajectories in Cichlid Facial Morphologies". *EvoDevo* 6.1, p. 25. DOI: 10.1186/s13227-015-0020-8.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. manual. Vienna, Austria: R Foundation for Statistical Computing.

Rahman, S. et al. (2013). "Embryonic Development of *Clarias Batrachus* under the Influence of Aeration and Water Flow". *Ecoprint An International Journal of Ecology* 18. DOI: 10.3126/eco.v18i0.9395.

Rand, D. A. et al. (2021). "Geometry of Gene Regulatory Dynamics". *Proceedings of the National Academy of Sciences* 118.38, e2109729118. DOI: 10.1073/pnas.2109729118.

Rigaud, S. et al. (2024). *clEsperanto/Pyclesperanto: 0.16.0*. Version 0.16.0. Zenodo. DOI: 10.5281/zenodo.14199059.

Rohde, L. A. et al. (2024). "Cell-Autonomous Timing Drives the Vertebrate Segmentation Clock's Wave Pattern". *eLife* 13. DOI: 10.7554/eLife.93764.2.

Romanos, M. et al. (2021). "Cell-to-Cell Heterogeneity in Sox2 and Bra Expression Guides Progenitor Motility and Destiny". *eLife* 10. Ed. by M. E. Bronner and V. Wilson, e66588. DOI: 10.7554/eLife.66588.

Roux, N. et al. (2019). "Staging and Normal Table of Postembryonic Development of the Clownfish (*Amphiprion Ocellaris*)". *Developmental Dynamics* 248.7, pp. 545–568. DOI: 10.1002/dvdy.46.

Row, R. H. et al. (2016). "The Zebrafish Tailbud Contains Two Independent Populations of Midline Progenitor Cells That Maintain Long-Term Germ Layer Plasticity and Differentiate in Response to Local Signaling Cues". *Development* 143.2, pp. 244–254. DOI: 10.1242/dev.129015.

Rusinkiewicz, S. and M. Levoy (2001). "Efficient Variants of the ICP Algorithm". In: *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*. Third International Conference on 3-D Digital Imaging and Modeling. Quebec City, Que., Canada: IEEE Comput. Soc, pp. 145–152. DOI: 10.1109/IM.2001.924423.

Ruvinsky, I., L. M. Silver, and R. K. Ho (1998). "Characterization of the Zebrafish Tbx16 Gene and Evolution of the Vertebrate T-box Family". *Development Genes and Evolution* 208.2, pp. 94–99. DOI: 10.1007/s004270050158.

Sáez, M., J. Briscoe, and D. Rand (2022a). "Dynamical Landscapes of Cell Fate Decisions". *Interface Focus* 12. DOI: 10.1098/rsfs.2022.0002.

Sáez, M. et al. (2022b). "Statistically Derived Geometrical Landscapes Capture Principles of Decision-Making Dynamics during Cell Fate Transitions". *Cell Systems* 13.1, 12–28.e3. DOI: 10.1016/j.cels.2021.08.013.

Salis, P. et al. (2021). "The Real Nemo Movie: Description of Embryonic Development in *Amphiprion Ocellaris* from First Division to Hatching". *Developmental Dynamics* 250.11, pp. 1651–1667. DOI: 10.1002/dvdy.354.

Sambasivan, R. and B. Steventon (2021). "Neuromesodermal Progenitors: A Basis for Robust Axial Patterning in Development and Evolution". *Frontiers in Cell and Developmental Biology* 8, p. 607516. DOI: 10.3389/fcell.2020.607516. pmid: 33520989.

Santos, M. E., J. F. Lopes, and C. F. Kratochwil (2023). "East African Cichlid Fishes". *EvoDevo* 14.1, p. 1. DOI: 10.1186/s13227-022-00205-5.

Saud, H. A. et al. (2021). "Molecular Mechanisms of Embryonic Tail Development in the Self-Fertilizing Mangrove Killifish *Kryptolebias Marmoratus*". *Development* 148.24, dev199675. DOI: 10.1242/dev.199675.

Saunders, D., C. Camacho-Macorra, and B. Steventon (2025). "Spinal Cord Elongation Enables Proportional Regulation of the Zebrafish Posterior Body". *Development* 152.1, dev204438. DOI: 10.1242/dev.204438.

Sawada, A. et al. (2000). "Zebrafish Mesp Family Genes, Mesp-a and Mesp-b Are Segmentally Expressed in the Presomitic Mesoderm, and Mesp-b Confers the Anterior Identity to the Developing Somites". *Development* 127.8, pp. 1691–1702. DOI: 10.1242/dev.127.8.1691.

Schindelin, J. et al. (2012). "Fiji: An Open-Source Platform for Biological-Image Analysis". *Nature Methods* 9.7 (7), pp. 676–682. DOI: 10.1038/nmeth.2019.

Schroeder, M. D., C. Greer, and U. Gaul (2011). "How to Make Stripes: Deciphering the Transition from Non-Periodic to Periodic Patterns in *Drosophila* Segmentation". *Development* 138.14, pp. 3067–3078. DOI: 10.1242/dev.062141.

Schröter, C. et al. (2008). "Dynamics of Zebrafish Somitogenesis". *Developmental Dynamics* 237.3, pp. 545–553. DOI: 10.1002/dvdy.21458.

Schröter, C. et al. (2012). "Topology and Dynamics of the Zebrafish Segmentation Clock Core Circuit". *PLoS biology* 10.7, e1001364. DOI: 10.1371/journal.pbio.1001364. pmid: 22911291.

Schulte-Merker, S. et al. (1992). "The Protein Product of the Zebrafish Homologue of the Mouse T Gene Is Expressed in Nuclei of the Germ Ring and the Notochord of the Early Embryo". *Development* 116.4, pp. 1021–1032. DOI: 10.1242/dev.116.4.1021.

Schulte-Merker, S. et al. (1994). "No Tail (Ntl) Is the Zebrafish Homologue of the Mouse T (Brachyury) Gene". *Development* 120.4, pp. 1009–1015. DOI: 10.1242/dev.120.4.1009.

Setu, S. and T. Ajith Kumar (2010). "Spawning Behaviour and Embryonic Developmentof Regal Damsel Fish, *Neopomacentrus Cyanomus* (Bleeker, 1856)". *World Journal of fish and Marine sciences* 2, pp. 410–415.

Shah, G. et al. (2019). "Multi-Scale Imaging and Analysis Identify Pan-Embryo Cell Dynamics of Germlayer Formation in Zebrafish". *Nature Communications* 10.1, p. 5753. DOI: 10.1038/s41467-019-13625-0.

Simsek, M. F. et al. (2023). "Periodic Inhibition of ERK Activity Drives Sequential Somite Segmentation". *Nature* 613.7942, pp. 153–159. DOI: 10.1038/s41586-022-05527-x. pmid: 36517597.

Sofroniew, N. et al. (2025). *Napari: A Multi-Dimensional Image Viewer for Python*. Version v0.6.0a0. Zenodo. DOI: 10.5281/zenodo.15029515.

Sørensen, S. R. et al. (2016). "Ontogeny and Growth of Early Life Stages of Captive-Bred European Eel". *Aquaculture* 456, pp. 50–61. DOI: 10.1016/j.aquaculture.2016.01.015.

Soroldoni, D. et al. (2014). "A Doppler Effect in Embryonic Pattern Formation". *Science* 345.6193, pp. 222–225. DOI: 10.1126/science.1253089.

Spiess, K. et al. (2024). "Approximated Gene Expression Trajectories for Gene Regulatory Network Inference on Cell Tracks". *iScience* 27.9. DOI: 10.1016/j.isci.2024.110840. pmid: 39290835.

Steventon, B. et al. (2016). "Species-Specific Contribution of Volumetric Growth and Tissue Convergence to Posterior Body Elongation in Vertebrates". *Development* 143.10, pp. 1732–1741. DOI: 10.1242/dev.126375.

Stooke-Vaughan, G. A. and O. Campàs (2018). "Physical Control of Tissue Morphogenesis across Scales". *Current Opinion in Genetics & Development*. Developmental Mechanisms, Patterning and Evolution 51, pp. 111–119. DOI: 10.1016/j.gde.2018.09.002.

Stooke-Vaughan, G. A. et al. (2025). "The Physical Roles of Different Posterior Tissues in Zebrafish Axis Elongation". *Nature Communications* 16.1, p. 1839. DOI: 10.1038/s41467-025-56334-7.

Strogatz, S. (2019). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Second edition, first issued in hardback. A Chapman & Hall Book. Boca Raton London New York: CRC Press. 513 pp.

Svardal, H. et al. (2020). "Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation". *Molecular Biology and Evolution* 37.4, pp. 1100–1113. DOI: 10.1093/molbev/msz294.

Talbot, W. S. et al. (1995). "A Homeobox Gene Essential for Zebrafish Notochord Development". *Nature* 378.6553, pp. 150–157. DOI: 10.1038/378150a0.

Tavolga, W. N. (1949). "Embryonic Development of the Platyfish (Platypoecilus), the Swordtail (Xiphophorus) and Their Hybrids". PhD thesis. United States – New York: New York University. 75 pp.

Taylor, S. E. and P. K. Dearden (2022). "The *Nasonia* Pair-Rule Gene Regulatory Network Retains Its Function over 300 Million Years of Evolution". *Development* 149.5, dev199632. DOI: 10.1242/dev.199632.

Tenin, G. et al. (2010). "The Chick Somitogenesis Oscillator Is Arrested before All Paraxial Mesoderm Is Segmented into Somites". *BMC Developmental Biology* 10.1, p. 24. DOI: 10.1186/1471-213X-10-24.

Terasaki, H. et al. (2006). "Transgenic Analysis of the Medaka *Mesp-b* Enhancer in Somitogenesis". *Development, Growth & Differentiation* 48.3, pp. 153–168. DOI: 10.1111/j.1440-169X.2006.00853.x.

Thisse, B. et al. (2004). "Spatial and Temporal Expression of the Zebrafish Genome by Large-Scale in Situ Hybridization Screening". *Methods in Cell Biology* 77, pp. 505–519. DOI: 10.1016/s0091-679x(04)77027-2. pmid: 15602929.

Thomson, L., L. Muresan, and B. Steventon (2021). "The Zebrafish Presomitic Mesoderm Elongates through Compaction-Extension". *Cells & Development*. Quantitative Cell and Developmental Biology 168, p. 203748. DOI: 10.1016/j.cdev.2021.203748.

Toh, K. et al. (2022). "Zebrafish Neuromesodermal Progenitors Undergo a Critical State Transition *in Vivo*". *iScience* 25.10, p. 105216. DOI: 10.1016/j.isci.2022.105216.

Tsai, H.-Y. et al. (2013). "Embryonic Development of Goldfish (*Carassius Auratus*): A Model for the Study of Evolutionary Change in Developmental Mechanisms by Artificial Selection". *Developmental Dynamics* 242.11, pp. 1262–1283. DOI: 10.1002/dvdy.24022.

Turner, G. F. (2007). "Adaptive Radiation of Cichlid Fish". *Current Biology* 17.19, R827–R831. DOI: 10.1016/j.cub.2007.07.026. pmid: 17925206.

Tzouanacou, E. et al. (2009). "Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis". *Developmental Cell* 17.3, pp. 365–376. DOI: 10.1016/j.devcel.2009.08.002.

Uriu, K. et al. (2017). "A Framework for Quantification and Physical Modeling of Cell Mixing Applied to Oscillator Synchronization in Vertebrate Somitogenesis". *Biology Open* 6.8, pp. 1235–1244. DOI: 10.1242/bio.025148.

Uriu, K. et al. (2021). "From Local Resynchronization to Global Pattern Recovery in the Zebrafish Segmentation Clock". *eLife* 10. Ed. by D. Y. Stainier and I. Riedel-Kruse, e61358. DOI: 10.7554/eLife.61358.

Urum, M. and E. Preger-Ben Noon (2025). "The Developmental and Genetic Basis of Male Genitalia Evolution in Drosophilids". *Current Opinion in Insect Science* 68, p. 101335. DOI: 10.1016/j.cois.2025.101335.

Van Eeden, F. J. M. et al. (1996). "Mutations Affecting Somite Formation and Patterning in the Zebrafish, Danio Rerio". *Development* 123.1, pp. 153–164. DOI: 10.1242/dev.123.1.153.

Venzin, O. F. et al. (2023). "Clock Driven Waves of Tbx6 Expression Prefigure Somite Boundaries". *bioRxiv*, p. 2023.11.09.566373. DOI: 10.1101/2023.11.09.566373.

Verd, B., A. Crombach, and J. Jaeger (2014). "Classification of Transient Behaviours in a Time-Dependent Toggle Switch Model". *BMC Systems Biology* 8.1, p. 43. DOI: 10.1186/1752-0509-8-43.

Verd, B., A. Crombach, and J. Jaeger (2017). "Dynamic Maternal Gradients Control Timing and Shift-Rates for *Drosophila* Gap Gene Expression". *PLOS Computational Biology* 13.2, e1005285. DOI: 10.1371/journal.pcbi.1005285.

Verd, B., N. A. Monk, and J. Jaeger (2019). "Modularity, Criticality, and Evolvability of a Developmental Gene Regulatory Network". *eLife* 8. Ed. by L. Altenberg and P. J. Wittkopp, e42832. DOI: 10.7554/eLife.42832.

Verd, B. et al. (2018). "A Damped Oscillator Imposes Temporal Order on Posterior Gap Gene Expression in *Drosophila*". *PLOS Biology* 16.2, e2003174. DOI: 10.1371/journal.pbio.2003174.

Virtanen, P. et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Vrat, V. (1949). "Reproductive Behavior and Development of Eggs of the Three-Spined Stickleback (*Gasterosteus Aculeatus*) of California". *Copeia* 1949.4, pp. 252–260. DOI: 10.2307/1438375. JSTOR: 1438375.

Walt, S. van der et al. (2014). "Scikit-Image: Image Processing in Python". *PeerJ* 2, e453. DOI: 10.7717/peerj.453.

Ward, L., S. E. Evans, and C. D. Stern (2017). "A Resegmentation-shift Model for Vertebral Patterning". *Journal of Anatomy* 230.2, pp. 290–296. DOI: 10.1111/joa.12540. pmid: 27580767.

Warga, R. M. et al. (2013). "Zebrafish Tbx16 Regulates Intermediate Mesoderm Cell Fate by Attenuating Fgf Activity". *Developmental Biology* 383.1, pp. 75–89. DOI: 10.1016/j.ydbio.2013.08.018.

Webb, A. B. et al. (2016). "Persistence, Period and Precision of Autonomous Cellular Oscillators from the Zebrafish Segmentation Clock". *eLife* 5. Ed. by T. T. Whitfield, e08438. DOI: 10.7554/eLife.08438.

Webb, A. B. et al. (2014). "Generation of Dispersed Presomitic Mesoderm Cell Cultures for Imaging of the Zebrafish Segmentation Clock in Single Cells". *Journal of Visualized Experiments : JoVE* 89, p. 50307. DOI: 10.3791/50307. pmid: 25078855.

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., D. Vaughan, and M. Girlich (2024). *Tidyr: Tidy Messy Data*. manual.

Wickham, H. et al. (2023). *Dplyr: A Grammar of Data Manipulation*. manual.

Windner, S. E. et al. (2015). "Tbx6, Mesp-b and Ripply1 Regulate the Onset of Skeletal Myogenesis in Zebrafish". *Development (Cambridge, England)* 142.6, pp. 1159–1168. DOI: 10.1242/dev.113431. pmid: 25725067.

Yamamoto, K., K. Yamauchi, and S. Kasuga (1975). "On the Development of the Japanese Eel, *Anguilla Japonica*". *Bulletin of the Japanese Society of Scientific Fisheries* 41.1, pp. 21–28. DOI: 10.2331/suisan.41.21.

Yasuhiko, Y. et al. (2008). "Functional Importance of Evolutionally Conserved Tbx6 Binding Sites in the Presomitic Mesoderm-Specific Enhancer of *Mesp2*". *Development* 135.21, pp. 3511–3519. DOI: 10.1242/dev.027144.

Ye, Z. and D. Kimelman (2020). "Hox13 Genes Are Required for Mesoderm Formation and Axis Elongation during Early Zebrafish Development". *Development* 147 (dev185298). DOI: 10.1242/dev.185298.

Ye, Z. et al. (2021). "Identification of *in Vivo* Hox13-binding Sites Reveals an Essential Locus Controlling Zebrafish *Brachyury* Expression". *Development* 148.11, dev199408. DOI: 10.1242/dev.199408.

Yokoi, H. et al. (2007). "Mutant Analyses Reveal Different Functions of *Fgfr1* in Medaka and Zebrafish despite Conserved Ligand–Receptor Relationships". *Developmental Biology* 304.1, pp. 326–337. DOI: 10.1016/j.ydbio.2006.12.043.

Yuikawa, T. et al. (2021). "Involvement of Oct4-type Transcription Factor Pou5f3 in Posterior Spinal Cord Formation in Zebrafish Embryos". *Development, Growth & Differentiation* 63.6, pp. 306–322. DOI: 10.1111/dgd.12742.

Zagorski, M. et al. (2017). "Decoding of Position in the Developing Neural Tube from Antiparallel Morphogen Gradients". *Science (New York, N.Y.)* 356.6345, pp. 1379–1383. DOI: 10.1126/science.aam5887. pmid: 28663499.

Zaucker, A. et al. (2014). "Description of Embryonic Development of Spotted Green Pufferfish (*Tetraodon Nigroviridis*)". *Zebrafish* 11.6, pp. 509–517. DOI: 10.1089/zeb.2014.0984.

Zhijing, S. U. N. et al. (2014). "Observation of embryonic and larval development of blood parrot cichlid fish *Cichlasoma synspilum*×*C.citrinellum*". *South China Fisheries Science* 10.3, pp. 38–46. DOI: 10.3969/j.issn.2095-0780.2014.03.006.

Zhou, Q.-Y., J. Park, and V. Koltun (2018). "Open3D: A Modern Library for 3D Data Processing". arXiv: 1801.09847.

Appendices

Appendix A: Experimental appendices

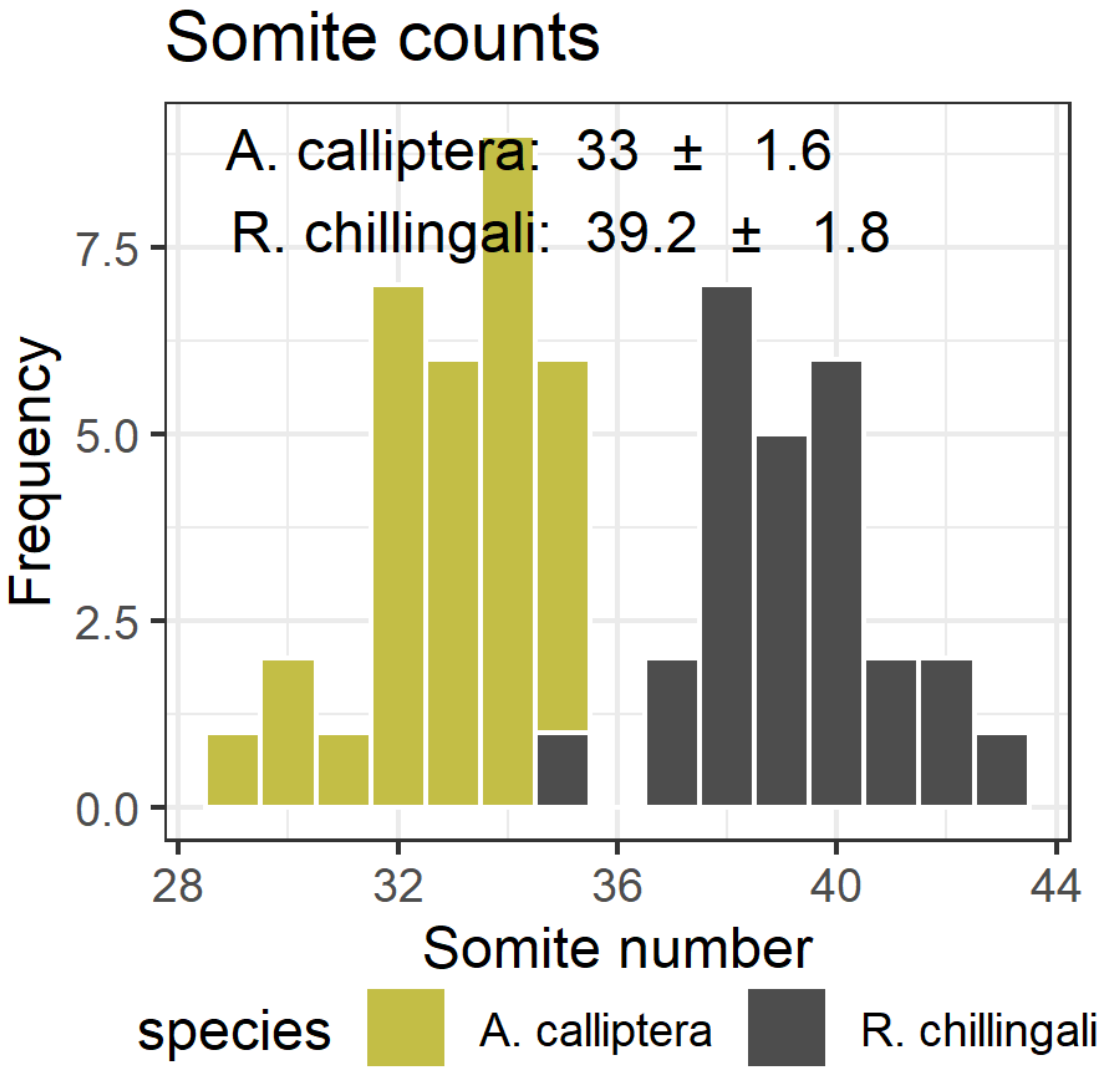


Figure A.1: Somite numbers in *A. calliptera* and *R. chillingali* A. Number of somite counts after the completion of segmentation in both *A. calliptera* and *R. chillingali*.

Table A.1: Statistics on PSM metrics

	variable stage					comparison		variable stage + species						
	R^2	slope	95 CI	F statistic	p-value	RSS model 1	RSS model 2	R^2	slope	95 CI	species difference	95 CI species diff.	p count	p species
PSM length	0.71	-13.02	-14.53 — -11.51	285.85	4.4e-33	570125	394490	0.80	-13.68	-14.95 — -12.4	77.95	56.6 — 99.3	1.1e-39	8.4e-11
notocord length	0.76	-10.13	-11.17 — -9.1	370.54	6.5e-38	266211	211735	0.81	-10.50	-11.43 — -9.56	43.41	27.77 — 59.05	4.4e-42	3.1e-07
PSM width	0.43	-3.43	-4.15 — -2.72	87.78	7.2e-16	129073	123839	0.45	-3.55	-4.26 — -2.83	13.45	1.49 — 25.42	3.3e-16	0.03
PSM depth	0.12	1.11	0.55 — 1.67	15.12	1.7e-04	78637	77451	0.13	1.06	0.49 — 1.62	6.41	-3.05 — 15.87	1.6e-04	n.s.
MPZ length	0.34	-2.89	-3.63 — -2.15	58.49	6.5e-12	137317	102837	0.50	-3.18	-3.83 — -2.53	34.54	23.63 — 45.44	1.6e-14	8.7e-09
Somite length	0.48	-0.75	-0.89 — -0.61	108.45	2.5e-18	5022	4797	0.51	-0.73	-0.87 — -0.59	-2.78	-5.14 — -0.43	9.5e-19	0.02
notocord width	0.30	-0.54	-0.69 — -0.39	50.77	9.4e-11	5560	5503	0.31	-0.55	-0.7 — -0.4	1.41	-1.12 — 3.93	9.4e-11	n.s.
length:width	0.16	-0.03	-0.04 — -0.01	21.73	8.4e-06	29	27	0.20	-0.03	-0.04 — -0.02	0.23	0.05 — 0.4	5.6e-06	0.01
length:depth	0.37	-0.35	-0.43 — -0.26	68.39	2.5e-13	1680	1680	0.37	-0.35	-0.43 — -0.26	0.14	-1.25 — 1.54	3.2e-13	n.s.
depth:width	0.23	-0.11	-0.15 — -0.07	33.69	5.7e-08	356	354	0.23	-0.11	-0.15 — -0.07	-0.25	-0.89 — 0.39	6.1e-08	n.s.

Table A.2: Probability of interaction between species and somite stage

	R^2	p species	p count	p interaction
PSM length	0.80	1.0e-10	2.3e-39	n.s.
notocord length	0.81	3.4e-07	1.0e-41	n.s.
PSM width	0.51	0.02	3.0e-17	6.2e-04
MPZ length	0.50	9.9e-09	2.0e-14	n.s.
Somite length	0.51	0.02	1.3e-18	n.s.
length:width	0.28	0.01	2.0e-06	4.6e-04

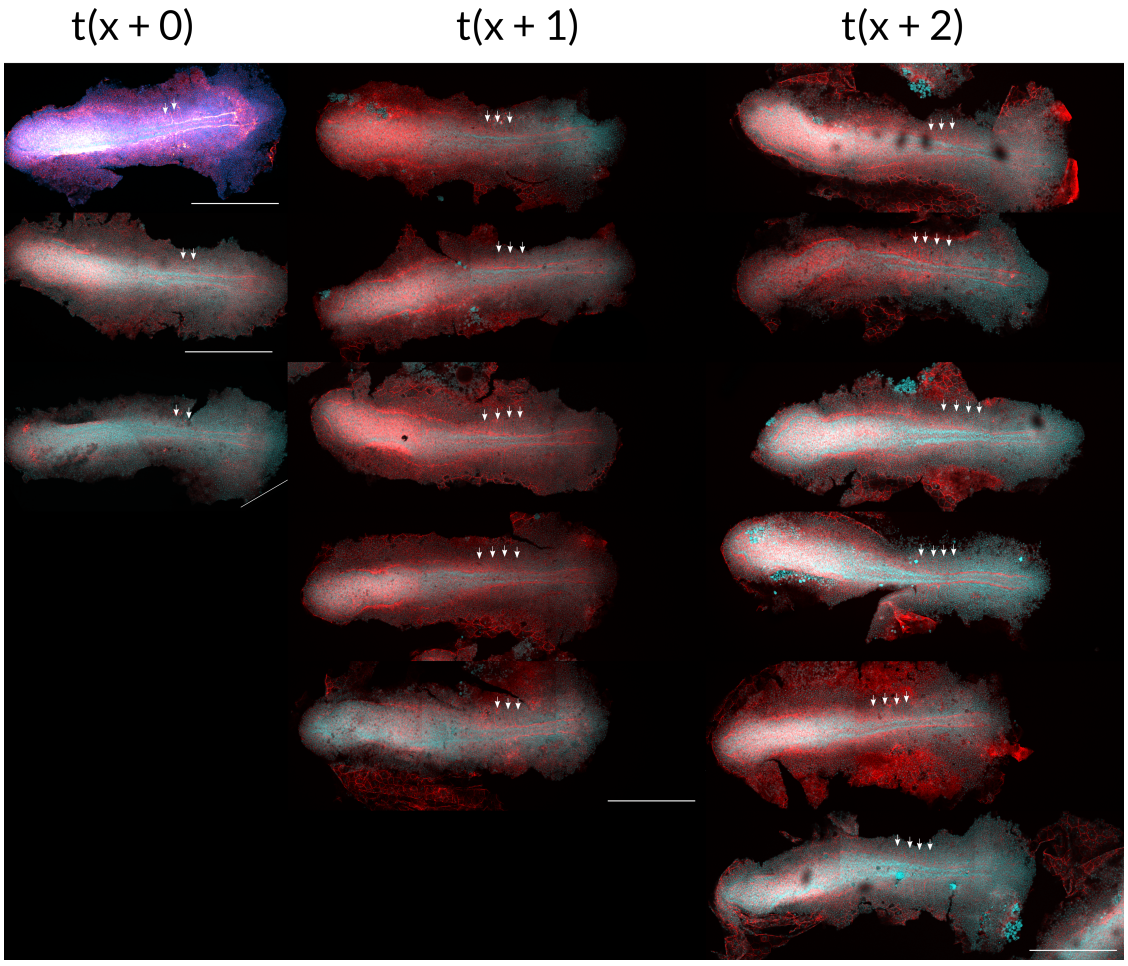


Figure A.2: *A. calliptera* somite numbers at the onset of segmentation. Embryos were fixed at hourly intervals, before being dissected and stained with DAPI (cyan) and phalloidin (red). Somites are indicated with arrows. Scale bar: 500 μm. All embryos are orinted anterior left.

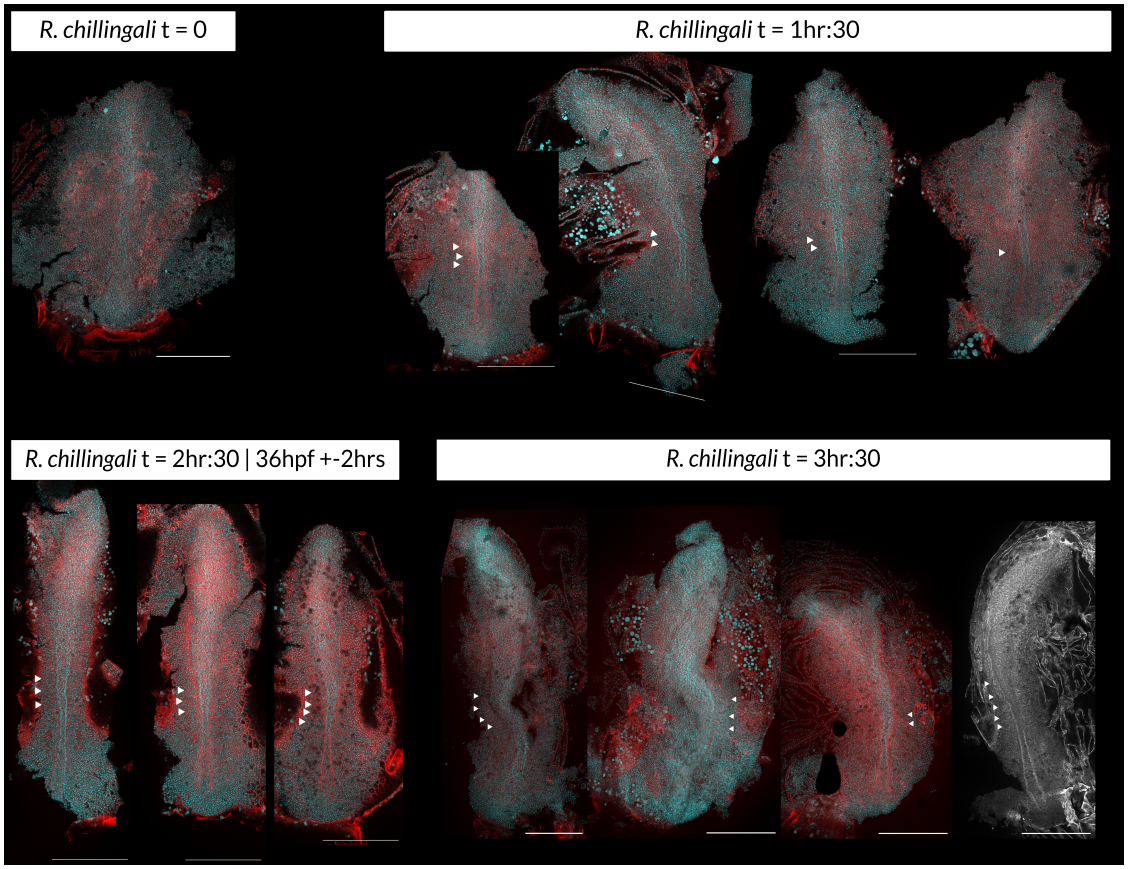


Figure A.3: *R. chillingali* somite numbers at the onset of segmentation. Embryos were fixed at hourly intervals, before being dissected and stained with DAPI (cyan) and phalloidin (red). Somites are indicated with arrows. Scale bar: 500 μm. All embryos are orinted anterior left.

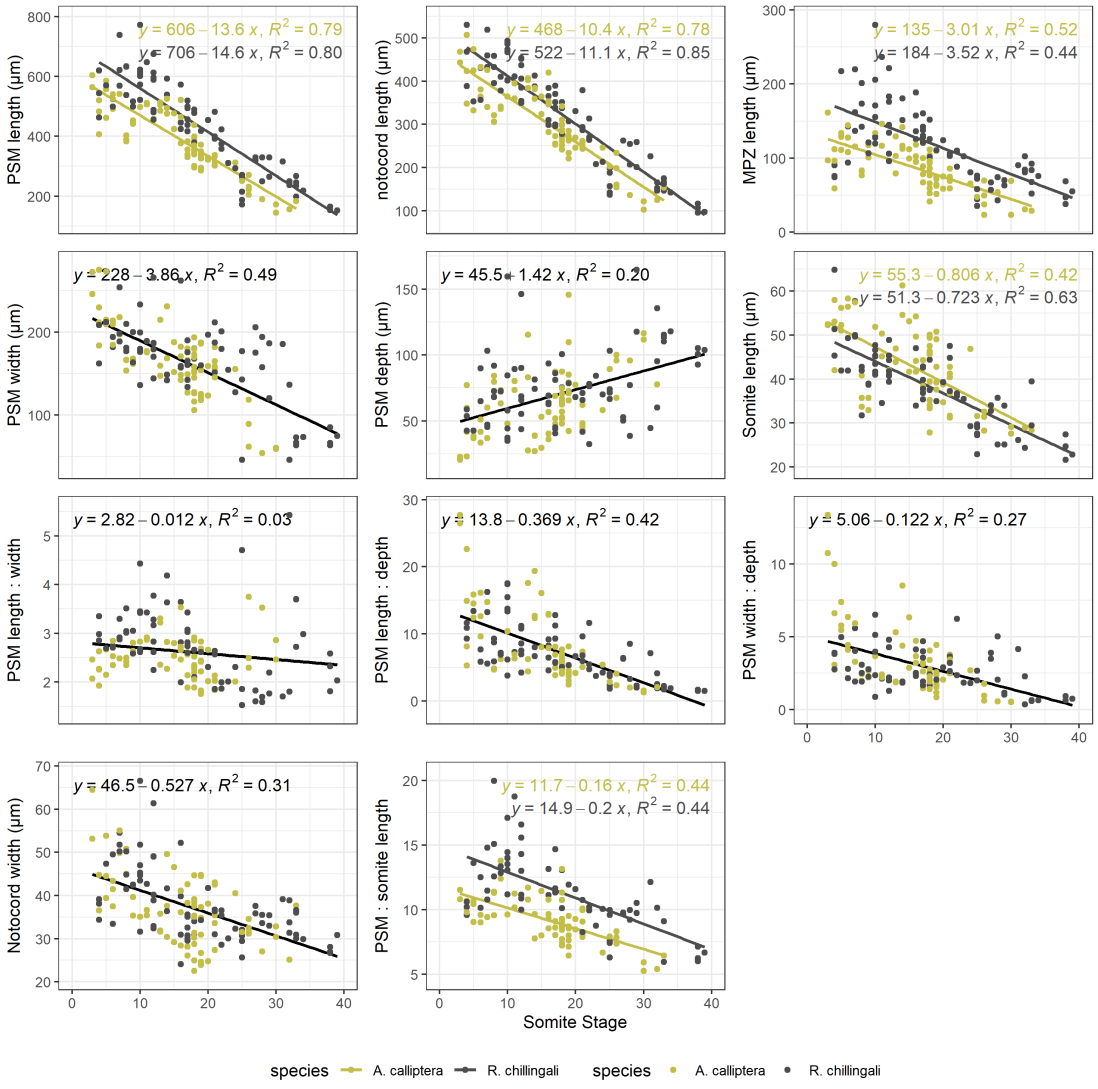


Figure A.4: PSM measurements used for PCA. Solid black lines indicate the trendline for measurements where species did not statistically differ, and different coloured lines indicate species-specific trendlines. *A. calliptera*: yellow. *R. chillingali*: grey.