Estimating complexity and adaptation in the embryo: a statistical developmental biology approach

Irepan Salvador-Martínez

Institute of Biotechnology

and

Division of Genetics
Department of Biosciences
Faculty of Biological and Environmental Sciences

and

Doctoral Programme in Integrative Life Science
University of Helsinki

ACADEMIC DISSERTATION

To be presented in ... text of a long permission notice. Text of a long permission notice. Text of a long permission notice. Text of a long permission notice. Text of a long permission notice. Text of a long permission notice.

Helsinki 2016

Supervisor
Isaac Salazar-Ciudad, University of Helsinki, Finland

Pre-examiners
Pavel Tomancak, Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany
Gregor Bucher, Georg-August-University Göttingen, Germany

Opponent
Johannes Jäger, Konrad Lorenz Institute, Austria

Custos
Name, University, Country

Advisory Committee
Jukka Jernvall, University of Helsinki
Osamu Shimmi, University of Helsinki
Mikael Fortelius, University of Helsinki

ISSN 1238-8645
ISBN 000-00-0000-0 (paperback)
ISBN 000-00-0000-0 (PDF)
Helsinki 2016
Unigrafia

Acknowledgements

Contents

List of publications

Abbreviations

α	Proportion of adaptive nucleotide substitutions
ω	dN/dS ratio
ω_α	Proportion of adaptive non-synonymous substitutions
N_e	Effective population size
s	Selection coefficient
2D	two-dimensional
3D	three-dimensional
A/P	anterior/posterior
Adh	Alcohol dehydrogenase
ANOVA	Analysis of Variance
CNS	Central Nervous System
D/V	dorsal/ventral
DFE	Distribution of Fitness Effects
dN	Non-synonymous divergence per site
DNA	Deoxyribonucleic acid
DNE	Dirichlet Normal Energy
dS	Synonymous divergence per site
evo-devo	Evolutionary developmental biology
GF	Growth Factor
GRN	Gene Regulatory Network
HG	Hourglass model
IQR	Inter Quartile Range
KW	Kruskal-Wallis test
miRNAs	microRNAs
MKT	MacDonald-Kreitman test
RNA	Ribonucleic acid
RTK	receptor tyrosine kinase

SEM	Standard error of the mean
SFS	Site Frequency Spectrum
SIGs	Signaling molecules
SNP	Single nucleotide polymorphism
ST	Synexpression territories
TF	Transcription Factor

Abstract

Embryonic development has amazed scientists and philosophers for centuries. Many reasons have been evoked for the perceivable complexity increase that transforms a single cell into a larva or an adult. During this process, at the level of gene expression, it its assumed that genes change from being expressed in large spatial domains of the embryo in early development to spatially restricted domains (e.g., tissues, cells) in late development. For some crucial developmental genes (e.g., Hox genes), the spatio-temporal expression dynamics has been thoroughly described. It is not clear however, if the dynamics are similar for the rest of the genes, or if there are differences between different types of genes or differences between species.

To some authors, adaptive reasons could be the cause for such increase in complexity. Adaptations can be estimated with molecular evolution methods, analysing the genes expressed in different developmental stages or regions in the embryo. These methods estimate adaptive changes at the DNA sequence level, assuming that a positive selected site would show less variance than other sites evolving neutrally. Importantly, different developmental stages might show distinct levels of positive or stabilizing selection, that could be related to inter-specific divergence patterns proposed by the von Baer's laws or the hourglass model (HG). The former states that the development of two species would be very similar in early stages and increasingly divergent in subsequent stages. In contrast, the latter states that development is less divergent at mid development.

In here, I analysed gene expression information to estimate both complexity and adaptation in the embryo using a statistical approach. To measure complexity, I analysed publicly available gene expression data (thousands of in situ hybridization experiments)in *Drosophila melanogaster* and *Ciona intestinalis*, and developed quantitative measures of complexity. To estimate adaptation, I combined the *D. melanogaster* transcriptomic data with populations genomic data (from the DGRP project). With the DFE-alpha method (which uses coding-region polymorphism and divergence to estimate the proportion of adaptive changes) , I charted a spatial map of adaptation of the fruit fly embryo's anatomy. Finally, I analysed the pattern of positive selection through the entire life cycle of *D. melanogaster* and how it correlated with specific genomic determinants (e.g., gene structure, codon bias)

Briefly, I found that *Drosophila* and *Ciona* complexity increases non-linearly with the major change before and after gastrulation, respectively, using three mathematical measures of gene expression compartmentalization in space. In both species, transcription factors and signalling molecules showed an earlier compartmentalization, consistent with their proposed leading role in pattern formation. In *Drosophila*, gonads and head showed high adaptation during embryogenesis, although pupa and adult male stages exhibit the highest levels of adaptive change, and mid and late embryonic stages show high conservation, showing an HG pattern. Furthermore, I propose that the explanation for the lack of conservation in the early stages and the HG-like pattern could be a relaxation of natural selection and gene structure complexity, respectively.

Review of the literature

This work is based and uses concepts from three main biology fields: developmental biology, evolutionary biology and genetics. Nowadays, the union of these scientific fields form multiple research programmes. Only in evolutionary developmental biology (evo-devo), the explicit union of the first two fields, at least four major research programmes have been recognized (Müller, 2007). However, these fields have not always gotten along well. Some decades ago, there was a clear conceptual and epistemological separation between evolutionary biology and developmental biology, even when embryology (which slowly transformed into developmental biology in the middle of the 20th century, see Horder, 2010) was considered crucial for the study of evolution in the 19th century.

In the following section, I will make a brief introduction of the scientific and philosophical origins of developmental biology, with special attention to its relations with evolutionary biology and genetics (for a comprehensive review on this issue, see: Amundson, 2005; Gilbert, 1991). Before that, it might be useful to define what development is, so firstly, I will address this apparently simple question.

1.1 What is development?

It seems that there is no unique or straightforward answer to this question. Sometimes, the study of development is implicitly considered to be the same as the the study of embryology (Horder, 2010). This could be problematic when considering organisms with complex life cycles. For example, holometabolous insects, in addition to embryonic development, undergo a complete metamorphosis (from pupa to adult), a process that could be considered a second embryonic development.

Currently, the most common definition of development refers to the set of processes through which an egg is transformed into an adult (Horder, 2010; Minelli, 2011). Already in 1880, Ernst Haeckel defined development in similar terms: "individual development, or the ontogenesis of every single organism, from the egg to the complete form is nothing but a growth attended by a series of diverging and progressive changes" (Haeckel, 1880).

Some authors criticize this egg-to-adult view to be an "adultocentric" view of development, and suggest instead to consider within the boundaries of development the whole life cycle of an organism (Gilbert, 2011; Minelli, 2011). Julian S. Huxley and Gavin R. de Beer said that development "is not merely an affair of early stages; it continues, though usually at a diminishing rate, throughout life" (Huxley and De Beer, 1963).

There have been recent attempts to construct a broader concept of development (Griesemer, 2014; Moczek, 2014; Pradeu, 2014) For example, Armin P. Moczek defines development as "the sum of all processes and interacting components that are required to allow organismal form and function, on all levels of biological organization, to come into being" (Moczek, 2014). The main challenge on adopting a new concept of development which is more inclusive, is to maintain its intuitiveness and applicability in scientific research.

Throughout this dissertation I will use the "common view" of development (Minelli, 2014), that considers the egg and the adult as the start and end of individual development respectively. However, and mainly for practical reasons, the major part of the analysis presented here (articles 1-3) is (ARE?) restricted to embryonic development.

1.2 On the history of developmental biology

In the last decades the scientific community has witnessed the flourishing of developmental biology. Since the 1980's crucial discoveries (Gilbert, 1998) have not only improved our understanding of the developmental process, but also changed the perspective of the explanatory role of development in biology.

However, developmental biology can not be considered a young scientific discipline, as its roots come from centuries ago, back from embryology and anatomy. The summary presented here grasps only the surface of this history, for further and deeper lecture, see (Gilbert, 1991; Amundson, 2005; Hall, 1999)

1.2.1 Aristotle

Before the 19th century, the major single contributor in the study of embryology was Aristotle. Some of his most important contributions to embryology are:

i. He organized and classified animals accordingly to their embryonic development after careful observation of the development in many species (Aristotle, 1979). Because of this, he can be considered the first comparative embryologist (Needham, 1959).
ii. For him, any developmental process was driven by "internal causes" that required a "soul" to guide it.
iii. He clearly defined two opposite theories of development, preformationism and epigenesis, from which he supported the latter.

After Aristotle, the preformationism-epigenesis debate would last centuries attracting many of the most important philosophers and naturalists.

1.2.2 The preformationism-epigenesis debate

Until the 18th century, supporters of epigenesis (like Wolffs and ??) saw development as starting from a formless embryo, with its form arising following a "vital" force (Amundson, 2005). During the 18th century, however, many rejected any vital force to explain development, leaving preformationism as the only possible solution to the problem of development (Jacob, 1973).

Defenders of preformationism, like Swammerdamm, said that the adult form was already present in the early embryo (or "germ") and that the process of development was just the unfolding of this pre-existent form (Amundson, 2005). Following this argumentation, it was said that all the germs in the future, present and past existed since the creation, nested one inside of another like Russian dolls, just waiting to be activated (Jacob, 1973).

Preformationism remained to be the main accepted idea in the 18th century, but some saw its consequences as impossible. Buffon refuted preformationism with a single calculation. He calculated the size that preformed germs of many future subsequent generations should have: for a sixth generation, he calculated, the germ should be smaller than the smallest possible atom (Buffon, 1807).

1.2.3 Haeckel, von Baer and the *Naturphilosophie*

In the 19th century, important contributions to embryology were made by advocates of *Naturphilosophie*. This philosophilcal movement, based in Kant and Goethe's ideas, aimed to classify nature into categories or classes. Among their classification efforts, they classified embryological phenomena and draw analogies between embryos of different taxonomic groups (Horder, 2010; Ghiselin, 2005).

The first pattern to be recognized, when comparing developmental trajectories of different species, was the Meckel-Serres law. This law, named so by E. S. Russell after two of their main proponents: Étienne Serres and Johann Friedrich Meckel (Russell, 1916), proposed that embryos followed a linear succession following the scala naturae. In this view (influenced by the *Naturphilosophie*), the embryonic development of a higher organism would be a succession of adult forms of lower organisms (Russell, 1916; Amundson, 2005).

Karl Ernst von Baer

K. E. von Baer, an Estonian naturalist considered the father of comparative embryology (Russell, 1916), refuted the Meckel-Serres law and formulated his own, known as von Baer's laws (von Baer, 1828). von Baer's laws stated that general characteristics develop before special characteristics (first law) and that, opposed to the Meckel-Serres law, the embryo of a "higher" animal never resembles the adult of another animal form, but only his embryo (fourth law). Importantly, von Baer's views were not evolutionary. The resemblance between developmental trajectories of different species was for him only a reflection of their relationship in the Natural System (Amundson, 2005). Ironically, in his "Origin of species", Darwin used and reinterpreted von Baer's observations on embryonic stages in different species to support common ancestry and therefore, evolution (Darwin, 1859).

Ernst Haeckel

Ernst Haeckel was one of the first who made explicit hypothesis about the connection between development and evolutionary patterns. He supported Darwinism and, in what is known as Haeckel's "Biogenetic Law", said that development (or ontogeny), is a brief summary of the slow and long phylogeny (Haeckel, 1874). In his view, similar to the parallelism view, a "higher" organism would pass through a series of conserved developmental stages that represent ancestral forms (this view is also known as the "recapitulation theory"). However, in contrast with the Meckel-Serres law, he recognized that this recapitulation was almost never complete, due to evolutionary modifications in development. He also classified two types of change in development, "heterochrony" and "heterotopy", concepts introduced by him that since then have been crucial in any discussion of the relationship between development and evolution (Horder, 2013):

> *"The falsification of the original course of development is based to a great extent on a gradually occurring displacement of the phenomena, which has been effected slowly over many millennia, by adapting to the changed conditions of embryonic existence. This displacement can affect both their location and time of appearance. Those former we call heterotopy, the latter heterochrony." (Haeckel, 1903).*

Haeckel's views were more complex than usually acknowledged (Richardson and Keuck, 2002). In fact, he said that it was not that all the mammalian eggs were the same, it

was just that with the available tools was impossible to detect the subtle, individual differences, "which are to be found only in the molecular structure" (Haeckel, 1903).

Now is evident that none of von Baer's or Haeckel's hypothesis can be considered "laws", as they are not universal. Nevertheless, the works of both Haeckel and von Baer represented the foundations of the comparative embryology field, which is in turn the basis of the modern evolutionary developmental biology (evo-devo).

1.2.4 *Entwicklungsmechanik*

Despite the great advances described above, embryology remained a descriptive science. It was not until the end of the 19th century when experimental embryology was born under the name of *Entwicklungsmechanik* (from the german "developmental mechanics"), with the experiments of Roux and Driesch (this is however a simplified version of the origins of *Entwicklungsmechanik*, for a more complete one, see Maienschein, 1991).

In the 1880's Wilhelm Roux, one of the co-founders of (and coiner of the term) *Entwicklungsmechanik*, performed a simple experiment to test Weismann's theory of inheritance. This theory stated that when a cell divides during development, "chromatin determinants" would be differentially inherited by the daughter cells (Weismann, 1893), determining its fate, i.e., if a cell inherits "muscle-determinants" it differentiates into a muscle cell. This notion of development was called "mosaic development". Importantly, in Weissman's theory, there is an explicit link between embryology and heredity (or genetics) (Gilbert, 1991). In fact, at that time any discussion of development had explicit genetics components, and viceversa (Gilbert, 1991). To test the mosaic development hypothesis, Roux killed one blastomere (by puncturing it with a hot needle) in 2-cell frog embryos and observed that, just as Weismann theory predicted, a half embryo was formed (Roux, 1888).

In 1892, in a further attempt to prove mosaic development, Hans Driesch separated the cells of a 2 cell sea urchin blastula with clear expectations of obtaining half sea urchin embryos. Instead of this, he was surprised to obtain two small sea urchin embryos (Driesch, 1892). One of Driesch's main conclusions was that the fate of a cell was not predetermined after cell division, but it depended on its location in the embryo (Driesch, 1894). Opposite to mosaic development, this type of development has been defined as "regulative development" (Gilbert, 2014).

The experiments of Roux and Driesch laid the foundations of a new scientific programme whose main purpose was to "research the causes, on which the formation, maintenance and regression of the organic forms are based" (Roux, 1897). Most importantly, they demonstrated that the problem of development was tractable and that hypotheses could be experimentally tested.

1.2.5 Spemann's *organizer*

In 1921 and 1922, Hans Spemann and Hilde Mangold perfomed what Slack has called "the most famous experiment in all of embryology" (Slack, 2012). They grafted (transplanted) a part of a gastrula amphibian embryo, the dorsal lip, into different positions of another host embryo. This resulted in the formation of a secondary embryo (that sometimes developed as a siamese twin), partly from the graft and partly from the host embryo (Spemann and Mangold, 1924). They named the dorsal lip region *organizer*. After its discovery, J. Huxley, G. de Beer, J. Needham and C. H. Waddington had a great

influence in spreading the importance of Spemann's findings (Horder, 2001). Conrad H. Waddington, a leading embryologist and geneticist mostly know for his 'epigenetic landscape' and 'genetic assimilation' concepts (Slack, 2002), wrote:

> *"The special importance of the organization centre is better conveyed by the name Spemann actually chose; it is that part of the embryo with respect to which all the rest is organized. In order to describe the behaviour of any part of a newt gastrula, it is necessary and sufficient to specify its relation to the organization centre. Spemann's name for his discovery may at first sight seem rather grandiloquent, but is really quite reasonable and accurate" (Waddington, 1962).*

However, how the organizer exerted its influence in its surroundings was not known. Waddington and many other embryologists around the world tried to characterize the chemical nature of the organizer (Waddington et al., 1935; Gilbert, 1991). Despite their efforts, they did not succeed and by the end of the 1930's "the sense of disappointment and disillusionment was manifest" (Horder, 2010), which caused the gradual lost of interest in the organizer problem (REF Holftreter in gilbert 1999)

1.2.6 The rise of genetics and its split from embryology

At the same time Spemann was investigating the organizer, genetics was advancing at a fast pace, establishing its own methods and concepts (Gilbert, 1991; Horder, 2001). Soon after the rediscovery of Mendel's laws in the 1900's there was an increased acceptance of the chromosomal theory of development. However, many embryologists did not accepted this theory. Gradually, genetics and embryology began to separate.

A crucial and unexpected contributor to this separation was Thomas Hunt Morgan. Morgan, who started his career as an embryologist, first rejected the chromosomal theory (or any particulate theory of development), considering it a modern preformationism view. He supported instead an epigenesis view, in which material differences in different eggs (such as chromosomes) "are too remotely connected with the end product of their development for us to think of those differences in terms of special or separate particles except in the purest symbolic fashion" (Morgan, 1910).

However, Morgan changed his views on chromosomes and heredity. After the results of his own research on developmental causes on sex determination, and the discovery of many mutations that segregated with the X-chromosome, he was forced to support the view he had been contending against for over a decade (Gilbert, 1978).

In his book "Theory of the Gene", Morgan declared the separation between embryology and genetics stating that "the theory of the gene is justified without attempting to explain the nature of the causal processes that connect the gene and the characters" (Morgan, 1926).

The new chromosomal theory combined in the 1940's with population genetics and other fields to form the Evolutionary Synthesis. Development, as it was considered irrelevant to the study of heredity, was excluded from the Evolutionary Synthesis (Amundson, 2005).

Ersnt Mayr, one of the most influential biologists of the 20th century, reinforced in the 1960's the exclusion of development from the Synthesis with his dichotomy of "proximal" and "ultimate" causes (Mayr, 1961). According to Mayr, "proximal" causes like development (or any physiological process) were not of interest for the evolutionary biologist (Mayr, 1961, 1993).

1.2.7 Developmental genetics

In the subsequent decades after the Synthesis, there were great advances in molecular biology and genetics. The unravelled DNA structure (REF) and the discovery of the gene regulation of protein synthesis (Jacob and Monod, 1961) lead to the acceptance of the central dogma: DNA must carry the information of Mendelian genes (Crick, 1958, 1970). Genes became the central focus in the study of evolution while development was considered for many to be just a readout of a genetic programme (see Keller, 2000).

Taking genes as responsible for the phenotype and assuming, like some evolutionary biologists had, that phylogenetically distant groups of animals developed and had evolved by completely different means (Carroll, 2005), homology between genes was not expected to be found. Mayr wrote:

> *"Much that has been learned about gene physiology makes it evident that the search for homologous genes is quite futile except in very close relatives. If there is only one efficient solution for a certain functional demand, very different gene complexes will come up with the same solution, no matter how different the pathway by which it is achieved" (Mayr, 1966).*

Mayr's prediction was incorrect. In the 1980's the Hox genes, a family of transcription factors, were shown to be conserved in arthropods and insects (McGinnis et al., 1984; Duboule and Dollé, 1989). Furthermore, Hox genes were shown to be involved in anterior-posterior patterning in many animals. Thus, not only genes were conserved between different animals, but their developmental role was also conserved. The concept of *developmental gene* was born, changing the discussion of development and how the gene was viewed with regard to evolution (Gilbert, 2000). The discovery of shared genetic regulatory mechanisms in structures that were not thought to be homologous based on their morphology (like animal eyes) was called "deep homology" (Shubin et al., 1997), a name making clear connections between developmental and evolutionary processes.

> *"Such homologies provide a profound insight into the evolutionary process. Studies of deep homology are showing that new structures need not arise from scratch, genetically speaking, but can evolve by deploying regulatory circuits that were first established in early animals." (Shubin et al., 2009)*

Development was not longer set aside of evolutionary discussions. However, some researchers were convinced that development, not only can be informative of evolutionary processes, but has a causal role in evolutionary change.

1.2.8 Evolutionary developmental Biology

In 1981, 48 researchers from very different scientific backgrounds (e.g., molecular biology, paleontology, developmental genetics, experimental embryology, mathematical biology) held a conference in Dahlem (Germany) with one goal: *"to examine how changes in the course of development can alter the course of evolution and to examine how evolutionary processes mold development"* (Bonner, 1982). The attendees, including Pere Alberch, Stephen Jay Gould, Lewis Wolpert and Eric Davidson, discussed for 5 days the role of development in evolutionary change from different levels: molecular, cellular, life cycle and evolutionary level.

The conference was a success and it gained attention even before its report was published (Lewin, 1981). One of the most important messages conveyed was that "developmental constraints" are important to evolutionary change (Alberch, 1982). The developmental constraint concept (defined as biases on the production of variant phenotypes

or limitations on phenotypic variability caused by the structure, character, composition, or dynamics of the developmental system; Maynard Smith et al., 1985), became central in evolutionary discussions in the subsequent years (Love, 2014) (although thereafter criticized for its negative implications; see Salazar-Ciudad, 2006; Love, 2014).

This and other concepts like "heterochrony", "canalization" and "phenotypic plasticity" played an important role during the conference. Since then, other concepts like "evolvability", "robustness", "modularity" and "variational properties" have also contributed to discuss the role of development in evolution (REFs).

Most importantly, these concepts formed part of a new conceptual framework (Love, 2014) that, together with the advances in developmental genetics, brought together again the fields of genetics, development and evolution, into a new scientific field: evolutionary developmental biology.

1.3 On the statistical approach in Biology

The statistical approach I have used in here, is nothing but new.

Darwin used a statistical approach to describe the action of natural selection (REF Darwin). For him, given the origination of small variations in natural populations, the occurrence of any advantageous variation in an individual, as slight it could be, would be reflected in a better chance of survival and to procreating their kind (Darwin). With many generations, the differential survival of the variants, would produce a change in the population mean. The effects of natural selection are thus only observable at the population level.

A more formal approach came from physics, more precisely from the study of diffusion of gases in the 19th century.

Against the main views of his contemporaries, which considered that all the particles in a gas move at the same speed, J. C. Maxwell proposed that each particle of a gas moved with different velocity and direction, both changing after the particles collision among them (REF Maxwell 1,2). The velocities in all directions are distributed among the particles according to a certain law. As it was impossible to observe the behaviour of all the particles, their properties could only be described at a statistical level, as the average movement of large numbers of gas particles.

For Boltzmann and Gibbs, which extended the studies on gas diffusion, the study of large numbers was not only important to overcome the problem of not being able to study each individual particles, also because their individual behaviour is not interesting at all (Jacob, logic of life). Knowing the movement and direction of each particle would not give more information than the population as a whole.

After the success of statistical mechanics, its methodology expanded to many other scientific fields. Laws could be applied to solve previously intractable problems by collecting sufficient information of a great number of cases of the same class and calculating its mean. The aim of the statistical approach is then to "obtain a law which transcends individual cases" (Jacob).

This novel approach changed biology drastically, transforming it into a quantitative science. As François Jacob said, "at the end of the nineteenth century, the study of living beings was no longer a science of order, but one of measurement as well".

1.4 Complexity

> *"The embryo in the course of development generally rises in organisation (...) I am aware that it is hardly possible to define clearly what is meant by the organisation being higher or lower. But no one probably will dispute that the butterfly is higher than the caterpillar."*
>
> Charles Darwin 1859

In this section, I will talk about the increase in complexity during embryonic development. I will first try to define in complexity in an organism in general terms, mentioning alternative definitions that have been already proposed. Then, I will explore the relations between complexity in evolution and development, and discuss the possibility of a trend in terms of complexity increase.

1.4.1 Defining complexity

Daniel W. McShea has provided an useful general definition of biological complexity. According to him, complexity is "the amount of differentiation among its parts or, where variation is discontinuous, the number of part types" (McShea, 1996, 2015). This definition can be used at different hierarchical levels of biological organization, e.g., tissues, cells, genes. Indeed, a measure of morphological complexity that has been favoured by some authors is the number of cell types that compose an organism (Valentine et al., 1994; Bell and Mooers, 1997; Bonner, 2004). Importantly, with this definition, complexity at different levels are not necessarily correlated.

This is related with the already acknowledged lack of correlation between the number of coding-genes and morphological complexity, sometimes referred as the "G-value paradox" (Hahn and Wray, 2002). Some decades ago, there was the expectation that the morphological complexity should correlate with the number of genes in an organism. With the release of the first eukaryotic genome sequences, such correspondence was not observed. Before that, the lack of correspondence between genome size and organism complexity (or "C-value paradox") was also noted.

An alternative definition of complexity includes not only the "number of parts" but also the "interaction among parts" (McShea, 1996; Arthur, 2010). This could be illustrated with the number of gene-gene interactions (e.g., expression regulation by a transcription factor binding to a promoter region of another gene), such that when comparing two different organisms that have same number of genes, one organism could be considered to be more complex than the other if the former has more gene-gene interactions than the latter.

As before, a high complexity at the molecular level would not necessarily imply a high complexity at a higher level. Related to this is the observation that In fact, it is acknowledged that, during evolution, gene-gene interactions underlying a phenotype can increase their complexity without affecting the phenotype (Müller and Newman, 1999; True and Haag, 2001; Salazar-Ciudad, 2009).

1.4.2 Complexity Increase in Development

The increase in complexity in an organism during embryogenesis is probably one of the most intuitive processes of animal development.

It is commonly seen even as one of its defining characteristics. Eric H. Davidson described the progressive increase in complexity as the "essence" of development (Davidson, 2001). Despite of the widely accepted view of complexity increase in development, there is no consensus of how to define it, much less on how to quantify it (Oyama, 2000).

Using the number of cell types, the increase of complexity during development is self-evident: in mammals, the embryo begins with one cell type and concludes with up to 200 cell types.

This definition of complexity is not exempt of complications, as there is no clear criteria of how to define a cell type or how to determine when a new cell type has formed during development. In addition, this definition does not take into account that embryos do not only get more cell types, but these cell types become organized in specific patterns in space and time in the embryo, which also could be considered as an increase in complexity over developmental time.

The notion of an increase in complexity during embryogenesis is tightly related to the concepts of embryo compartmentalization and pattern formation.

Compartments in development

The concept of compartmentalisation in the embryo was firstly proposed in an analysis of the wing imaginal disc in Drosophila melanogaster (García-Bellido et al., 1973). Using clonal analysis, Garcia Bellido et al., found that different parts of the fly wing were subsequently determined in the imaginal disc by the formation of differentiated populations of cells that do not intermix between them, from initially homogeneous contiguous cells. They demonstrated that the imaginal disc is initially divided in two compartments: anterior and posterior, that is subsequently subdivided into smaller compartments defining specific parts of the wing (García-Bellido et al., 1973). It was later proposed that the compartmentalisation was specified by a genetic code or address: "in effect, a binary ZIP code representing the decisions of key regulatory genes" (García-Bellido et al., 1979). Importantly, the subsequent formation of compartments in the embryo would represent an increase in complexity using the number of cell types.

Developmental pattern

A developmental pattern can be defined as the specific distribution of cell types in a specific temporal window of embryonic development (Salazar-Ciudad and Jernvall, 2004). Therefore, development can be conceptualized as the continuous transformation of one pattern into another. This relates to the compartments definition, as the earliest pattern transformations usually establish the main axis or "compartments" of the embryo. For example, the anterior/posterior (A/P) and dorsal/ventral (D/V) axes in the fruit fly. Later pattern transformations would define smaller compartments of the embryo, e.g, limbs, fingers or internal organs. However, the definition of pattern is different from the compartments one, in that the a pattern transformation does not necessarily involves changes in gene expression. A new pattern could be formed by a morphogenetic process, e.g., migration.

As development proceeds, spatial compartments are progressively specified at an increasing finer resolution (Davidson, 2001). Thus, a great proportion of pattern transformation involve the partition of specific embryo compartments into smaller sub-compartments.

Complexity at the molecular level

If we consider again the number of cells as the complexity measure, it could be expected that the increase in complexity over developmental time (as the number of cell types augments), should be associated with an underlying increase in complexity at the molecular level (Arthur, 2010), following the reasoning that:

i. In development, the morphological complexity increases with time, as new cell types form.
ii. Different cell-types are characterized by the differential expression of genes.
iii. Therefore, the more cell-types an organism is formed of, more different combinations of expressed genes has to have (with the gene regulatory complexity this must entail).

The above reasoning has lead to some researchers to propose that the complexity of an organism resides in the regulatory machinery that ends into the differentiation of the diverse cell types (Davidson, 2001).

Furthermore, the increasing compartmentalization of the embryo during development can be conceptualized as the progressive spatial restriction of gene expression to subsequently smaller regions in the embryo. Sean Carroll defines this process(Carroll et al., 2001) as:

i. In early development, genes have a broad expression in the embryo and define the main axes of the body.
ii. Later, genes define smaller compartments like organs and appendages (field-specific selector genes).
iii. Finally, genes become expressed in specific cell types like muscle and neural cells (cell-type specific selector genes).

It is important to note that this would imply that, in general, the area of expression of a gene in the embryo (relative to the area of the whole embryo).

At the molecular level, the definition of "interaction among parts" and "number of parts" can be easily associated, as the the differential gene expression in the various number of cell types are determined in great manner by the interaction of between genes and their *cis*-regulatory regions (DNA regions usually close to a gene which contains specific sequence motifs where proteins bind and affect its expression). The interaction between genes and their *cis*-regulatory regions is also referred as gene regulatory networks (GRNs) or "regulatory architecture" of the genome (Davidson, 2001).

In addition to the interaction between *cis*-regulatory regions and genes, there are other gene expression regulatory mechanisms that have been proposed to be crucial in the origin of complex organisms. This is the case of the microRNAs (miRNAs), non-coding RNA molecules that negatively regulate gene expression. After the observation that miRNAs are found only in protostomes and deuterostomes and not in sponges or cnidarians, and that they are specifically expressed in certain cell-types, tissues or organs, it was proposed that regulation of gene expression by miRNAs could have played a significant role in the origins of complex organs and "body plans" (Sempere et al., 2006).

Different types of developmental genes

It is widely acknowledged that the spatio-temporal regulation of gene expression in development is crucial for the progressive compartmentalization of the embryo. Therefore, it is useful to identify which type of genes are directly involved in this process. More than fifty years ago, Jacques Monod and François Jacob (Jacob and Monod, 1961) published in a seminal work a model of the genetic regulatory mechanism in bacteria. The most important conclusion of this paper was the existence of "regulator" genes that control the production rate of proteins from "structural" genes, and that mutations in "regulator" genes affect the regulatory mechanism but not the structure of the regulated protein. In the same paper they suggested that these regulator genes may affect the synthesis of several different proteins (Jacob and Monod, 1961).

Nowadays the process of gene activation is known in great detail. The so-called "regulator genes" are now known as transcription factors, proteins that bind to DNA to promote or repress the transcription of a gene.

The information for the spatio-temporal regulation of gene expression during cell differentiation requires however more than transcription factors, as the differentiation of a cell depends in a great manner of extracellular signals from its neighbouring context (Gilbert, 2014). The molecule network involved in cell-cell communication, from the reception of a extracellular signal to the ultimate transcription of genes (usually going through many intermediate signal "transducers"), is known as a signalling pathway (see Figure 1.1). Due to the importance of both transcription factors as signalling pathway genes in cell differentiation, a brief description of each follows.

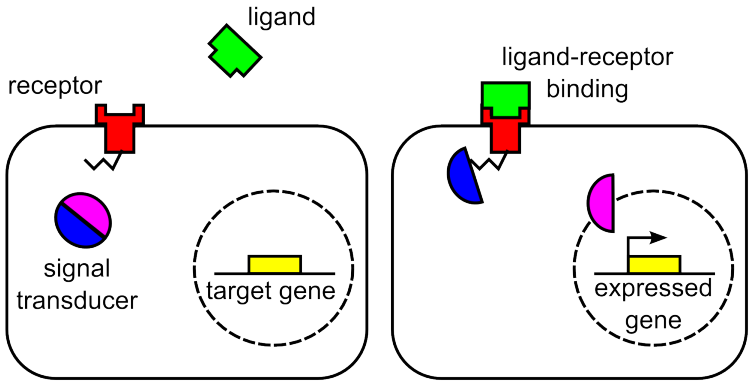

Figure 1.1: Scheme of an hypothetical signalling pathway. Left: The extracellular ligand (green) is not bound to the membrane receptor (red) so the signal transducer protein (blue/magenta) is inactivated. In the nucleus (dashed circle) the target gene (yellow box) is inactive. Right: As the ligand binds to the receptor, the cytoplasmic domain (depicted as a zigzag tail) of the receptor change to an active conformation, cleaving the signal transducer. Part of the signal transducer acts as a transcription factor, going into the nucleus where activates the transcription of the target gene after binding to its regulatory region.

Transcription factors The transcription factors (TFs) are proteins that bind to specific regulatory regions, to induce or repress the expression of a gene. Based on the secondary structure of the protein binding domain, TFs can b e classified in four main

families: helix-turn-helix, helix-loop-helix, zinc finger and leucine zipper ((Carroll et al., 2001).

The members of each family has been recognised in playing different roles in development. For example, it has been observed that in diverse metazoan species C2H2 zinc-figers TFs are over-represented in early development, as opposite to Homeobox TFs which are under-represented in the same period (Schep and Adryan, 2013). Hox genes (a subset of the Homeobox TF family) are involved in the A/P patterning of many metazoan groups. Intriguingly, these genes were found to have spatial collinearity in mice and flies (REF). That means that the A/P expression of the Hox genes reflects their physical order along the chromosome. At the time of its discovery, collinearity of Hox genes were considered as a master plan for A/P patterning in animals (REF). However, after Hox genes were investigated in more species it became clear that in some species with Hox genes collinearity is not always present, and some species do not have Hox genes at all (REF).

Signalling pathway genes Signalling pathways are usually a complex network of molecules including extracellular diffusible signals, membrane receptors, intermediate signal transducer molecules and transcription factors. Signalling pathways usually begin with a extracellular signal that causes a conformational change in its cell membrane receptor after binding to it. The new conformation results in enzymatic activity in the cytoplasmic domains of the receptor protein, that phosphorylate other cytoplasmic proteins. Finally, one or more activated transcription factors induce or repress specific gene activity (Gilbert, 2014).

Signalling pathways recurrently used during animal development are the Wnt, FGF and Shh pathways (for a detailed description of each signalling pathway, see (Gilbert, 2014). For example, the Shh pathway plays a fundamental role in the fruit fly segment polarization(REF) and wing development (REF), and in vertebrate limb (REF) and tooth development (Jernvall et al., 2000).

1.4.3 Complexity Increase in Evolution

The increase in complexity in evolution has has been a topic of interest for more than a century. Early views of evolution saw the increase in complexity as inexorable, with all the species descending from simpler ancestral forms (Lamarck, 1809; Haeckel, 1874), and with the human species as the latest and more perfect product of the evolution of animals (Haeckel, 1874).

Recent views recognize that within a phylum, complexity of the species can increase or decrease. Using the number of cell-types as complexity measure, there are clear examples of taxa that have decreased their complexity over time, specially in parasites. An example would be the animal group formerly known as the "Mesozoa", which are worm-like parasites of marine invertebrates. Because of their simple morphology (based on their small number of cell types), these animals were thought to be "living fossils" or intermediate forms between Protozoans and Metazoans. Now, even when they remain poorly studied animals, it is thought that they are degenerate descendants of more complex ancestors, probably some lophotrochozoan group (Arthur, 2010). The Orthonectida, for example, is a phylum of parasites of marine invertebrates with only two types of cells, external ciliated and internal reproductive cells without any internal organs. Molecular phylogenetic analysis provided evidence that these animals are more

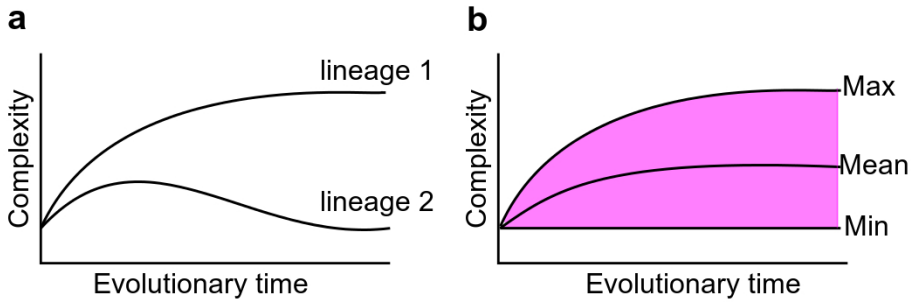

Figure 1.2: a) Two lineages with different complexity change through their evolutionary trajectories. b) Representation of the minimum, mean and maximum complexity of many lineages over evolutionary time in which the minimum stay constant while the mean and maximum increase. Redrawn from (Arthur, 2010).

closely related to tripoblastic animals than to protists or diplobastic taxa (Hanelt et al., 1996). These animals most probably evolved from a more complex free living animal and decreased their morphological complexity after they adopted a parasitic life style.

So, it seems that there is no unique trend to increase the complexity over time, i.e., in a specific lineage, complexity might decrease, increase or stay the same (see Figure 1.2a).

Is there a trend towards increasing complexity?

As the organisms have a wide range of morphological complexities, it is useful to use an statistical approach to identify a trend towards increasing complexity. If we consider a minimum complexity statistic, it can be said that complexity has remained more or less constant in evolution, as low complexity (unicellular) organisms like bacteria, have been present from 3.5 billion years.

If we look at the mean or maximum statistics, a trend for increasing complexity would be apparent, as the initial complexity would have increased with the appearance of simple multicellular organisms (only few cell types) and would have increase more with the appearance of organisms composed of hundreds of cell types.

It is important to notice that this apparent trend towards increasing complexity does not necessarily imply that it has been selected for (McShea, 2015). Even a "passive" mechanism, could result in an apparent complexity increase.

In this case, given that the first organisms would have been simple, and there is a low complexity boundary (it is not possible to be simpler than an unicellular organims), even if the complexity of different lineages would have changed in a random walk fashion (see Figure 1.2b) the mean and maximum complexity would increase in evolution (Gould, 1996).

1.4.4 Other definitions of complexity

Complexity in informational terms

Davidson (2001) used the GRN concept in addition to others to explain development (and evolution) on informational terms. He said that development (which is the outcome

of spatial and temporal series of differential gene expression) is controlled by a hardwired regulatory program built into the DNA and the metric of complexity is the diversity of the programs of gene expression that are "installed and executed" as the embryo develops. As Davidson, other authors have used informational/computational analogies to define development (Apter and Wolpert, 1965; Monod, 1963; Mayr, 1997)

To illustrate how the complexity of a regulatory network or "program" can increase in evolution (but a similar case could be said for development), Davidson describes an imaginary example: an early evolutionary state consists of a small gene battery (set of functionally linked genes expressed in concert) encoding proteins used for some differentiated cell type, which is activated by a small number of genes encoding transcription factors. The network activating the gene battery is itself controlled by a single upstream gene. In subsequent evolutionary states, the whole structure is said to become more complex as: the battery of genes is now used in some pattern formation system, new batteries of genes appear, new regulatory genes and new *cis*-regulatory regions are introduced (Davidson, 2001).

Even when in this kind of examples would seem easy to discern a simple GRN from a complex one just from its topology, the high intricacy of real biological systems make this an extremely difficult if not impossible task.

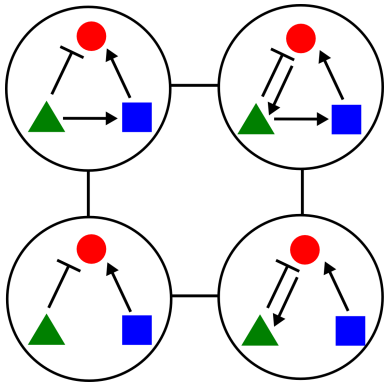

Figure 1.3: Some GRNs . Cotterel and Harpe calculated... (Cotterell and Sharpe, 2010)

There are also important critiques of this informational approach. First, using an informational analogy to describe development implies the distinction between a "hardware" and a "software". The "hardware" would consist of the genome structure, regulatory components, cells, organs, etc., and the "software" would be the GRNs or the set of instructions that directs the performance of specific operations. For biological systems, this distinction is misleading, as there are recurrent feedback between its "hardware" and "software", so that the structure of development processes change through development (Oyama, 2000; Salazar-Ciudad, 2009; Jaeger and Sharpe, 2014).

Complexity in terms of dynamical systems theory

Estimating the combinatorial possibilities of a small set of regulatory genes, considering that each gene can regulate (whether positively or negatively) more than one gene's expression in addition to its own, could result in an astronomic number of possible gene topologies (see Figure 1.3). Also feedback effects and non-linear regulation of gene

expression make the prediction of changes in regulatory states hard or even impossible to predict (Jaeger and Sharpe, 2014).

To overcome this limitations, some authors have propose to use dynamical systems theory, which deals with a complex system with many interacting components (a dynamical system), by representing its state as a point in a multidimensional space (Alberch, 1991; Forgacs and Newman, 2005; Jaeger and Sharpe, 2014). To illustrate this we could think of a specific cell type, with n number of genes, which its cell state depends on the expression of each of the genes. The simpler case we could imagine would be a cell with only two genes. In this case, the cell would be in a two-dimensional "state space" (also called "phase space").

Importantly, the dynamical system is governed by the relations between its components (Forgacs and Newman, 2005). In our example the relations would be represented by the interaction between genes, namely the gene regulatory network (GRN). If in our example the expression level of one gene is affected by the expression level of the other gene, the system will not stay in any particular state, but it will change until it reaches a "stable steady state", in which the level of both genes are at equilibrium. Given a specific GRN, the number of stable states would represent the possible differentiation states a cell can achieve (REF Slack book)

Many modifications of the GRN would not have consequences in the ultimate differentiation state, as it will converge to the same "attractor" point. However, some modifications (or mutations) could produce a change in the "state space" leading to the formation of a new stable state (i.e., a new cell type). So within the framework of the dynamical systems theory, and keeping the number of cell types as our measure of complexity, there would be an increase in complexity when a mutation would change the gene regulatory machinery so that a new stable steady state is formed.

1.5 Adaptation

Measuring adaptation is a central theme in evolutionary biology. In many texts however, adaptation is not clearly defined (leading to ambiguity) or is often used in biological irrelevant contexts (REF Dobzhansky 1968). Even in Darwin's *Origin*, where it is a central concept, adaptation is not properly defined throughout the text (Darwin, 1859).

Adaptation

Usually adaptation can refer to two different things, to a trait that confers an advantage to its bearer, and to the process to become adapted. Simpson (1955) said that:

"*an* adaptation is a characteristic of an organism advantageous to it or to the conspecific group in which it lives, while adaptation or the process of adaptation is the acquisition within a population of such individual adaptation" (italics by Simpson; REF Simpson)

We say an organism is adapted to an environment when it is able to live and reproduce in it. A related concept is adaptedness, or "the degree to which an organism is adapted to an environment" REF Dobzhansky.

Since Darwin, the concepts of adaptation and natural selection usually come together. There has been a long standing controversy on the relationship between natural selection and adaptation. For some authors like Barton et al., adaptation is "a trait

that functions to increase fitness and that evolved for that function" and that can be only caused by natural selection (REF Barton). Other authors however, consider that an adaptation might originate only by chance and natural selection only addresses the spread of adaptive variants (Williams, Gould and Vrba, Alberch).

Natural selection

Charles Darwin, in its 1859's *Origin*, defined Natural selection as follows:

> *" Owing to this struggle (for life), variations, however slight and from whatever cause proceeding, if they be in any degree profitable to the individuals of a species (...) will tend to the preservation of such individuals, and will generally be inherited by the offspring. The offspring, also, will thus have a better chance of surviving, for, of the many individuals of any species which are periodically born, but a small number can survive. I have called this principle, by which each slight variation, if useful, is preserved, by the term Natural Selection"* (Darwin, 1859).

More recently, and following the Darwininan concept of natural selection, Jhon A. Endler (1986), defined it as a process in which, given that a population has:

a. **variation** among individuals in some attribute or trait;
b. **fitness differences** (consistent relationship between that trait and mating ability, fertilizing ability, fertility, fecundity, and, or, survivorship);
c. **inheritance** (consistent relationship, for that trait, between parents and their offspring, which is at least partially independent of common environmental effects).

Then:

i. the trait frequency distribution will differ among age classes or life-history stages, beyond that expected from ontogeny;
ii. if the population is not at equilibrium, then the trait distribution of all offspring in the population will be predictably different from that of all parents, beyond that expected from conditions a and c alone.

Conditions a, b, and c are necessary and sufficient for the process of natural selection to occur, and these lead to deductions i and ii (Endler, 1986).

Condition a relates to phenotypic changes across generations. Importantly, a phenotypic change , whether a new character or a modification of an existing character in the adult/larva, is produced from a change in development. For example, the difference in the beak size and shape between the famous Galapagos Darwin's finches (REF Darwin), a classic example of adaptive change under natural selection, has been shown to be regulated by the differential expression of the genes CaM and BMP4 during development. Indeed, a proposed model of the role of BMP4 and CaM role in beak size and shape explains both elongated and deep/wide beaks of these finches (Abzhanov et al., 2006).

So, even when natural selection acts in the adult phenotype (or in the larva, in the case of species with a feeding larva stage), we should be able to find changes in development that would explain an adaptive change in the adult or larva.

Methods to detect natural selection

There are many different methods designed to detect natural selection in natural populations. Jhon A. Endler classified ten different methods with diverse ability to detect

natural selection (Endler, 1986). Some of these methods test directly the conditions (b and c) required by natural selection, while others test the predicted outcome result of natural selection in a population.

Among the latter we find the molecular methods. The molecular methods are based on the assumption that if changes leading to an adaptation are (at least partially) caused by mutations in gene regulatory or coding sequences, the effects of natural selection could be traceable looking at the adaptive changes in the genes expressed in different times and locations during development. There is an entire field within evolutionary biology, namely molecular evolution, dedicated to explain the sequence changes in molecules as DNA, RNA and proteins.

In the next sections, due to its relevance in this work I will only focus on the molecular methods to detect natural selection.

1.5.1 Molecular evolution

The theoretical basis of the molecular evolution field includes concepts from evolutionary biology and population genetics. At the DNA level, any transmissible change in the sequence is considered a mutation. The most simple change is a point mutation or single nucleotide polymorphism (SNP), which is a change in a single nucleotide in the DNA sequence of a locus of two individuals. If the individuals belong to the same species, this mutation is referred as polymorphism. In contrast, divergence refers to the mutations when individuals from different species are taken into account. SNPs occur in non-coding and coding DNA sequence. A single point mutation that occurs in a coding sequence can be classified in two categories, depending on the effect of this mutation in the protein sequence: i) synonymous mutation and ii) non-synonymous mutation. A synonymous mutation does not affect the amino-acid sequence of the protein, albeit it can affect its function (Kimchi-Sarfaty et al., 2007) or the gene transcriptional efficiency (REF). A non-synonymous mutation does affect the amino-acid sequence of the protein whether by changing a single amino-acid (missense mutation) or by producing a stop codon (non-sense mutation) which results in a truncated version of the protein.

As the non-synonymous mutations can affect dramatically the structure and function of the protein, it is expected that most of non-synonymous mutations would have a negative fitness effect. However, it is also expected that a fraction of non-synonymous mutations, or adaptive substitutions, would have a positive fitness effect that (depending on the strength of the fitness effect) could lead to the fixation of that mutation in the population.

An important branch of the molecular evolution field is dedicated to the identification of adaptive substitutions in a species, which has lead to the development of many statistical tests. Importantly, these tests are based on the neutral theory of evolution, proposed by Kimura (Kimura, 1968).

1.5.2 Neutral theory of evolution

In 1968, Mooto Kimura calculated the average rate of nucleotide substitutions in the evolutionary history of mammals. The result of his calculations was that, on average, one nucleotide has been substituted every 2 years. For him, this very high rate of substitution was only explainable if most mutations were almost neutral in natural selection (Kimura, 1968). This was in contrast with the prevailing view at the time that practically no

mutations are neutral (REF). More importantly, the neutral theory provided a set of testable predictions, providing a null-hypothesis of molecular evolution. This allowed the development of statistical methods to detect adaptive changes, i.e., we can say that a sequence has been under positive selection if the amount of changes exceeds the number of changes expected only by neutral evolution. One of the most popular tests is the McDonald-Kreitman test (MKT), which estimates the proportion of the adaptive substitution resulted from natural selection.

1.5.3 McDonald-Kreitman test

John H. McDonald and Martin Kreitman developed this test in 1991 when analysing the divergence in the Adh locus in three Drosophila species (McDonald and Kreitman, 1991). The main assumption of the MKT is that the substitutions in a protein are neutral if the inter-specific ratio of non-synonymous (Dn) to synonymous (Ds) changes is equal to the intra-specific ratio of non-synonymous (Pn) to synonymous (Ps) changes (i.e. Dn/Ds = Pn/Ps). Any departure from these equality would imply the action of positive or negative selection. If some of the changes are result from positive selection, the ratio of non-synonymous to synonymous variation within species should be lower than the ratio of non-synonymous to synonymous variation between species (i.e. Dn/Ds $>$ Pn/Ps). In the case that the observed ratio of non-synonymous to synonymous variation between species is lower than the ratio of non-synonymous to synonymous variation within species (i.e. Dn/Ds $<$ Pn/Ps) then negative selection is at work.

Since mutations under positive selection spread through a population rapidly, they don't contribute to polymorphism but do have an effect on divergence.

Although the MKT has been proved robust to many sources of error (e.g., variation to mutation rate across the genome), it can underestimate the proportion of adaptive changes in the presence of slightly deleterious mutations (Messer and Petrov, 2013; Eyre-Walker et al., 2006). Recently, more sophisticated methods based on the MKT have been developed to correct for underestimation of adaptive evolution in the presence of slightly deleterious mutations.

1.5.4 Distribution of Fitness Effects

To have a more precise estimate of the proportion of adaptive substitutions it is important to consider the relative contributions of the different types of mutations, based on their fitness effects. Because even when for simplicity the mutation effects are usually classified in advantageous, neutral, and deleterious, there is actually a continuum of selective effects, from strongly deleterious, to highly adaptive mutations (Eyre-Walker and Keightley, 2007), with weakly deleterious, neutral and slightly adaptive mutations in between.

The relative frequencies of all these type of mutations is called the Distribution of Fitness Effects (DFE). The DFE has other practical implications, like predicting the effects on the genetic variation in a population with low population size. In order to know the DFE, a few experimental approaches exist. The most direct method is whether to induce (Sanjuán et al., 2004) or to collect (MUKAI, 1964) spontaneous mutations and assay their effects (fitness) in the laboratory. As can be expected, this experiments require many generations to gather sufficient data, so these approaches have been used mainly in micro organisms (Eyre-Walker and Keightley, 2007). A caveat of

these experimental approaches is that, in order to identify the effect of a mutation, its effect has to be detectable in a fitness assay. Therefore, these methods give valuable information for mutations with relatively large effects.

An alternative approach is to infer the DFE by analysing patterns of DNA sequence differences at intra and inter-specific level (polymorphism and divergence respectively). The methods using this approach rely mainly on two assumptions:

i. the probability that a mutation spreads to a certain freq in a population (or to fixation) depends on the strength of selection (positive or negative) acting on it. Severely deleterious mutations have lower probability to reach a high frequency in a population.

ii. the efficiency of selection depends on the effective population size. With a high effective population size, selection is more efficient and a smaller proportion of mutation will behave as effectively neutral.

The "absolute strength" of selection on a mutation is then measured as $N_e s$, the product of the effective population size (N_e) by the selection coefficient (s) of the mutation. Mutations with $N_e s$ much less than 1 are effectively neutral, while $N_e s$ greater than 100 have no chance to appear as polymorphism.

1.5.5 DFE-alpha

Eyre-Walker and collaborators (Eyre-Walker and Keightley, 2009) proposed a method to estimate both the DFE and the proportion of adaptive nucleotide substitutions (α) using polymorphism and divergence data. More specifically, they use the polymorphism site frequency spectrum (SFS) to estimate the DFE and then use this estimated DFE to estimate the proportion of substitutions under positive selection between species. This method, assumes that there are two types of nucleotide sites: i) sites at which all mutations are neutral and ii) sites at which some of the mutations are subject to selection (positive or negative). Also it is assumed that any new adaptive mutation in a population would not be detected in the polymorphic phase but only in the divergent one, and that the DFE can be represented with a gamma distribution. The advantage of using a gamma distribution is that very different distributions (e.g., normal, exponential, leptokurtic) can be represented using only a shape parameter and the mean of the distribution (Figure 1.4).

The divergence at the neutral sites is then proportional to the mutation rate per site and the predicted divergence at the selected sites, in the absence of advantageous mutations, is proportional to the product of the mutation rate and the average fixation probability of a selected mutation, which is inferred based on the DFE and other parameters estimated from the polymorphism data analysis (Eyre-Walker and Keightley, 2009). The difference between the observed and predicted divergences therefore estimates the divergence due to adaptive substitutions.

Using this method Eyre-Walker and collaborators estimated that approximately 50% of amino acid substitutions and approximately 20% of substitutions in introns are adaptive (Eyre-Walker and Keightley, 2009).

Messer and Petrov performed molecular evolution simulations to test if the estimates of different tests, like the MKT and the more sophisticated DFE-alpha, are accurate under different realistic gene-structure and selection scenarios (Messer and Petrov, 2013). More specifically, the authors wanted to test how accurate these methods are in presence of genetic draft (stochastic effects generated by recurrent selective sweeps at closely

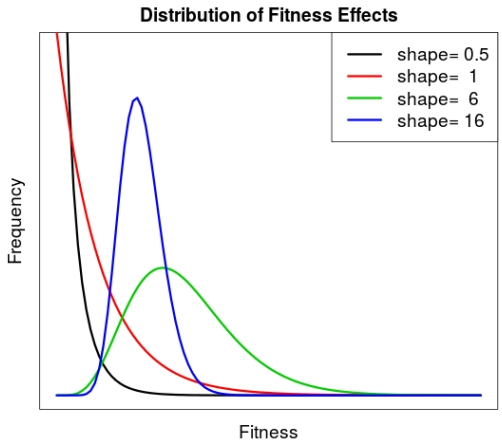

Figure 1.4: Example of different Distribution of Fitness Effects (DFE) represented by a gamma distribution. Many distributions can be represented by modifying the shape parameter of a gamma distribution, from a leptokurtic (shape parameter less than 1) to an exponential (shape parameter equal to 1) or a skewed normal distribution (shape greater than 1).

linked sites) and background selection (interference among linked sites by lightly deleterious polymorphisms).

They found that in the presence of slightly deleterious mutations, MKT estimates of α are severely underestimated. They also found that the DFE-alpha is very accurate to calculate alpha when changes is demography are considered (Messer and Petrov, 2013), with the caveat that DFE can highly overestimate the demography changes when genetic draft (stochastic effects generated by recurrent selective sweeps at closely linked sites) or background selection (interference among linked sites caused by lightly deleterious polymorphisms) are present (Messer and Petrov, 2013).

This is because genetic draft leaves similar signatures to a recent population expansion, namely distortions of the SFS at synomymous sites, and the DFE-alpha interprets these SFS distortions as being a consequence of demography and attempts to correct for it. Importantly, this caveat in using the DFE-alpha is only relevant when demography changes wants to be estimated, but not when alpha is the parameter of interest, as in here.

1.6 Drosophila

The fruit fly, *Drosophila melanogaster*, has been a great valuable tool for biological research. Its use as a model system dates back to the beginning of the 20th century. In 1908, Thomas H. Morgan (see section X) started to grow flies in large quantities to study gene mutations. At that time, the gene concept was an abstract one, as the nature and location of the genes was still disputed. The main advantages of using flies were their rapid generation time, it was easy to culture and cheap to maintain (Arias, 2008). In his lab, at the University of Columbia, Morgan encountered a fly with white eyes (the wild-type eye color is red), which became a subject of his research for many years. Eventually, he discovered that the allele of the gene, that he called *white*, was

located in a sex chromosome, demonstrating for the first time the sex-linkage of genes (Morgan, 1919). Morgan's students also demonstrated that mutations were inducible with X-rays and introduced the use of "balancer" chromosomes to keep stable stocks of mutants (Arias, 2008). The research carried in Morgan's lab laid the basis modern genetics, and its fly room became a central node in the genetics research, establishing Drosophila as a organism model.

However, *Drosophila*'s development was difficult to study, as the embryos were not large enough to experimentally manipulate them, and not transparent enough to visualize with a microscope (Gilbert, 2014).

In 1976, E. Lewis published a seminal work, in which he determined the effects of mutations in the Bithorax complex (BX-C). He determined that the BX-C consisted of distinct genetic elements and there was a correlation in the order of the mutations within the complex and the A/P order of the body affected by them (Lewis, 1978), a phenomenon called spatial co-linearity.

Lewis discoveries were complemented with the discovery of the Hox genes (REF), a family of transcription factors that was shown to be conserved with vertebrates (REF). Hox genes in *D. melanogaster* are arranged in two clusters, the Antennapedia (ANT-C) and the Bithorax cluster. Molecular biology techniques allowed finally to study fly genes and their effect on embryogenesis, unravelling the mysteries of *Drosophila*'s development.

In the next subsections, I will describe briefly: 1) the *D. melanogaster* life cycle with special focus on its embryonic development, 2) the anterio-posterior patterning cascade as it serves as an example of the gradual compartmentalization of the embryo and 3) the blastoderm fatemap and the relation of fate maps with gene expression maps.

1.6.1 *D. melanogaster* life cycle

Drosophila melanogaster is a holometabolous insect, which means that it goes through a complete metamorphosis, i.e., the larva and the adult forms are very different. Its embryonic development is very fast, the larva hatches after around 20 hours (at X degrees). The larva grows and passes through two moults before becoming a resting stage called a pupa in which the body is remoulded to form the adult.

Much of the adult body is formed from the imaginal discs and the abdominal histoblasts which are only present as undifferentiated buds in the larva.

Developmental stages

In here, I will briefly summarized the embryonic development of D. melanogaster, for a comprehensive lecture, see (Campos-Ortega and Hartenstein, 1985; Gilbert, 2014)

Campos Ortega has divided the embryonic development of Drosophila in 17 stages. The main events of each stage and its timing under laboratory conditions are presented in Table X.

TABLE

1.6.2 Anterio-Posterior patterning

A milestone on the embryogenesis research on Drosophila took place in 1980, when Eric Wieschaus and Christiane Nüsslein-Volhard identified crucial genes involved in the early patterning of the Drosophila embryo. They systematically searched for embryonic lethal

mutants, identifying 15 loci that altered the segmentation pattern of the embryo when mutated (Nüsslein-Volhard and Wieschaus, 1980), which they separated in tree groups based on their phenotype: "pattern duplication in each segment (segment polarity mutants; six loci), pattern deletion in alternating segments (pair-rule mutants; six loci) and deletion of a group of adjacent segments (gap mutants; three loci)" (Nüsslein-Volhard and Wieschaus, 1980).

All these genes form part of the A/P patterning cascade, whose hierarchical regulation is currently well known.

Maternal effect genes

The first A/P pattern of the embryo is determined in the egg chamber, during oogenesis. The oocyte nucleus transports *Gurken* protein close to the posterior part of the egg chamber. The follicle cells in that region receive the Gurken signal (Gurken is homologue of the vertebrate epidermal growth factor [EGF], see Neuman-Silberberg and Schüpbach, 1993), which determines their fate as posterior cells. This signal provokes the polarization of the microtubules in an A/P axis, that facilitates the transportation of mRNAs or proteins to specific parts of the oocyte. Among these molecules are the mRNAs of the *bicoid* and *nanos*, which are transported to the anterior pole and posterior pole of the oocyte, respectively. These and other genes, which are known as maternal effect genes, specify the A/P axis regulating specific target genes.

The maternal effect genes are classified in three different groups depending on their localization (anterior, posterior and terminal groups). Each group is briefly described below.

Anterior group After its anchorage to the anterior region of the embryo, the *bicoid* mRNA is translated forming a gradient from the anterior to the posterior part of the embryo. This protein determines the position of the anterior structures of the embryo acting as a *morphogen*, i.e., different levels of Bicoid protein determines different cell fates in the anterior part of the embryo (Driever and Nüsslein-Volhard, 1988). Bicoid is a transcription factor that regulates many target genes in a concentration-dependent manner. Foe example, expression of target genes in the head region require 1)high concentration of Bicoid and 2)the expression of *hunchback*, a gene that is activated at moderate levels of Bicoid (Simpson-Brose et al., 1994).

Posterior group The *nanos* mRNA, that is localized in the posterior region of the embryo, also generates a protein gradient. Nanos inhibits the translation of *hunchback* (Tautz, 1988) by forming a complex with other ubiquitous proteins in the embryo (Cho et al., 2006). The inhibition of *hunchback* by Nanos causes an anterior to posterior gradient of the former. Another gene of the posterior group is the transcription factor *caudal*. Contrary to *nanos* or *bicoid* mRNA, *caudal* mRNA is distributed in the whole embryo. Caudal gradient is formed by translation repression by the Bicoid protein. Caudal activates genes that determine the abdominal fate. The opossing gradients of Bicoid and Caudal will activate zygotic genes at different positions along the A/P axis of the blastoderm embryo.

Terminal group In mutants of terminal group genes the acron and telson are not formed (Klingler et al., 1988), which are the most anterior and posterior regions of the

embryo, repectively. The boundaries of these structures are defined by the Torso signal. Torso is a tyrosine kinase receptor that, although is uniformly expressed along the surface membrane of early embryos, it is only activated at both poles (Casanova and Struhl, 1989).

GAP gene network

Gap genes were named like that as their mutants lacked some segments in the embryo, leaving a "gap" in the embryo. These genes constitute a dynamical gene network of transcription factors directly activated or repressed by the A/P gradients of the maternal effects genes. Their expression consist of one or two broad domains in the embryo, whose boundaries are defined by five basic regulatory mechanisms (Jaeger et al., 2004):

i. Activation of gap genes by Bicoid and/or Caudal
ii. Auto-activation
iii. Strong repression between mutually exclusive gap genes
iv. Repression between overlapping gap genes
v. Repression by Tolloid

Some of the gap genes are *hunchback*, *knirps*, *krüppel* and *giant*. Importantly, the patterning by gap gene network occurs in the late syncitial blastoderm stage, that corresponds to stage 4 and 5 in Campos-Ortega (Campos-Ortega and Hartenstein, 1985). This allows that the nuclei, still not surrounded by membranes, regulate each other expression with transcription factors.

Pair-rule genes

Gap genes control then the expression of the pair-rule genes, when the embryo is still in the syncitial blastoderm stage. Pair-rule genes were name like that as they are expressed in a regularly spaced striped pattern, each stripe corresponding to one parasegment. Parasegments, which are considered the segmental unit in the embryo, do not correspond to the segments of the larva or adult, instead, parasegments and segments are out of phase (a segment is composed by anterior and posterior compartments, a parasegment is composed by the posterior compartment of a segment and the anterior compartment of the next segment; Lawrence, 1992; Arias, 2008).

Pair-rule genes are divided in primary and secondary pair rule genes. The former are regulated by gap and maternal effect genes, while the latter are also regulated from the primary pair rule genes (Chipman, 2015). The stripes of expression of pair-rule genes are modularly regulated by specific enhancers, each for one stripe or a pair of stripes. This was firstly discovered in the stripe 2 of the *even-skipped* (*eve*) gene. Genetic studies determined that *eve* stripe 2 is activated by Bicoid and Hunchback, while repressed by Giant and Krüppel (Small et al., 1991; Stanojevic et al., 1991). Another pair rule gene that is expressed in an alternate manner to *eve* is *fushi-tarazu* (*ftz*).

Segment polarity genes

The next step in the A/P patterning cascade is the activation of the segment polarity genes. At this point blastoderm cells are already formed. Therefore, further pattern formation requires cell-cell communication (Gilbert, 2014). Segment polarity genes are regulated directly by gap and pair-rule genes and, as their name suggests, are expressed

in the anterior or posterior side of the embryo para-segments. The genes involved in this cell to cell communication are members of the Wnt/Wingless (Wg) and Hedgehog (Hh) signalling pathways.

The expression of Hh is regulated by the *engrailed* gene, which in turn is activated is the cells where high levels of Even-skipped gene or Fushi Tarazu. Additional regulatory inputs drive the expression of *engrailed* in the anterior part of each parasegment. Expression of *wingless* is repressed by both Ftz and Eve and repressed by Odd-paired, so Wg is express only in one row of cells, adjacent to the cells expressing *en*. A forward feedback loop (Chipman, 2015). Interaction between the Wg and Hh signaling pathways, reinforce each other expression (Ingham et al., 1991; Heemskerk et al., 1991), maintaining the pattern formed by pair-rule genes and forming a stable boundary between anterior and posterior compartments of each para-segment.

Hox genes

As mentioned before, the discovery of the Hox genes, its conservation among metazoans, and the conservation of their role in the A/P patterning, was a milestone for the emerging field of developmental genetics. But before that, Hox genes were identified as having a role in homeotic transformations. Homeotic transformations were first described by William Bateson (Bateson, 1894) as a kind of natural variation found in animals, where one part of the body was transformed into another part of the body.

> *"the case of the modification of the antenna of an insect into a foot, of the eye of a Crustacean into an antenna, of a petal into a stamen, and the like, are examples of the same kind" (Bateson, 1894).*

In 1915, Calvin Bridges (who was a student of T. H. Morgan and also worked in the famous Fly room at Columbia University) isolated a spontaneous mutation in *Drosophila*, in which part of the haltere (the small posterior flight appendage) was transformed into a wing tissue. Bridges called this mutant *bithorax*. As we mentioned, it was Ed Lewis who then showed that there were several genes responsible for homeotic mutations, that these genes were arranged in two gene complexes in *Drosophila*, and that the order of the genes in the chromosome correlated with the order of the segments in the A/P axis (Lewis, 1978). It was then showed that Hox genes contain a highly conserved sequence of 180 base pairs, the homeobox, which codes for a DNA binding domain known as the homeodomain. Some examples of these genes are *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*) and *Abdominal-B* (*Abd-B*). Gap genes activate and repress Hox genes, resulting in a specific expression of the latter in each of the Drosophila segments (Jaeger, 2011). Then, Hox genes interact between them to refine the A/P expression boundaries (Hughes and Kaufman, 2002). Hox genes have been said to be "selector genes" (García-Bellido et al., 1973; Carroll et al., 2001) as they seem to determine cell fate.

1.6.3 Fate map

The "fate map" is a very important concept in developmental biology. Its name refers to the practice of cartography (or map making), i.e., contructing two-dimensional (2D) representations of a usually three-dimensional (3D) space. In the case of a fate map, the prospective fate is mapped onto the 2D representation of usually an early embryo (Gilbert, 2007). The first fate maps were constructed by tracking cell lineages to identify cell fate, only by observation. In 1905, Conklin tracked the cell lineage of the tunicate

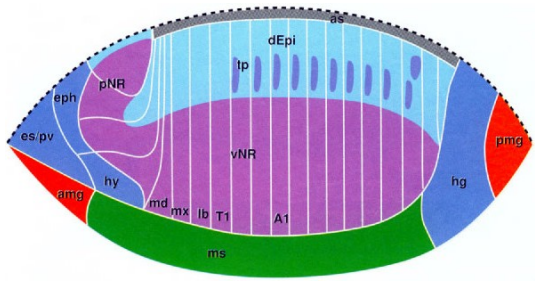

Figure 1.5: Fate map of the *Drosophila melanogaster* blastoderm The fate map is projected onto a planimetric reconstruction of the blastoderm. The upper dashed line represents the dorsal midline and the lower margin represents the ventral midline. A1 Abdominal segment 1; amg anterior midgut rudiment (endoderm); as amnioserosa; dEpi dorsal epidermis; eph epipharynx; es esophagus; hg hindgut; hy hypopharynx; lb labium; md mandible; ms mesoderm; mx maxilla; pmg posterior midgut rudiment (endoderm); pNR procephalic neurogenic region; pv proventriculus; vNR ventral neurogenic region; T1 thoracic segment 1; tp tracheal placodes. Diagram from Hartenstein (1993)

embryo, providing the first fate map (Conklin, 1905). In the 1930's Hörstaduius provided the fate map of the sea urchin embryo, also by cell lineage tracking. These two species were good to perform cell lineage tracking, as the number of cells of the embryo is relatively small and there are no major morphogenetic movements.

The first method to create a fate map of the *Drosophila* blastoderm were based on the analysis of gynandromorphs (Janning, 1978). Gynandromorphs are genetic mosaics of both male and female cells. Alfred H. Sturtevant analysed many gynandromorphs of *Drosophila simulans* and calculated how frequently two different parts of the embryo were of the same sex. He concluded that two different parts that were more often from different sexes should come from spatially separated cleavage nuclei (Janning, 1978). Importantly, this technique assumes that the position of a cell in the early embryo correlates with its developmental fate. Garcia-Bellido and Merriam improved the work of Sturtevant and using data from 379 gynandromorphs, calculated distances (in "sturt" units, in memory of Sturtevant) between the different embryo parts and construct the first fate map (Garcia-Bellido and Merriam, 1969). Also, histological methods (which consisted in following back to the blastoderm the location of larval organ precursors) and cell ablation methods (killing cells in the blastoderm and correlate its position with the position of the defects detected later) were used to create a fate map of the *Drosophila* blastoderm (Campos-Ortega and Hartenstein, 1985).

In the 1980's José A. Campos-Ortega and Volker Hartenstein combined a labelling techniques (injecting horseradish peroxidase) and histological methods to create a very precise fate map (Campos-Ortega and Hartenstein, 1985), which is still considered a standard modern reference (see Figure 1.5).

Gene expression maps

Techniques such as mRNA in situ hybridization allow to map gene expression patterns directly in the embryo. In situ hybridization is based on labelled probes that are complementary to the mRNA (or DNA) that is wanted to map (Gall and Pardue, 1969). The probe accumulates then only where the mRNA of interest is found. Another technique to map gene expression is the use of a reporter gene. A reporter gene, which codes

for a protein that can be easily identified (like the green fluorescence protein or beta-galactosidase), is linked to the regulatory region of the gene of interest so the reporter gene is going to be expressed where the gene of interest is expressed. Gene expression maps can also be used to create (or refine) fate maps (Gilbert, 2007). For example, if a gene is know to be expressed only in mesoderm precursors, mapping their gene expression in the early embryo will reveal where such mesodermal precursors are located.

Taking advantage of recent high-throughput methods of in situ hybridization (Tomancak et al., 2002; Weiszmann et al., 2009), the expression pattern of thousands of genes through *Drosophila* embryogenesis have been systematically determined (Tomancak et al., 2002, 2007; Hammonds et al., 2013), and publicly available databases have been developed (Tomancak et al., 2002; Kumar et al., 2011) so any researcher can see where and when a gene is expressed in the embryo.

These databases are suitable for computational image analysis, as the protocols used to produce the images are standardized (Tomancak et al., 2002) and the images can be aligned to an anatomical view (e.g., dorsal, lateral) (Kumar et al., 2011). With the expression patterns of thousands of genes, gene expression maps can be made using clustering techniques, showing regions where the expression of genes is more similar. Frise and collaborators made such analysis, processing thousands of in situ images of the blastoderm embryo and projected them into a virtual representation of the embryo made of ca. 300 triangles (Frise et al., 2010). After clustering the triangles based on their expression similarity, they produce a co-expression map that resembled the fate map shown in Figure 1.5.

Importantly, fate maps (done by lineage tracking, histological or ablation methods) and gene expression maps do not necessarily have to totally coincide. Fate maps inform about which cells in the early embryo will give rise to different cell types or tissues, even when at such early stage the cells can be genetically equivalent.

1.7 The Hourglass model in *Drosophila*

As I described briefly in section 1.2.3, von Baer stated in his "laws" that within a group of animals the general characteristics appear earlier in development, while the most special appear in late development (von Baer, 1828). This would lead to low morphological variation at early development, gradually increasing as development proceeds.

Other authors (Medawar, 1954; Slack et al., 1993; Duboule, 1994; Raff, 1996) proposed an alternative pattern in which there is great variation in early and late development, while the mid-development would show less variation. This pattern of variation (or conservation) has been called 'phylotypic egg-timer' (Duboule, 1994) and 'developmental hourglass' (Raff, 1996).

Duboule's concept of 'phylotypic egg-timer' was based in the concept of 'phylotypic stage' of Sander (1983), who coined this term to describe the convergence into a conserved segmented germ band stage in insects from very divergent early development (Sander, 1996). In vertebrates, there has been controversy around what should be the phylotypic stage (Ballard, 1981; Slack et al., 1993; Duboule, 1994). Richardson (1995) argued that indeed there is no single conserved stage in vertebrate's development and instead he proposed the term 'phylotypic period' instead.

Initially, two explanations for the hourglass model were proposed. Denis Duboule, after observing that the expression of the Hox genes seemed to coincide with the phy-

lotypic stage, he considered that this could not be a coincidence and proposed that the activation of the Hox genes was the cause for the morphological invariance (Duboule, 1994). In contrast, Rudolf A. Raff proposed that the phylotypic stage was the result of complex interaction between developmental modules at this stage (Raff, 1996).

There is an ongoing discussion about whether the hourglass model (HG), the von Baer law or some other pattern fits the divergence among developmental stages in phylogeny (Richardson et al., 1997; Poe and Wake, 2004; Kalinka and Tomancak, 2012).

Also, it is not clear if the HG, that seems to fit well in vertebrates and arthropods, would apply to other phyla (Raff, 1996) (Salazar-Ciudad, 2010). Salazar-Ciudad (2010) has proposed that different patterns of variation throughout development in metazoan groups would correlate with different developmental types (a classification based on the relative use of signalling and morphogenetic events).

Recently, the HG have received support from different gene expression studies. Kalinka et al. (2010) used micro-arrays for six Drosophila species and quantified expression divergence at different developmental stages. They found that gene expression was most conserved during the extended germ-band stage (considered the phylotypic period) and that the non-synonymous divergence per site (dN) correlated with their divergence measures. They also showed that most genes fit best to models incorporating stabilizing selection and proposed that natural selection acts to conserve patterns of gene expression during mid-embryogenesis (Kalinka et al., 2010).

The HG seems also to be reflected in the age of the transcriptome (mid-embryonic stage shows the older transcriptome; Domazet-Lošo and Tautz, 2010) and in the conservation of the regulatory regions (most conserved for genes expressed in mid-development; Piasecka et al., 2013).

Studies measuring the conservation of genes at the DNA sequence level also seem to support the HG. Davis et al. (2005) assessed whether proteins expressed at different times during *D. melanogaster* development varied systematically in their rates of evolution (comparing with *D. pseudoobscura*) and found that proteins expressed early in development and particularly during mid-late embryonic development evolve slower. This suggests, according to the authors, that embryonic stages from 12 to 22 hours are highly conserved between *D. melanogaster* and *D. pseudoobscura*, which is consistent with the HG. In a similar study, Mensch et al. (2013) calculated the dN/dS ratio for more than 2,000 genes among six *Drosophila* species, separating genes in three categories: maternal genes (genes whose products are left by the mother in the egg), genes expressed in early development and genes expressed in late development. They found that maternal genes and lately expressed zygotic genes show higher dN/dS ratios (i.e., are less conserved) than early expressed zygotic genes. Finally, it has also been found that genes expressed in the adult have higher dN/dS ratios than genes expressed in the pupa and those of the pupa have higher dN/dS ratios that those expressed in the embryo (Artieri et al., 2009).

Some limitations of these last studies is that they classify the genes in a few broad temporal categories that do not permit to precisely determine the temporal dynamics of conservation and that are based only in divergence data (dN/dS ratios between two species). A study that integrates polymorphism data from natural populations would improve the evolutionary interpretation of these patterns, as it would allow to estimate what proportion of the dN are adaptive (as explained in section 1.5.5). Measuring adaptation is specially relevant as some authors have argued that the HG is caused by different selection pressures in early and late development (Slack et al., 1993; Kalinka

and Tomancak, 2012; Wray, 2000)

1.8 Ciona

1.8.1 Ciona as a model

The ascidian *Ciona intestinalis*, a marine invertebrate animal, has a long history in developmental biology and evoutionary biology. Darwin highlighted the importance of the ascidians due to their close phylogenetic relationship to the vertebrates (REF). Also, it provided one of the first evidences of localized determinants of cell specification (Conklin, 1905). Although their adult form is a sessile filter feeder, its tadpole larva has characteristic features of the chordate group: a dorsal neural tube, a notochord surrounded by muscle and a ventral endodermal strand (Satoh, 1994). Ascidians show morphogenetic movements during gastrulation and neurulation similar to vertebrates and both share common genetic regulators of cell specification (REF). Their relative short life cycle, almost transparent body and rapid development facilitate many genetic techniques and are partly responsible for the re-emergence of *C. intestinalis* as model organism in developmental biology (Levin et al., 2012).

1.8.2 Current knowledge about Ciona development

The sequencing of the *C. intestinalis* genome (Dehal et al., 2002) facilitated its comparison with other vertebrate sequenced genomes and the analysis of gene expression through its life cycle. The *C. intestinalis* genome is only 160Mb and contains 16,000 genes, a gene number similar to the invertebrate *D. melanogaster* genome and only is half of the genes found in some vertebrates (REF). This low number of genes (compared to vertebrates) can be explained by the finding that many gene families or subfamilies have only one representative in *C. intestinalis* (Dehal et al., 2002). Relevant efforts have been made to describe the spatial expression patterns of individual genes (REF). The spatial expression patterns of >1,000 cDNA clones have been described using whole-mount in situ hybridization techniques at different developmental stages (Imai et al., 2004). Importantly, the developmental stages included cover a wide temporal range, e.g., blastula, gastrula and tapole stages (REF). Taking advantage of the ascidian invariant cleavage pattern and well described lineage analysis (Conklin, 1905; Nishida, 1987), the spatial expression of many genes have been described at the single cell level up to the early gastrula stage (REF), making this an invaluable resource to investigate the spatio-temporal dynamics of gene expression.

Aims of the study

Material and Methods

Results and Discussion

4.1 Comparative study between *Drosophila* and *Ciona* (I and II)

4.1.1 Compartmentalization

I estimated the degree of compartmentalization calculating the relative area or volume of expression of genes during development. My intention here was not to focus on individual genes, but to get a global overview of the embryo compartmentalization and differentiation processes based on expression data of thousands of genes, i.e., using a statistical approach.

One would expect, and it has been implicitly assumed (Carroll et al., 2001) (Davidson, 2001) that the compartmentalization of the embryo (as I measure it here) increases during development. However, the specific temporal dynamics of this increase in any species is not known. Neither is clear if the dynamics should be similar for different species, or for different groups of genes. As the development of *Ciona* and *Drosophila* are very different and it would be impossible to compare them stage-by-stage, I focused here in three major developmental periods: pre-gastrula, gastrula, and post-gastrula stages. These periods are easily recognizable in both species facilitating the comparative analysis.

I found that in both species, the relative area or volume decreased in a non-linear way (see Figs X). However, the timing of the major decrease was different. In *Drosophila* the major decrease occurred at very early development, from maternal to early gastrula stage (Fig. 4.1). Practically half of the genes in follows this decrease pattern: 46% of the genes were characterized as having a non-linear decrease in their relative area. In contrast, in *Ciona* the volume of expression decreases mostly after gastrulation (between the 112-cell and the early tailbud stage). However less dramatic, I found significant differences between the 32-cell and 64-cell stages, and between the 64-cell and 112-cell stages.

The difference in the timing of the major change on compartmentalization between species must relate to differences in their specific development. The earlier compartmentalization of *Drosophila* is most probably due to its derived early development, namely,

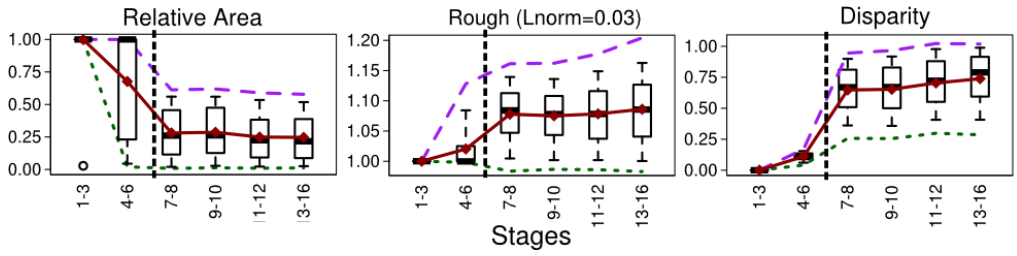

Figure 4.1: Measures in *Drosophila*. Distribution plot of the relative area of expression (left), roughness (center) and disparity (right) for all genes in each stage. Diamonds represent the mean, boxes the Inter Quartile Range (IQR). Whiskers 10 and 90 percentiles. Dashed line represents the max values and dotted line the min values (mean of the last and first decile, respectively). Stages on the x-axis, vertical dashed line represents gastrulation entry.

the syncytial blastoderm. During the blastoderm stage, approximately 4,000 cell nuclei can 'communicate' with each other only by TFs (Jaeger, 2011). The direct cross regulation of gene expression facilitates a rapid and highly dynamic process which seems to be responsible for the early spatial restriction of a great proportion of developmental genes. In contrast, *Ciona*'s early embryonic patterning is based on maternal determinants and signalling events mostly between neighbouring cells (Lemaire, 2009), which act in a combinatorial way (Hudson et al., 2007) to establish a unique TF combination in more than half of the blastomere pairs before gastrulation (Imai et al., 2006) determining most of their fates. Thus, even when in *Ciona* most of the cell fates are already determined (by the specific combination of a fraction of TFs) and the embryo can be said to be already highly compartmentalized, this is not evident at the global level of gene expression, which I am measuring here.

Therefore, the 'delay' of compartmentalization observed in *Ciona* could be explained by the relatively slower process of signal transduction (as in *Ciona*) compared to the gap gene network (in *Drosophila*).

4.1.2 Disparity

As the relative area (or volume) of expression informs on how genes are expressed in progressively smaller regions in the embryo, the disparity can inform about how different regions of the embryo express increasingly different combinations of genes. Therefore, both measures reflect slightly different aspects of complexity that are independent from each other. A decrease in the volume of expression of genes does not necessarily imply an increase in spatial disparity: genes could decrease their volume of expression but end up restricted to the same parts of the embryo. If the majority of genes would be expressed ubiquitously (this is large volume), however, then the mean disparity between its regions would be necessarily low.

My results show that in each species, the global disparity pattern is similar to the relative area or volume patterns. Therefore, in *Drosophila* the disparity increases mostly in the transition from the maternal to early gastrula and in Ciona this major change occurs after gastrulation.

It is important to notice that these measures should not necessarily correlate, as it could be that between two stages the relative area of expression decreases but not the

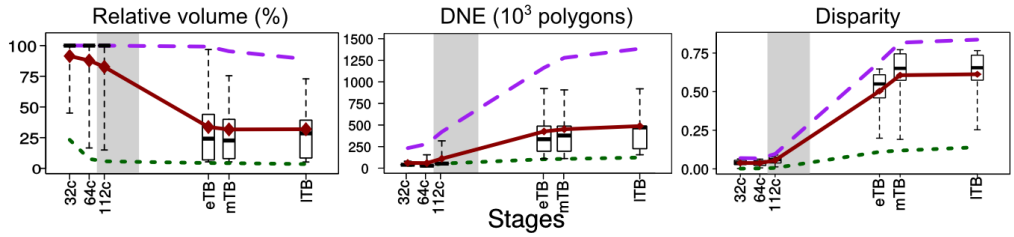

Figure 4.2: Measures in *Ciona*. Distribution plot of the relative volume of expression (left), DNE (center) and disparity (right) for all genes in each stage. Diamonds represent the mean, boxes the IQR. Whiskers 10 and 90 percentiles. Dashed line represents the max values and dotted line the min values (mean of the last and first decile, respectively). Stages on the X-axis (s32c, 32-cells; s64c, 64-cells; s112c, 112-cells; eTB, early tailbud; mTB, mid tailbud; lTB, late tailbud). Grey area represents gastrulation period.

disparity if the genes are expressed in the same part of the embryo or vice-versa. In *Ciona* I found an example of such case, when there is no perfect correspondence between the relative volume and the disparity of expression: disparity increased significantly between early to mid-tailbud stages but no significant differences between the relative volume of expression of these stages were found (II, Fig. 3A). This means that, on average, genes are expressed in a similar number of tissues in these stages, but in the mid tailbud the combination of genes expressed in these tissues are more different between each other.

This shows that the disparity measure is useful specially when is complemented with the relative area (or volume) measure to describe the compartmentalization of the embryo.

4.1.3 The leading role of TFs and GFs (and other signalling molecules)

I wanted the test in both species if TFs and GFs showed an earlier compartmentalization or greater disparity when compared to the rest of the genes. This would be expected from their allegedly leading role in early pattern formation.

Using a GOterm analysis in *Drosophila*, I found that TFs (GO:0003700) and GFs (GO:0008083) showed smaller relative area of expression that the rest of genes in the blastoderm stage (Fig. 4.1). The TFs are also expressed in smaller areas than the rest of the genes in all subsequent stages, while the GFs are expressed in smaller areas at the blastoderm (stage 4-6) and extended germ band stages (stage 9-10 and 11-12) (I, Fig.4). A similar result for TFs was reported in Drosophila by Hammonds et al. (2013). They made an extensive analysis of TFs expression using manual annotation of gene expression based on an anatomical controlled vocabulary and classifying every gene as ubiquitous, patterned, ubiquitous-patterned, or maternal (from the BDGP database; Tomancak et al., 2007). They found that the fraction of TFs expressed in a restricted pattern (assigned to a tissue) was higher, when compared to other genes, in all zygotic stages with the exception of the stage 13-16.

The results I show for stages 4-6, 7-8, 9-10 and 11-12 are consistent with Hammonds et al., as the higher proportion of the TF genes showing a restricted or tissue-specific expression pattern would imply that TFs are expressed in smaller areas in the embryo. For the 13-16 stage, contrary to these authors, I showed that the TFs are highly compartmentalized. This might indicate a limitation of the annotation method used by Hammonds et al., to capture the high spatial compartmentalization of the TFs in this stage.

In *Ciona*, I performed a similar analysis using the categorization of TFs and signaling molecules (SIGs) made by Imai et al. (2004). SIGs consist of genes of receptor tyrosine kinase (RTK) pathways such as FGFs and intracellular signalling molecules such as MAPK, Notch, Wnt, TGFβ, Hedgehog and genes in the JAK/STAT pathways (Imai et al., 2004).

As expected, TFs volume of expression decreased faster than non-TFs. The TFs showed lower volume of expression in the 64-cell and 112-cell stages (II, Fig. 3B). The results are similar for maternal and zygotic genes (maternal/zygotic classification based on Matsuoka et al., 2013; II,Fig. S1). I then compared TF families and found that six TF families showed lower relative volume in the early gastrula (BZIP, T-box, bHLH, HMG, Nuclear Receptor, and 'Other-TFs') but only T-box genes showed a lower relative volume from the 32-cell stage until gastrula (II,Fig. S2).

The results obtained for the T-box gene family (conserved in metazoan and several non-metazoan lineages (Sebé-Pedrós et al., 2013)) are consistent with the known important role these genes have in diverse metazoan species early cell fate specification (reviewed in: Papaioannou, 2014; Showell et al., 2004. Examples of T-box genes in *Ciona* are Tbx6 and *brachyury*, crucial for muscle tissue formation (Mitani et al., 1999; Nishida, 2005) and for notochord specification (Yasuo and Satoh, 1998), respectively. I also found that the SIGs showed significant lower relative volume of expression than the rest of the genes in the 32-cell, 64-cell, and 112-cell stages (II, Fig. 3B). Specifically, in the 64-cell stage RTK-MAPK, Wnt and TGFβ families showed significant higher disparity in the 64 cells stage, suggesting a predominant role of these pathways in the patterning of the embryo at this stage. This is consistent with known short range induction events by nodal and various FGFs, which are part of the TGFβ and RTK-MAPK signalling pathways, respectively (Lemaire et al., 2008).

In general, the fact that in these two species that display a very different development TFs and GFs (or SIGs in the case of *Ciona*) are more compartmentalized than the rest of the genes precisely in the stage before entering gastrulation, is consistent with these genes having a special role in pattern formation and compartmentalization. Therefore, my results support the hypothesis of the leading role of TFs and GFs in driving pattern formation and compartmentalization in the early embryo.

4.1.4 2D and 3D roughness analyses

I wanted to test the hypothesis of gene expression spatial patterns becoming more complex during development. In here, with gene expression spatial pattern I mean the spatial distribution of the cells or tissues expressing a specific gene.

Considering that I had information in 2D in *Drosophila* and in 3D in *Ciona*, it was necessary to apply a specific method for each species. For *Drosophila*, I developed a 'roughness' measure (Salvador-Martínez and Salazar-Ciudad, 2015) which accounts for the curvature of the contour in a 2D gene expression pattern, normalizing it with the contour of a circle of the same perimeter. In *Ciona*, I applied a similar measure of curvature in 3D, called 'Dirichlet normal energy' (DNE), which quantifies the deviation of a surface from being planar (Bunn et al., 2011). Both measures not only inform about the overall imbrication or convolution of the shape of a gene expression pattern, but also do it at different spatial scales.

In the following paragraphs, to improve the readability of the text, I will refer to the roughness measure implemented in *Drosophila* as 2D roughness and to the DNE measured used in *Ciona* as 3D roughness.

The results show that both 2D and 3D roughness increase in a non-linear way during development. As with the compartmentalization and disparity, what changes between species is where the major change is found.

In *Drosophila*, the major change is found in the transition from the blastoderm to the early gastrula (Fig. 4.1). When analysing the maximal values (mean of the last decile) it can be seen that they increase initially in the pre-gastrula, reach a stationary phase at mid-embryogenesis and finally increase in the last stages. As I mentioned in the literature review (section X), the maximal values are informative about the overall morphological spatial complexity of the embryo in a given stage. When comparing roughness at different spatial scales (I, Methods), I found that in the last three stages

the roughness values are significantly higher at smaller spatial scales is significantly higher that at the higher spatial scales.(Fig. S2 in article I).

In Ciona, the 3d roughness increase throughout development (II, Fig. 5), with the major change between the 112-cell and the early tailbud (Fig. 4.2) The max (mean of the last decile) values increase substantially already between the 64 and 112 cells stages (with 1000 and 10000 polygonal faces), while the min values (mean of the first decile) remain practically constant during development, showing that the most complex patterns in each stage get increase their DNE value but there is always a proportion of very simple expression pattern. Also, I found that at low spatial scales (1000 and 10000 polygons per mesh; II, Fig. 5) I found that the mean DNE of the late tailbud is higher than at the mid tailbud (one-way ANOVA pvals < 0.05).

In summary, this results show that the complexity of distribution in space of cells/tissues expressing a gene increases through development, and that these complexity (measured with the 2D and 3D roughness) increase in both *Ciona* and *Drosophila* in a similar way than the other two measures, compartmentalization and disparity. Also, by analysing the roughness at different scales, I found compelling evidence that complexity may be increasing not only through all the development but also that it does at finer spatial scales over time.

4.1.5 Synexpression territories

I wanted to explore in both species the relative degree of similarities between different parts of the embryo within and between different developmental stages.

To do this I used two different approaches, based on the differences between the databases I used for each species. In *Drosophila* I used the polar regions with which I computationally divided the embryo. In *Ciona*, I took advantage of the available information at the individual cell/tissue.

In both cases I used a clustering algorithm to produce dendrogram representing the relative degrees of similarities between all regions of different stages at the same time (Fig. 4.3 and Fig. 4.4). I will refer to the regions that clustered together as 'synexpression territories' (STs).

In *Drosophila*, after cutting the dendrogram at a specific threshold and filtering STs with less than 50 genes expressed with a minimum specificity (see methods in I for a detailed description), 30 STs were selected for further analyses (Fig. 4.3 B).

Finally, I grouped the STs in eight 'meta-territories', as I wanted not only to see how the regions in the embryo formed different STs, but also how different STs cluster with each other, as this is informative of the degree of differentiation between stages. If STs cluster with other STs in the same stage, it would mean that the majority of genes change their expression in a similar way over time independently of where they are. If STs cluster with other STs in the same part of the embryo in successive stages, it would mean that this part of the embryo has expression dynamics independent from other parts of the embryo, which would be expected in already differentiated cells/tissues.

The results show that stages 1-3 and 4-6 each one form a ST. If a cut-off is selected so that stage 4-6 is divided in four sub-territories (I,Fig. S3) the embryo splits in four parts: anterior, posterior, dorsal and ventral. This correspond to a nearly Cartesian system one could expect from the two signalling systems known in the earliest patterning in *Drosophila* (the A/V and D/V signalling cascades; (Gilbert, 2014)). The STs seem to coincide with the known embryo fate map (see Fig. 4.3 D; Hartenstein, 1993) and many

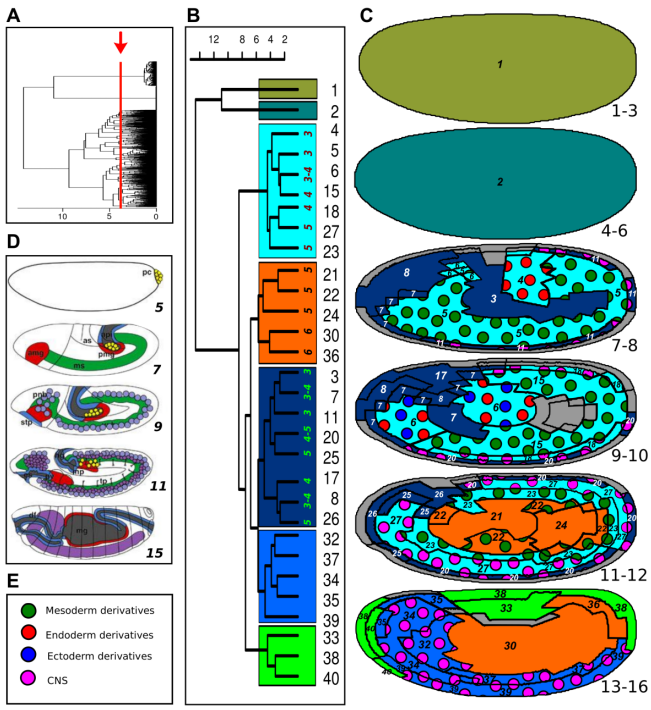

Figure 4.3: Synexpression territories (ST). (A) Dendrogram produced by hierarchical clustering on a similarity matrix (pearson's correlation) of all the embryo regions of the six stages. Red line shows the cut-off to produce 40 STs. (B) Dendrogram reconstructed using only territories with at least 50 genes with a minimum specificity (I, methods). The coloured boxes show the main branches of the dendrogram. The number indicated inside the boxes represent the stages each ST corresponds to (3 is stage 7-8, 4 is stage 9-10, 5 is stage 11-12 and 6 is stage 13-16). The ST number is at the right. (C) STs mapped onto the embryo. Gray regions have less than 50 genes expressed. Background color refers to which 'meta-territory' (in B) each ST is part of. Coloured circles represent GOterm enrichment of a specific tissue/germ layer derivative (shown in E). Stages in the lower-left part of each embryo. From stage 7-8, the ST number (as in B) is indicated. (D) Hartenstein's embryo schemes (Hartenstein, 1993) with their respective stages in the left upper part. (E) Colour code of specific tissue/germ layer derivative used in C.)

of them are enriched with GOterms that coincide with their expected fate. For example, in stage 7-8 (just after gastrulation) there is a ST that corresponds spatially with the germband and is enriched with mesodermal GOterms (Fig. Fig. 4.3 C).

There are two meta-territories that appear in the last stage (light blue and green, Fig. 4.3 C), which suggests that the tissues/organs related to those STs differentiate quite late. One meta-territory is enriched with terms related to epidermis such as cuticle development ('chitin catabolic process' [GO:0006032] and 'cuticle development' [GO:0042335] STs 33 and 38), which coincides with cuticle deposition by epithelial cells during stage 16 (Ostrowski et al., 2002). The other meta-territory corresponds spatially with the CNS of the embryo and is indeed enriched with CNS GO-terms. The CNS territory is enriched with GOterms like 'dendrite morphogenesis' (GO:0048813) and 'axon guidance' (GO:0007411).

In *Ciona*, because gene expression information in tailbud stages is based on tissues

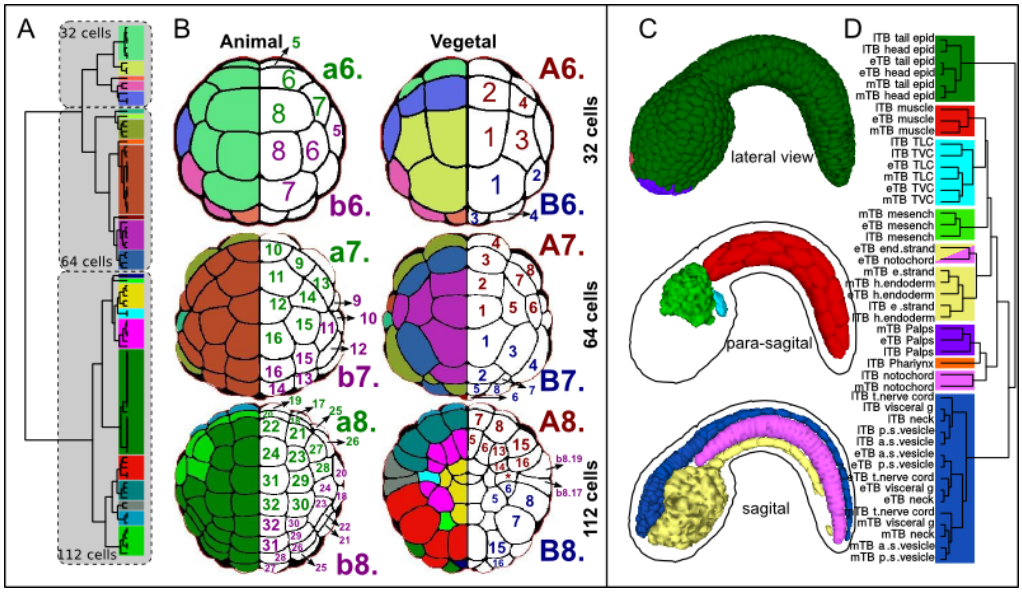

Figure 4.4: Ciona synexpression territories. (A) Dendrogram produced by hierarchical clustering of cells in 32-cell, 64-cell and 112-cell stages. Dashed boxes show that STs cluster by stage. Coloured boxes show the cut-off to produce 24 STs. (B) Names of cells (Conklin nomenclature; Conklin, 1905) indicated with a prefix shown at right. STs in the 32 cells, 64 cells and 112 cells stages (top, middle and bottom, respectively). Colour refers to which ST of the dendrogram (in A) each cell is part of. Animal view based on Nicol and Meinertzhagen (1988) and vegetal view based on Cole and Meinertzhagen (2004). The cell marked with a star (*) is the A7.6 cell, that in this analysis represents their descendant cells (A8.11 and A8.12). C) Dendrogram produced by hierarchical clustering of tissues in early, mid and late tailbud stages. The coloured boxes show the cutoff to produce 10 STs. (D) STs in the tailbud stages shown in a lateral, para-sagital and sagital views of a mid tailbud 3D embryo model (from Nakamura et al., 2012). Colour refers to which ST of the dendrogram (in C) each tissue is part of.

and not on individual cells as the early stages, I analysed the STs of these stages separately (II, Methods).

If in the early stages, three 'meta-territories' are formed, each one would correspond to one stage, i.e., STs in early stages cluster by stage. Thus, even if at the first three stages a high proportion of blastomeres express a nearly unique combination of transcriptional factors (Imai et al., 2006), the bulk change in gene expression is common to all blastomeres. Within each early stage, STs coincides very well with the know fate map (II, Fig 6A; II, Fig. S8), with some exceptions I will describe in the next subsection.

In contrast, in tailbud stages practically all STs cluster by tissue/cell type, which indicates that the in early tailbud, most tissues are already quite differentiated. This is consistent with studies analysing these stages at the level of individual or small sets of genes (Corbo et al., 1997; Di Gregorio and Levine, 1999).

My analysis in the early stages is similar to the one made by Imai et al. (2006), who used the expression profile of 53 zygotically TFs in single cells in the 16, 32, 64, and 112-cell stages, to perform a hierarchical clustering (for each stage separately). It is different in two aspects: I performed the clustering using the blastomeres of different stages and my analysis is not restricted to TFs. As I said previously, using various stages is informative of the overall differentiation process and can be used to discern between

differentiation scenarios, as the differences between early and tailbud stages I found here.

The main difference between species is that, in *Drosophila* the differentiation process continues throughout whole embryogenesis (as new STs were formed until the last stage I analysed) and different organs differentiate at different developmental times. In contrast, the *Ciona* embryo seems to be already genetically differentiated at the early tailbud (as the STs of all the tissues in the tailbud stages cluster together) so the last embryo stages consist only of moderate morphogenetic movements (mainly cell elongation; Hotta et al., 2007). Therefore, the ST analysis is a valuable tool, based on differential gene expression, to get a global perspective on the local differentiation of the embryo.

4.2 Main spatio-temporal profiles of gene expression in *Drosophila* (I)

With a time series cluster analysis (Ernst and Bar-Joseph, 2006) of the relative area of expression, I found the eight main spatio-temporal profiles of gene expression in the embryonic development of *Drosophila* (I, Fig. 5). As expected, the most common profile (n=297 genes) follows the global profile of non-linear decrease in the first stages (I, Fig 5).

Among the rest of profiles, I found both linear increase and decrease profiles and a 'hill-like' profile (initial increase and further decrease with the higher values at stage 7-8) The linear decrease profile (n=167 genes) was enriched with 'mitotic cell cycle' (GO:0000278), 'RNA processing' (GO:0006396) and 'chromatin modification' (GO:0016568) GOterm genes, highlighting biological processes that first are present in the whole embryo and become more and more restricted in space as development proceeds. The 'mitotic cell cycle' term, for example, most likely relates to the fast mitotic cycles in the earliest embryo. During stage 1-3 nine fast and synchronic mitotic divisions take place in the entire embryo, then in stage 4-6 mitotic divisions 10-13 occur more slowly, almost synchronically. The 14th cycle, zygotically controlled, is long and of different durations in the embryo.

With a temporal co-expression cluster analysis using microarray data through the life cycle of *D. melanogaster*, Arbeitman et al. (2002) found that most cell cycle genes were expressed at high levels during the first 12h, but only a few are expressed at high level thereafter. My analysis is consistent with this, as I found that the profile of linear decrease (I, Fig. 5A) is enriched with such genes. In this sense, this study is complementary to Arbeitman et al., and adds the spatial dimension to their temporal expression profiles.

4.3 Discrepancies between fate map and STs (II)

I found a few cases in *Ciona* in which cells with the same fate where contained in different STs. This would be the case of: 1) cells whose fate is disproportionally affected or determined by a small number of genes (as this analysis reflect quantitative differences at the level of hundreds of expressed genes but can not distinguish between the relative importance of each gene) or 2) cells that although having a restricted fate at a certain stage their differentiation is not complete (at the level of gene expression).

4.3 Discrepancies between fate map and STs (II)

This analysis could not be made in *Drosophila* as the gene expression data is not a the single level resolution.

An example of the latter is a ST in the 112-cell stage (in magenta; 4.4 B; II, Fig. S8) that contains precursors of the notochord (A8.5, A8.6, A8.13, and A8.14, B8.6) and mesenchyme (B8.5) (Tokuoka et al., 2004). The latter come from a secondary notochord/mesenchyme bipotential cell (B7.3). It has been reported that the expression of Twist-like 1, necessary for mesenchyme differentiation, starts at this stage (Imai, 2003). This evidence, together with the inclusion of the mesenchyme cell in this otherwise exclusively notochord territory (primary and secondary), seems to indicate that the differentiation of cell pair B8.5 as mesenchyme is still incomplete at this stage.

Gene expression dynamics in cell-lineages

I analysed the gene expression similarity between lineage-related cells (i.e., between daughters cells and between mother/descendants cells) in the early stages (II, Fig. 8). In general, cells are more closely genetically to their sister cells than to their mother/descendants, which is reflected in the clustering of STs by stages discussed before. I found also that at the 64-cell stage, cells that show more genes expressed differently than their ancestors are neural fated cells, which might be related with the fact the unrestricted state of these cells at this stage (i.e., their descendants will give rise to different cell fates).

4.4 Adaptation in *Drosophila* embryogenesis (III and IV)

I combined the Synexpression Territories (STs) approach with genome-wide coding-region polymorphism data (from the DGRP database) and the coding-region divergence between *D. yakuba* and *D. melanogaster* in order to estimate the proportion of adaptive non-synonymous substitutions (ω_α) in the genes expressed in each ST (n=589 genes; III, Methods).

Using this approach, I could chart a spatial map of natural selection acting on *Drosophila*'s embryo anatomy. I complemented this with a analysis using available annotation of gene expression (n=2,835 genes) using a controlled vocabulary of anatomical structures from the BDGP database (Tomancak et al., 2007).

The results showed a few STs with significant higher or lower ω_α (permutation test; III, Methods)

4.4.1 STs or anatomical terms with high ω_α

STs 13 and 32 (ST number comes from the hierarchical clustering algorithm), which showed a higher ω_α, seem to correspond to the forming foregut and hindgut (stage 11-12) and to the CNS (stage 13-16) respectively. To explore if ST 32 high ω_α was indeed related to the CNS, I separated the genes CNS or not-CNS related. I found that both groups showed a high ω_α, which suggests that in addition to the CNS, another structure in the anterior region would be under positive selection. Using the anatomical terms approach, no anatomical terms related to the CNS were found to have high ω_α with the initial criteria. I therefore applied a more stringent criterion to consider genes as part of an anatomical term (before a gene could have a maximum of seven anatomical terms associated instead of a more stringent number of three) and found that 'Embryonic brain' showed high ω_α (permutation test, $p = 0.046$). Also, with the anatomical terms approach, I found that genes associated with 'Gonads', in the last stage, clearly showed evidence of adaptive evolution (III, Figure 2), which is consistent with previously reported high rates of adaptive substitution in the testes (Akashi, 1994; Civetta and Singh, 1995; Nuzhdin et al., 2004; Pröschel et al., 2006)

Finally, ST 24 (stage 11-12) that seems to corresponds to part of the trunk mesoderm marginally significant high ω_α ($p = 0.061$). A similar result was obtained with for the anatomical term 'Trunk mesoderm' in stage 9-10 ($p = 0.087$).

4.4.2 STs or anatomical terms with low ω_α

STs 20 and 29 with showed low ω_α, seem to correspond to the forming midgut (stage 11-12) and to the forming larval digestive system (stage 13-16) respectively. Similar results are found when using the anatomical term approach, as low ω_α was found in many anatomical terms related to the digestive system in the last stage: 'Embryonic midgut', 'Embryonic salivary gland', 'Embryonic hindgut', 'Embryonic proventriculus'. Also, combining three related anatomical terms, 'Embryonic foregut', 'Embryonic epipharynx' and 'Embryonic hypopharynx', that separately did not have enough genes to be considered in the analysis, showed low ω_α.

The lack of adaptive change in the forming digestive system might reflect their relative enrichment in metabolic genes (Marianes and Spradling, 2013). The coding regions of metabolic genes have been found to be more conserved than non-metabolic genes (Peregrín-Alvarez et al., 2009).

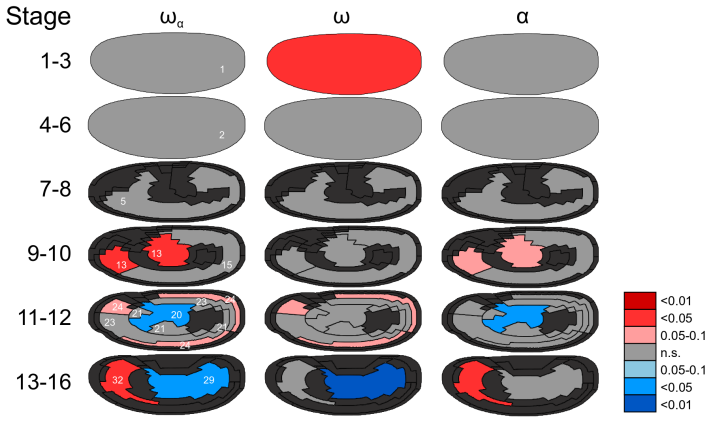

Figure 4.5: ω_α on embryonic territories over space and time. Territories drawn in red in the central column mark significantly high ω_α while those in blue mark significantly low ω_α in space in each of the 6 developmental stages (rows). Other columns depict α, the proportion of base substitutions fixed by natural selection, and ω, the rate non-synonymous substitutions relative to the mutation rate. Territories in dark gray are territories without enough specific genes to be analyzed. The statistical was calculated by a permutation test using all the genes analyzed (see Material and methods). Territory 13 in stage 9-10 (ω_α: 0.059, p = 0.045). Territory 20 from stage 11-12 (ω_α: 0.022, p = 0.048; α: 0.259, p = 0.028). Territory 24 from stage 11-10 (ω_α: 0.070, p = 0.061). Territory 29 from stage 13-16 (ω_α: 0.037, p = 0.047; ω: 0.074, p < 0.001). Territory 32 from stage 13-16 (ω_α: 0.068, p = 0.044; α: 0.71, p = 0.04).

4.4.3 Transcriptome age index and other genomic determinants

Then, I wanted to test if the 'age' of the genes, also differed between different parts of the embryo and how this related to the adaptation results. For that, I used the phylostratigraphic maps of *D. melanogaster* (Drost et al., 2015), that assign a phylogenetic age to each protein-coding gene based on the phylogenetic level at which orthologs for a gene are found (a young found only in Drosophilids would be very young, of age 1). Also, I used a modified version of the Transcriptome Age Index (TAI; Domazet-Lošo and Tautz, 2010) and applied it to the polar regions and STs (see III, methods). TAI is low for regions expressing old genes and large for regions expressing young genes. I found that STs with low ω_α express (on average) older genes (high TAI values). Similar results were found for anatomical structures (III, Fig. S2). I also found that in stage 13-16 the mean phylogenetic age of the genes expressed in the endoderm is lower than in other germ-layers, specially compared to the ectoderm (Fig. 4.6). Similar TAI results between germ-layers were found by Domazet-Loso et al. (2007) but without stages comparisons. The correlation between adaptation and gene age fit the expectation that older genes would more likely perform essential functions than younger genes, and that, as older genes would have been moulded by natural selection for longer times would be therefore more close to optimality (assuming that this function is conserved). Therefore,

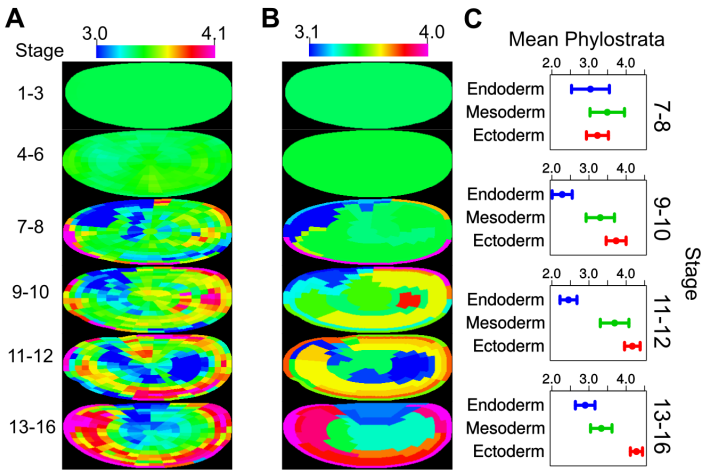

Figure 4.6: The center of the embryo expresses older genes. (A) Heatmaps showing the transcriptome age index (TAI) im polar regions (B) Heatmaps showing the TAI for STs. (C) Mean phylostrata of genes assigned to each germ layer. Circles represent the mean and whiskers the SEM.

more room for changes would be expectable in embryo regions with a larger proportion of younger genes.

I also found that embryo polar regions with high ω_α have low codon bias (previously reported by Sharp, 1991; Betancourt and Presgraves, 2002; Haerty et al., 2007) and that, as Plotkin and Kudla (2011) previously found, regions high codon bias show have high levels of gene expression (average RNAseq levels per region; III, methods). To clarify the relation between these three variables, I fitted a multivariate linear regression and found that embryo regions with high ω_α exhibit low codon bias relative to what would be expected from their gene expression levels (III, Fig 5). The negative correlation between codon bias and protein adaptation that I found would be expected given that, an adaptive aminoacid change in a protein would be probably different from a change that would increase codon usage efficiency (Hershberg and Petrov, 2008; Presnyak et al., 2015).

4.4.4 Selective constraint in late embryogenesis

Analysing the RNAseq developmental data from the modENCODE project (Graveley et al., 2011) with the DFE-alpha method, I found that from hour 10 until 24 of embryogenesis show significant low ω_α and ω (see Fig X in section X), which would be consistent with the low rate of adaptive change seen in many anatomical structures in stage 13-16 (as stage 13-16 of BDGP roughly maps to RNA-seq samples em10-12 hr, em12-14 hr, em14-16 hr, and em16-18 hr of modENCODE; Hammonds et al., 2013). Therefore, by combining different approaches, I could identify that the proteins produced in late embryogenesis change less their aminoacid sequence (i.e., are more conserved). This phenomenon, of some proteins evolving slower, has been called 'selective constraint' and has been linked to the higher degree of functionality of such proteins (Kimura, 1983). Most importantly, I could identify which specific anatomical structures expressed genes with a higher degree of conservation.

4.5 Adaptation trough *Drosophila* life cycle (IV)

Concluding Remarks

References

Abzhanov, A., W. P. Kuo, C. Hartmann, B. R. Grant, P. R. Grant, and C. J. Tabin (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature 442*(7102), 563–7.

Akashi, H. (1994). Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. *Genetics 136*(3), 927–35.

Alberch, P. (1982). Developmental Constraints in Evolutionary Processes. In J. T. Bonner (Ed.), *Evolution and Development*, Volume 22 of *Dahlem Workshop Reports*, pp. 313–332. Springer Berlin Heidelberg.

Alberch, P. (1991). From genes to phenotype: dynamical systems and evolvability. *Genetica 84*(1), 5–11.

Amundson, R. (2005). *The Changing Role of the Embryo in Evolutionary Thought: Roots of Evo-Devo*. Cambridge Studies in Philosophy and Biology. Cambridge University Press.

Apter, M. and L. Wolpert (1965). Cybernetics and development I. Information theory. *Journal of Theoretical Biology 8*(2), 244–257.

Arbeitman, M. N., E. E. M. Furlong, F. Imam, E. Johnson, B. H. Null, B. S. Baker, M. A. Krasnow, M. P. Scott, R. W. Davis, and K. P. White (2002). Gene expression during the life cycle of Drosophila melanogaster. *Science (New York, N.Y.) 297*(5590), 2270–5.

Arias, A. M. (2008). Drosophila melanogaster and the Development of Biology in the 20th Century. In C. Dahmann (Ed.), *Drosophila: Methods and Protocols*, pp. 1–25. Totowa, NJ: Humana Press.

Aristotle (1979). *Generation of Animals*. Harvard University Press.

Arthur, W. (2010). *Evolution: A Developmental Approach*. Wiley.

Artieri, C. G., W. Haerty, and R. S. Singh (2009). Ontogeny and phylogeny: molecular signatures of selection, constraint, and temporal pleiotropy in the development of Drosophila. *BMC biology 7*(1), 42.

Ballard, W. W. (1981). Morphogenetic Movements and Fate Maps of Vertebrates. *American Zoologist 21*(2), 391–399.

Bateson, W. (1894). *Materials for the study of variation treated with especial regard to discontinuity in the origin of species*. London: Macmillan and co.,.

Bell, G. and A. Mooers (1997). Size and complexity among multicellular organisms. *Biological Journal of the Linnean Society 60*(3), 345–363.

Betancourt, A. J. and D. C. Presgraves (2002). Linkage limits the power of natural selection in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America 99*(21), 13616–20.

Bonner, J. T. (Ed.) (1982). *Evolution and Development: Report of the Dahlem Workshop on Evolution and Development Berlin 1981, May 10–15*. Dahlem Workshop Report. Springer Berlin Heidelberg.

Bonner, J. T. (2004). Perspective: The Size-Complexity Rule. *Evolution 58*(9), 1883–1890.

Buffon, G. L. L. (1807). *Buffon's Natural history: Containing a theory of the earth, a general history of man, of the brute creation, and of vegetables, minerals, &c. &c.* Number v. 3 in Buffon's Natural History: Containing a Theory of the Earth, a General History of Man, of the Brute Creation, and of Vegetables, Minerals, &c. &c. Printed for the Proprietor, and sold by H. D. Symonds.

Bunn, J. M., D. M. Boyer, Y. Lipman, E. M. St Clair, J. Jernvall, and I. Daubechies (2011). Comparing Dirichlet normal surface energy of tooth crowns, a new technique of molar shape quantification for dietary inference, with previous methods in isolation and in combination. *American journal of physical anthropology 145*(2), 247–61.

Campos-Ortega, J. A. and V. Hartenstein (1985). *The Embryonic Development of Drosophila melanogaster*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Carroll, S. B. (2005). *Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom*. ISSR library. W.W. Norton & Company.

Carroll, S. B., J. K. Grenier, and S. D. Weatherbee (2001). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Malden, MA.: Blackwell Science.

Casanova, J. and G. Struhl (1989). Localized surface activity of torso, a receptor tyrosine kinase, specifies terminal body pattern in Drosophila. *Genes & development 3*(12B), 2025–38.

Chipman, A. D. (2015). Hexapoda: Comparative Aspects of Early Development. In *Evolutionary Developmental Biology of Invertebrates 5*, pp. 93–110. Vienna: Springer Vienna.

Cho, P. F., C. Gamberi, Y. A. Cho-Park, I. B. Cho-Park, P. Lasko, and N. Sonenberg (2006). Cap-dependent translational inhibition establishes two opposing morphogen gradients in Drosophila embryos. *Current biology : CB 16*(20), 2035–41.

Civetta, A. and R. S. Singh (1995). High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in Drosophila melanogaster and Drosophila virilis group species. *Journal of molecular evolution 41*(6), 1085–95.

Cole, A. G. and I. A. Meinertzhagen (2004). The central nervous system of the ascidian larva: mi-

totic history of cells forming the neural tube in late embryonic Ciona intestinalis. *Developmental biology 271*(2), 239–62.

Conklin, E. G. (1905). The organization and cell-lineage of the ascidian egg. *Journal of the Academy of natural sciences of Philadelphia 13*, 1–119.

Corbo, J., A. Erives, A. Di Gregorio, A. Chang, and M. Levine (1997). Dorsoventral patterning of the vertebrate neural tube is conserved in a protochordate. *Development 124*(12), 2335–2344.

Cotterell, J. and J. Sharpe (2010). An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. *Molecular systems biology 6*, 425.

Crick, F. (1958). On Protein Synthesis. In *The Symposia of the Society for Experimental Biology*, pp. 138–166.

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature* (227), 561–563.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray.

Davidson, E. H. (2001). *Genomic Regulatory Systems: In Development and Evolution*. Academic Press.

Davis, J. C., O. Brandman, and D. A. Petrov (2005). Protein evolution in the context of Drosophila development. *Journal of molecular evolution 60*(6), 774–85.

Dehal, P., Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A. Di Gregorio, M. Gelpke, D. M. Goodstein, N. Harafuji, K. E. M. Hastings, I. Ho, K. Hotta, W. Huang, T. Kawashima, P. Lemaire, D. Martinez, I. A. Meinertzhagen, S. Necula, M. Nonaka, N. Putnam, S. Rash, H. Saiga, M. Satake, A. Terry, L. Yamada, H.-G. Wang, S. Awazu, K. Azumi, J. Boore, M. Branno, S. Chin-Bow, R. DeSantis, S. Doyle, P. Francino, D. N. Keys, S. Haga, H. Hayashi, K. Hino, K. S. Imai, K. Inaba, S. Kano, K. Kobayashi, M. Kobayashi, B.-I. Lee, K. W. Makabe, C. Manohar, G. Matassi, M. Medina, Y. Mochizuki, S. Mount, T. Morishita, S. Miura, A. Nakayama, S. Nishizaka, H. Nomoto, F. Ohta, K. Oishi, I. Rigoutsos, M. Sano, A. Sasaki, Y. Sasakura, E. Shoguchi, T. Shin-i, A. Spagnuolo, D. Stainier, M. M. Suzuki, O. Tassy, N. Takatori, M. Tokuoka, K. Yagi, F. Yoshizaki, S. Wada, C. Zhang, P. D. Hyatt, F. Larimer, C. Detter, N. Doggett, T. Glavina, T. Hawkins, P. Richardson, S. Lucas, Y. Kohara, M. Levine, N. Satoh, and D. S. Rokhsar (2002). The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. *Science (New York, N.Y.) 298*(5601), 2157–67.

Di Gregorio, A. and M. Levine (1999). Regulation of Ci-tropomyosin-like, a Brachyury target gene in the ascidian, Ciona intestinalis. *Development 126*(24), 5599–5609.

Domazet-Loso, T., J. Brajković, and D. Tautz (2007). A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends in genetics : TIG 23*(11), 533–9.

Domazet-Lošo, T. and D. Tautz (2010). A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature 468*(7325), 815–8.

Driesch, H. (1892). Entwicklungsmechanische Studien: I. Der Werthe der beiden ersten Furchungszellen in der Echinogdermenentwicklung. Experimentelle Erzeugung von Theil- und Doppelbildungen. *Zeitschrift für wissenschaftliche Zoologie* (53), 160–178.

Driesch, H. (1894). *Analytische theorie der organischen entwicklung*. W. Engelmann.

Driever, W. and C. Nüsslein-Volhard (1988). The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell 54*(1), 95–104.

Drost, H.-G., A. Gabel, I. Grosse, and M. Quint (2015). Evidence for Active Maintenance of Phylotranscriptomic Hourglass Patterns in Animal and Plant Embryogenesis. *Molecular biology and evolution 32*(5), 1221–31.

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development 1994*(Supplement), 135–142.

Duboule, D. and P. Dollé (1989). The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. *The EMBO journal 8*(5), 1497–505.

Endler, J. A. (1986). *Natural Selection in the Wild*. Monographs in population biology. Princeton University Press.

Ernst, J. and Z. Bar-Joseph (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC bioinformatics 7*, 191.

Eyre-Walker, A. and P. D. Keightley (2007). The distribution of fitness effects of new mutations. *Nature reviews. Genetics 8*(8), 610–8.

Eyre-Walker, A. and P. D. Keightley (2009). Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Molecular biology and evolution 26*(9), 2097–108.

Eyre-Walker, A., M. Woolfit, and T. Phelps (2006). The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics 173*(2), 891–900.

Forgacs, G. and S. A. Newman (2005). *Biological Physics of the Developing Embryo*. Cambridge University Press.

References

Frise, E., A. S. Hammonds, and S. E. Celniker (2010). Systematic image-driven analysis of the spatial Drosophila embryonic expression landscape. *Molecular systems biology 6*, 345.

Gall, J. G. and M. L. Pardue (1969). Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proceedings of the National Academy of Sciences of the United States of America 63*(2), 378–383.

García-Bellido, A., P. A. Lawrence, and G. Morata (1979). Compartments in Animal Development. *Scientific American 241*(1), 102–111.

Garcia-Bellido, A. and J. R. Merriam (1969). Cell lineage of the imaginal discs in Drosophila gynandromorphs. *Journal of Experimental Zoology 170*(1), 61–75.

García-Bellido, A., P. Ripoll, and G. Morata (1973). Developmental Compartmentalisation of the Wing Disk of Drosophila. *Nature 245*(147), 251–253.

Ghiselin, M. T. (2005). Homology as a relation of correspondence between parts of individuals. *Theory in biosciences = Theorie in den Biowissenschaften 124*(2), 91–103.

Gilbert, S. F. (1978). The embryological origins of the gene theory. *Journal of the History of Biology 11*(2), 307–351.

Gilbert, S. F. (Ed.) (1991). *A Conceptual History of Modern Embryology*. Boston, MA: Springer US.

Gilbert, S. F. (1998). Conceptual breakthroughs in developmental biology. *Journal of Biosciences 23*(3), 169–176.

Gilbert, S. F. (2000). Genes Classical and Developmental. In P. J. Beurton, R. Falk, and H.-J. Rheinberger (Eds.), *The Concept of the Gene in Development and Evolution*. Cambridge University Press.

Gilbert, S. F. (2007). Fate maps, gene expression maps, and the evidentiary structure of evolutionary developmental biology. In J. Maienschein and M. D. Laubichler (Eds.), *From Embryology to Evo-Devo : A History of Developmental Evolution.*, Dibner Institute Studies in the History of Science and Technology, pp. 357–374. The MIT Press.

Gilbert, S. F. (2011). Expanding the Temporal Dimensions of Developmental Biology: The Role of Environmental Agents in Establishing Adult-Onset Phenotypes. *Biological Theory 6*(1), 65–72.

Gilbert, S. F. (2014). *Developmental Biology* (10th ed.). Sinuaer Associates.

Gould, S. J. (1996). *Full House: The Spread of Excellence From Plato to Darwin*. New York: Harmony Books.

Graveley, B. R., A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin, L. Yang, C. G. Artieri, M. J. van Baren, N. Boley, B. W. Booth, J. B. Brown, L. Cherbas, C. a. Davis, A. Dobin, R. Li, W. Lin, J. H. Malone, N. R. Mattiuzzo, D. Miller, D. Sturgill, B. B. Tuch, C. Zaleski, D. Zhang, M. Blanchette, S. Dudoit, B. Eads, R. E. Green, A. Hammonds, L. Jiang, P. Kapranov, L. Langton, N. Perrimon, J. E. Sandler, K. H. Wan, A. Willingham, Y. Zhang, Y. Zou, J. Andrews, P. J. Bickel, S. E. Brenner, M. R. Brent, P. Cherbas, T. R. Gingeras, R. a. Hoskins, T. C. Kaufman, B. Oliver, and S. E. Celniker (2011). The developmental transcriptome of Drosophila melanogaster. *Nature 471*(7339), 473–9.

Griesemer, J. (2014). Reproduction and scaffolded developmental processes: an integrated evolutionary perspective. In *Towards a Theory of Development*, pp. 183–202. Oxford University Press.

Haeckel, E. (1874). *Anthropogenie oder Entwickelungsgeschichte des Menschen*. Engelmann, Leipzig.

Haeckel, E. (1880). *The history of creation*, Volume 1. New York: Appleton and Company.

Haeckel, E. (1903). *Anthropogenie: oder, Entwickelungsgeschichte des menschen* (5th ed.). Leipzig: W. Engelmann.

Haerty, W., S. Jagadeeshan, R. J. Kulathinal, A. Wong, K. Ravi Ram, L. K. Sirot, L. Levesque, C. G. Artieri, M. F. Wolfner, A. Civetta, and R. S. Singh (2007). Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. *Genetics 177*(3), 1321–35.

Hahn, M. W. and G. A. Wray (2002). The g-value paradox. *Evolution and Development 4*(2), 73–75.

Hall, B. K. (1999). *Evolutionary Developmental Biology*. Springer.

Hammonds, A. S., C. A. Bristow, W. W. Fisher, R. Weiszmann, S. Wu, V. Hartenstein, M. Kellis, B. Yu, E. Frise, and S. E. Celniker (2013). Spatial expression of transcription factors in Drosophila embryonic organ development. *Genome biology 14*(12), R140.

Hanelt, B., D. Van Schyndel, C. M. Adema, L. A. Lewis, and E. S. Loker (1996). The phylogenetic position of Rhopalura ophiocomae (Orthonectida) based on 18S ribosomal DNA sequence analysis. *Molecular biology and evolution 13*(9), 1187–91.

Hartenstein, V. (1993). *Atlas of Drosophila Development*. Cold Spring Harbor Laboratory Press.

Heemskerk, J., S. DiNardo, R. Kostriken, and P. H. O'Farrell (1991). Multiple modes of engrailed regulation in the progression towards cell fate determination. *Nature 352*(6334), 404–10.

Hershberg, R. and D. A. Petrov (2008). Selection on codon bias. *Annual review of genetics 42*, 287–99.

Horder, T. (2010). History of Developmental Biology. In *Encyclopedia of Life Sciences*. Chichester, UK: John Wiley & Sons, Ltd.

Horder, T. (2013). Heterochrony. In *eLS*. Chichester, UK: John Wiley & Sons, Ltd.

Horder, T. J. (2001). The organizer concept and modern embryology: Anglo-American perspectives. *The International journal of developmental biology 45*(1), 97–132.

Hotta, K., K. Mitsuhara, H. Takahashi, K. Inaba, K. Oka, T. Gojobori, and K. Ikeo (2007). A web-based interactive developmental table for the ascidian Ciona intestinalis, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. *Developmental dynamics : an official publication of the American Association of Anatomists 236*(7), 1790–805.

Hudson, C., S. Lotito, and H. Yasuo (2007). Sequential and combinatorial inputs from Nodal, Delta2/Notch and FGF/MEK/ERK signalling pathways establish a grid-like organisation of distinct cell identities in the ascidian neural plate. *Development 134*(19), 3527–3537.

Hughes, C. L. and T. C. Kaufman (2002). Hox genes and the evolution of the arthropod body plan. *Evolution & development 4*(6), 459–99.

Huxley, J. and G. De Beer (1963). *The elements of experimental embryology*. Cambridge comparative physiology. Hafner Pub. Co.

Imai, K. S. (2003). A Twist-like bHLH gene is a downstream factor of an endogenous FGF and determines mesenchymal fate in the ascidian embryos. *Development 130*(18), 4461–4472.

Imai, K. S., K. Hino, K. Yagi, N. Satoh, and Y. Satou (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development (Cambridge, England) 131*(16), 4047–58.

Imai, K. S., M. Levine, N. Satoh, and Y. Satou (2006). Regulatory blueprint for a chordate embryo. *Science (New York, N.Y.) 312*(5777), 1183–7.

Ingham, P. W., A. M. Taylor, and Y. Nakano (1991). Role of the Drosophila patched gene in positional signalling. *Nature 353*(6340), 184–7.

Jacob, F. (1973). *The Logic of Life*. New York: Pantheon Books.

Jacob, F. and J. Monod (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology 3*, 318–56.

Jaeger, J. (2011). The gap gene network. *Cellular and molecular life sciences : CMLS 68*(2), 243–74.

Jaeger, J., M. Blagov, D. Kosman, K. N. Kozlov, Manu, E. Myasnikova, S. Surkova, C. E. Vanario-Alonso, M. Samsonova, D. H. Sharp, and J. Reinitz (2004). Dynamical analysis of regulatory interactions in the gap gene system of Drosophila melanogaster. *Genetics 167*(4), 1721–37.

Jaeger, J. and J. Sharpe (2014). On the concept of mechanism in development. In A. Minelli and T. Pradeu (Eds.), *Towards a Theory of Development*, pp. 56–78. OUP Oxford.

Janning, W. (1978). Gynandromorph Fate Maps in Drosophila. In W. J. Gehring (Ed.), *Results and Problems in Cell Differentiation*, Chapter 1, pp. 1–28. Springer-Verlag Berlin Heidelberg.

Jernvall, J., S. V. Keränen, and I. Thesleff (2000). Evolutionary modification of development in mammalian teeth: quantifying gene expression patterns and topography. *Proceedings of the National Academy of Sciences of the United States of America 97*(26), 14444–8.

Kalinka, A. T. and P. Tomancak (2012). The evolution of early animal embryos: conservation or divergence? *Trends in ecology & evolution 27*(7), 385–93.

Kalinka, A. T., K. M. Varga, D. T. Gerrard, S. Preibisch, D. L. Corcoran, J. Jarrells, U. Ohler, C. M. Bergman, and P. Tomancak (2010). Gene expression divergence recapitulates the developmental hourglass model. *Nature 468*(7325), 811–4.

Keller, E. F. (2000). Decoding the Genetic Program: Or, Some Circular Logic in the Logic of Circularity. In P. J. Beurton, R. Falk, and H.-J. Rheinberger (Eds.), *The Concept of the Gene in Development and Evolution*. Cambridge University Press.

Kimchi-Sarfaty, C., J. M. Oh, I.-W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar, and M. M. Gottesman (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science (New York, N.Y.) 315*(5811), 525–8.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature 217*(5129), 624–6.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Klingler, M., M. Erdélyi, J. Szabad, and C. Nüsslein-Volhard (1988). Function of torso in determining the terminal anlagen of the Drosophila embryo. *Nature 335*(6187), 275–7.

Kumar, S., C. Konikoff, B. Van Emden, C. Busick, K. T. Davis, S. Ji, L.-W. Wu, H. Ramos, T. Brody, S. Panchanathan, J. Ye, T. L. Karr, K. Gerold, M. McCutchan, and S. J. Newfeld (2011). FlyExpress: visual mining of spatiotemporal patterns for genes and publications in Drosophila embryogenesis. *Bioinformatics (Oxford, England) 27*(23), 3319–20.

Lamarck, J. (1809). *Zoological philosophy*. Univ. of Chicago Press, Chicago.

Lawrence, P. A. (1992). *The Making of a Fly: The Genetics of Animal Design*. Wiley.

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

Lemaire, P., W. C. Smith, and H. Nishida (2008).

Ascidians and the plasticity of the chordate developmental program. *Current biology : CB 18*(14), R620–31.

Levin, M., T. Hashimshony, F. Wagner, and I. Yanai (2012). Developmental Milestones Punctuate Gene Expression in the Caenorhabditis Embryo. *Developmental Cell 22*(5), 1101–1108.

Lewin, R. (1981). Seeds of change in embryonic development. *Science 214*(4516), 42–44.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature 276*(5688), 565–70.

Love, A. C. (2014). *Conceptual Change and Evolutionary Developmental Biology*. Boston Studies in the Philosophy and History of Science. Springer Netherlands.

Maienschein, J. (1991). The Origins of Entwicklungsmechanik. In S. F. Gilbert (Ed.), *A Conceptual History of Modern Embryology*, pp. 43–61. Boston, MA: Springer US.

Marianes, A. and A. C. Spradling (2013). Physiological and stem cell compartmentalization within the Drosophila midgut. *eLife 2*, e00886.

Matsuoka, T., T. Ikeda, K. Fujimaki, and Y. Satou (2013). Transcriptome dynamics in early embryos of the ascidian, Ciona intestinalis. *Developmental biology 384*(2), 375–85.

Maynard Smith, J., R. Burian, S. Kauffman, P. Alberch, J. Campbell, B. Goodwin, R. Lande, D. Raup, and L. Wolpert (1985). Developmental Constraints and Evolution: A Perspective from the Mountain Lake Conference on Development and Evolution. *The Quarterly Review of Biology 60*(3), 265–287.

Mayr, E. (1961). Cause and effect in biology. *Science (New York, N.Y.) 134*(3489), 1501–6.

Mayr, E. (1966). *Animal Species and Evolution*. Belknap Press of Harvard University Press.

Mayr, E. (1993). *One Long Argument: Charles Darwin and the Genesis of Modern Evolutionary Thought (Questions of Science)*. Harvard University Press.

Mayr, E. (1997). *Evolution and the Diversity of Life: Selected Essays*. Selected Essays. Belknap Press of Harvard University Press.

McDonald, J. H. and M. Kreitman (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature 351*(6328), 652–4.

McGinnis, W., M. S. Levine, E. Hafen, A. Kuroiwa, and W. J. Gehring (1984). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature 308*(5958), 428–433.

McShea, D. W. (1996). Perspective: Metazoan Complexity and Evolution: Is There a Trend? *Evolution 50*(2), 477.

McShea, D. W. (2015). Three Trends in the History of Life: An Evolutionary Syndrome. *Evolutionary Biology*.

Medawar, P. (1954). The significance of inductive relationships in the development of vertebrates. *Journal of Embryology and Experimental ... 2*(June), 172–174.

Mensch, J., F. Serra, N. J. Lavagnino, H. Dopazo, and E. Hasson (2013). Positive selection in nucleoporins challenges constraints on early expressed genes in Drosophila development. *Genome biology and evolution 5*(11), 2231–41.

Messer, P. W. and D. A. Petrov (2013). Frequent adaptation and the McDonald-Kreitman test. *Proceedings of the National Academy of Sciences of the United States of America 110*(21), 8615–20.

Minelli, A. (2011). Animal Development, an Open-Ended Segment of Life. *Biological Theory 6*(1), 4–15.

Minelli, A. (2014). Developmental disparity. In *Towards a Theory of Development*, pp. 227–245. Oxford University Press.

Mitani, Y., H. Takahashi, and N. Satoh (1999). An ascidian T-box gene As-T2 is related to the Tbx6 subfamily and is associated with embryonic muscle cell differentiation. *Developmental Dynamics 215*(1), 62–68.

Moczek, A. P. (2014). Towards a theory of development through a theory of developmental evolution. In *Towards a Theory of Development*, pp. 218–226. Oxford University Press.

Monod, J. (1963). Genetic Repression, Allosteric Inhibition, and Cellular Differentiation. In M. Locke (Ed.), *Cytodifferentiation and Macromolecular Synthesis*, pp. 30–64. New York: Academic Press.

Morgan, T. H. (1910). Chromosomes and Heredity. *The American naturalist* (44), 449–496.

Morgan, T. H. (1919). The physical basis of heredity.

Morgan, T. H. (1926). *The theory of the gene*. Yale University Press.

MUKAI, T. (1964). THE GENETIC STRUCTURE OF NATURAL POPULATIONS OF DROSOPHILA MELANOGASTER. I. SPONTANEOUS MUTATION RATE OF POLYGENES CONTROLLING VIABILITY. *Genetics 50*, 1–19.

Müller, G. B. (2007). Evo-devo: extending the evolutionary synthesis. *Nature reviews. Genetics 8*(12), 943–9.

Müller, G. B. and S. A. Newman (1999). Generation, integration, autonomy: three steps in the evolution of homology. *Novartis Foundation symposium 222*, 65–73; discussion 73–9.

Nakamura, M. J., J. Terai, R. Okubo, K. Hotta, and K. Oka (2012). Three-dimensional anatomy of the Ciona intestinalis tailbud embryo at single-cell resolution. *Developmental biology 372*(2), 274–84.

Needham, J. (1959). *A history of embryology*. New York: Abelard-Schuman.

Neuman-Silberberg, F. S. and T. Schüpbach (1993). The Drosophila dorsoventral patterning gene gurken produces a dorsally localized RNA and encodes a TGF alpha-like protein. *Cell 75*(1), 165–74.

Nicol, D. and I. A. Meinertzhagen (1988). Development of the central nervous system of the larva of the ascidian, Ciona intestinalis L. II. Neural plate morphogenesis and cell lineages during neurulation. *Developmental biology 130*(2), 737–66.

Nishida, H. (1987). Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. *Developmental biology 121*(2), 526–41.

Nishida, H. (2005). Specification of embryonic axis and mosaic development in ascidians. *Developmental dynamics : an official publication of the American Association of Anatomists 233*(4), 1177–93.

Nüsslein-Volhard, C. and E. Wieschaus (1980). Mutations affecting segment number and polarity in Drosophila. *Nature 287*(5785), 795–801.

Nuzhdin, S. V., M. L. Wayne, K. L. Harmon, and L. M. McIntyre (2004). Common pattern of evolution of gene expression level and protein sequence in Drosophila. *Molecular biology and evolution 21*(7), 1308–17.

Ostrowski, S., H. A. Dierick, and A. Bejsovec (2002). Genetic Control of Cuticle Formation During Embryonic Development of Drosophila melanogaster. *Genetics 161*(1), 171–182.

Oyama, S. (2000). *The Ontogeny of Information: Developmental Systems and Evolution*. Science and Cultural Theory. Duke University Press.

Papaioannou, V. E. (2014). The T-box gene family: emerging roles in development, stem cells and cancer. *Development (Cambridge, England) 141*(20), 3819–33.

Peregrín-Alvarez, J. M., C. Sanford, and J. Parkinson (2009). The conservation and evolutionary modularity of metabolism. *Genome biology 10*(6), R63.

Piasecka, B., P. Lichocki, S. Moretti, S. Bergmann, and M. Robinson-Rechavi (2013). The hourglass and the early conservation models–co-existing patterns of developmental constraints in vertebrates. *PLoS genetics 9*(4), e1003476.

Plotkin, J. B. and G. Kudla (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nature reviews. Genetics 12*(1), 32–42.

Poe, S. and M. H. Wake (2004). Quantitative tests of general models for the evolution of development. *The American naturalist 164*(3), 415–22.

Pradeu, T. (2014). Regenerating theories in developmental biology. In A. Minelli and T. Pradeu (Eds.), *Towards a Theory of Development*, pp. 15–32. Oxford University Press.

Presnyak, V., N. Alhusaini, Y.-H. Chen, S. Martin, N. Morris, N. Kline, S. Olson, D. Weinberg, K. Baker, B. Graveley, and J. Coller (2015). Codon Optimality Is a Major Determinant of mRNA Stability. *Cell 160*(6), 1111–1124.

Pröschel, M., Z. Zhang, and J. Parsch (2006). Widespread adaptive evolution of Drosophila genes with sex-biased expression. *Genetics 174*(2), 893–900.

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

Richardson, M. K. (1995). Heterochrony and the phylotypic period. *Developmental biology 172*(2), 412–21.

Richardson, M. K., J. Hanken, M. L. Gooneratne, C. Pieau, a. Raynaud, L. Selwood, and G. M. Wright (1997). There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development. *Anatomy and embryology 196*(2), 91–106.

Richardson, M. K. and G. Keuck (2002). Haeckel's ABC of evolution and development. *Biological Reviews of the Cambridge Philosophical Society 77*(4), 495 – 528.

Roux, W. (1888). Beiträge zur Entwickelungsmechanik des Embryo. Über die künstliche Hervorbringung halber Embryonen durch Zerstörung einer der beiden ersten Furchungskugeln, sowie über die Nachentwickelung (Postgeneration) der fehlenden Körperhälfte. *Virchows Arch. pathol. Anat. Physiol. klin. Med. 114*(1), 113–153.

Roux, W. (1897). *Programm und Forschungsmethoden der Entwickelungsmechanik der Organismen*. Leipzig: W. Engelmann.

Russell, E. S. (1916). *Form and function : a contribution to the history of animal morphology*. London: J. Murray.

Salazar-Ciudad, I. (2006). Developmental constraints vs. variational properties: how pattern formation can help to understand evolution and development. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 306B*(2), 107–125.

Salazar-Ciudad, I. (2009). Looking at the origin of phenotypic variation from pattern formation gene networks. *Journal of biosciences 34*(4), 573–87.

Salazar-Ciudad, I. (2010). Morphological evolution and embryonic developmental diversity in metazoa. *Development (Cambridge, England) 137*(4), 531–9.

Salazar-Ciudad, I. and J. Jernvall (2004). How different types of pattern formation mechanisms affect the evolution of form and development. *Evolution & development 6*(1), 6–16.

Salvador-Martínez, I. and I. Salazar-Ciudad (2015). How complexity increases in development: An analysis of the spatial-

temporal dynamics of 1218 genes in Drosophila melanogaster. *Developmental Biology 405*(2), 328–339.

Sander, K. (1983). The evolution of patterning mechanisms: gleanings from insect embryogenesis and spermatogenesis. In B. C. Goodwin, N. Holder, and C. C. Wylie (Eds.), *Development and Evolution*. Cambridge, UK: Cambridge University Press.

Sander, K. (1996). Pattern formation in insect embryogenesis: The evolution of concepts and mechanisms. *International Journal of Insect Morphology and Embryology 25*(4), 349–367.

Sanjuán, R., A. Moya, and S. F. Elena (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proceedings of the National Academy of Sciences of the United States of America 101*(22), 8396–401.

Schep, A. N. and B. Adryan (2013). A comparative analysis of transcription factor expression during metazoan embryonic development. *PloS one 8*(6), e66826.

Sebé-Pedrós, A., A. Ariza-Cosano, M. T. Weirauch, S. Leininger, A. Yang, G. Torruella, M. Adamski, M. Adamska, T. R. Hughes, J. L. Gómez-Skarmeta, and I. Ruiz-Trillo (2013). Early evolution of the T-box transcription factor family. *Proceedings of the National Academy of Sciences of the United States of America 110*(40), 16050–5.

Sempere, L. F., C. N. Cole, M. A. McPeek, and K. J. Peterson (2006). The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *Journal of experimental zoology. Part B, Molecular and developmental evolution 306*(6), 575–88.

Sharp, P. M. (1991). Determinants of DNA sequence divergence between Escherichia coli and Salmonella typhimurium: codon usage, map position, and concerted evolution. *Journal of molecular evolution 33*(1), 23–33.

Showell, C., O. Binder, and F. L. Conlon (2004). T-box genes in early embryogenesis. *Developmental dynamics : an official publication of the American Association of Anatomists 229*(1), 201–18.

Shubin, N., C. Tabin, and S. Carroll (1997). Fossils, genes and the evolution of animal limbs. *Nature 388*(6643), 639–48.

Shubin, N., C. Tabin, and S. Carroll (2009). Deep homology and the origins of evolutionary novelty. *Nature 457*(7231), 818–23.

Simpson-Brose, M., J. Treisman, and C. Desplan (1994). Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. *Cell 78*(5), 855–65.

Slack, J. M., P. W. Holland, and C. F. Graham (1993). The zootype and the phylotypic stage. *Nature 361*(6412), 490–2.

Slack, J. M. W. (2002). Conrad Hal Waddington: the last Renaissance biologist? *Nature reviews. Genetics 3*(11), 889–95.

Slack, J. M. W. (2012). *Egg & Ego: An Almost True Story of Life in the Biology Lab*. Springer New York.

Small, S., R. Kraut, T. Hoey, R. Warrior, and M. Levine (1991). Transcriptional regulation of a pair-rule stripe in Drosophila. *Genes & development 5*(5), 827–39.

Spemann, H. and H. Mangold (1924). über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren. *Archiv für Mikroskopische Anatomie und Entwicklungsmechanik 100*(3-4), 599–638.

Stanojevic, D., S. Small, and M. Levine (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science (New York, N.Y.) 254*(5036), 1385–7.

Tautz, D. (1988). Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres. *Nature 332*(6161), 281–4.

Tokuoka, M., K. S. Imai, Y. Satou, and N. Satoh (2004). Three distinct lineages of mesenchymal cells in Ciona intestinalis embryos demonstrated by specific gene expression. *Developmental biology 274*(1), 211–24.

Tomancak, P., A. Beaton, R. Weiszmann, E. Kwan, S. Shu, S. Lewis, S. Richards, M. Ashburner, V. Hartenstein, S. Celniker, and G. Rubin (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. *Genome Biology 3*(12), 1–14.

Tomancak, P., B. P. Berman, A. Beaton, R. Weiszmann, E. Kwan, V. Hartenstein, S. E. Celniker, and G. M. Rubin (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. *Genome biology 8*(7), R145.

True, J. R. and E. S. Haag (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evolution and Development 3*(2), 109–119.

Valentine, J. W., A. G. Collins, and C. P. Meyer (1994). Morphological Complexity Increase in Metazoans. *Paleobiology 20*(2), 131–142.

von Baer, K. E. (1828). *Über Entwickelungsgeschichte der Thiere. Beobachtung und Reflexion*. Königsberg.

Waddington, C. H. (1962). *How animals develop*. New York: Harper Torchbooks.

Waddington, C. H., J. Needham, W. W. Nowinski, and R. Lemberg (1935). Studies on the Nature of the Amphibian Organization Centre. I.–Chemical Properties of the Evocator. *Proceedings of the Royal Society B: Biological Sciences 117*(804), 289–310.

Weismann, A. (1893). *The germ-plasm. A Theory of Heredity*. New York: Charles Scribner's Sons.

Weiszmann, R., A. S. Hammonds, and S. E. Celniker (2009). Determination of gene expression patterns using high-throughput RNA in situ hybridization to whole-mount Drosophila embryos. *Nature protocols 4*(5), 605–18.

Wray, G. A. (2000). The evolution of embryonic patterning mechanisms in animals. *Seminars in cell & developmental biology 11*(6), 385–93.

Yasuo, H. and N. Satoh (1998). Conservation of the developmental role of Brachyury in notochord formation in a urochordate, the ascidian Balocynthia roretzi. *Developmental biology 200*(2), 158–70.