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

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.

University of Helsinki
Finland

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

Copyright © 2016 Irepan Salvador-Martínez
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
N_e	Effective population size
s	Selection coefficient
A/P	anterior/posterior
Adh	Alcohol dehydrogenase
BDGP	Berkeley Drosophila Genome Project
BMP4	Bone morphogenetic protein 4
CaM	Calmoduline
D/V	dorsal/ventral
DFE	Distribution of Fitness Effects
DGRP	Drosophila Genome Reference Panel
DNA	Deoxyribonucleic acid
Evo-devo	Evolutionary Developmental Biology
GRN	Gene Regulatory Network
miRNAs	microRNAs
MKT	MacDonald-Kreitman test
RNA	Ribonucleic acid
SFS	Site Frequency Spectrum
SNP	Single nucleotide polymorphism
TF	Transcription Factor

1 Abstract

The fact that complexity

2 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 (Muller, 2006). However, these fields have not always gotten along well. Some decades ago, there was a (CLEAR? CONCEPTUAL, 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). But before that, it might be useful to define what development is, so firstly, I will address this apparently simple question.

2.1 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 sin-

gle 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.

2.2 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)

2.2.1 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.

2.2.2 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 as the main accepted idea in the 18th century, but some scientists 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, its germ would be smaller than the smallest possible atom (Buffon, 1807).

2.2.3 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 view was not an evolutionary one. 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 classified two types of change in development introducing "heterochrony" and "heterotopy", concepts 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. They only apply to some characters, stages and levels of phylogenetic inclusiveness (Richardson and Keuck, 2002). 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).

2.2.4 *Entwicklungsmechanik*

Despite the great advances described above, embryology remained a descriptive science most of the 19th century. 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 famous 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, when Weissman proposed his theory, 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).

Even when Roux's hypothesis was proved wrong by Driesch just a few years after being proposed, his work laid the foundations of a new scientific programme whose main purpose development mechanism was to "research the causes, on which the formation, maintenance and regression of the organic forms are based" (Roux, 1897). Most importantly, he demonstrated that the problem of development was tractable and that hypotheses could be experimentally tested.

2.2.5 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 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).

> *"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. Conrad H. Waddington (a leading embryologist and geneticist mostly know for his 'epigenetic landscape' and 'genetic assimilation' concepts; Slack, 2002) and many other researchers around the world tried to characterize the chemical nature of the organizer (Waddington et al., 1935; Gilbert, 1991). However, embryologists did not succeed in its characterization and by the end of the 1930's "the sense of disappointment and

disillusionment was manifest" (Horder, 2010). This lack of success led to the gradual lost of interest in the organizer problem (REF Holftreter in gilbert 1999)

2.2.6 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, which was not accepted by many embryologists. 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, 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 1926, in his book "Theory of the Gene", Morgan declare 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, that stated that the study of genetics was completely different from that of development, combined in the 1940's with population genetics and other fields to form the Evolutionary Synthesis. As it was considered irrelevant to the study of heredity, the entire scientific field of development was excluded from the Evolutionary Synthesis (Amundson, 2005).

– calls for reconciliation (waddington urges for integration of genetics and embryology)

2.2.7 Developmental genetics

The second major approach in development was the genetic approach (examples)

– finally the discovery of the organizer molecules!

– HOX genes!

–

These discoveries prompted discussions on the role of development on the evolutionary theory.

The main idea was that all evolutionary change is produced by a change in development (Alberch).

Evolutionary developmental biology (evo-devo) was born.

2.2.8 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".

2.3 Complexity

The notion of an increase in complexity as an evolutionary trend has been for long part of the evolutionary thought. Advocates to this idea have used many arguments to support it. For example, adaptive reasons have been suggested, so that the increase in complexity should have been driven by natural selection (Bonner, 1988; Carroll et al., 2001).

There are however some complications to accept the existence of this trend. In the first place, before accepting the existence of such trend, we should define complexity. More specifically, we should be able to measure the complexity of an organism in order to compare it to another one. Furthermore, even if we find evidence of an increase of complexity in particular lineages, it would not mean that it is generalized trend (for a great review on this topic see (McShea, 1996))

2.3.1 Defining complexity

A general definition could be "the number of component parts" of an organism (McShea, 1996; Arthur, 2010). These "parts" might be body segments (e.g., of an insect) or genes. It is doubtful to say that some centipede is more complex than a beetle, just based in the different number of segments they have. Also, it is already acknowledge that there is no relation between the number of coding-genes and morphological complexity. This lack of correspondence, sometimes referred as the "G-value paradox" (Hahn and Wray, 2002), became evident with the release of the first eukaryotic genome sequences. Decades before, 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. Again this definition is disputable, as it is acknowledged that during evolution gene-gene interactions (or gene regulatory network) underlying a phenotype can increase their complexity without affecting the phenotype itself (Müller and Newman, 1999; True and Haag, 2001; Salazar-Ciudad, 2009).

A measure of morphological complexity that has been favoured by some authors (perhaps because of its intuitiveness), is the number of cell types that compose an organism (Bell and Mooers, 1997; Bonner, 2004; McShea, 1996).

2.3.2 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. Animals of the group formerly known as the "Mesozoa" are worm-like parasites of marine invertebrates. Because of their simple morphology, 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 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, now is clear 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 2.1a). However, if we could depict the change in complexity in all lineages (see Figure 2.1b), we probably would see that

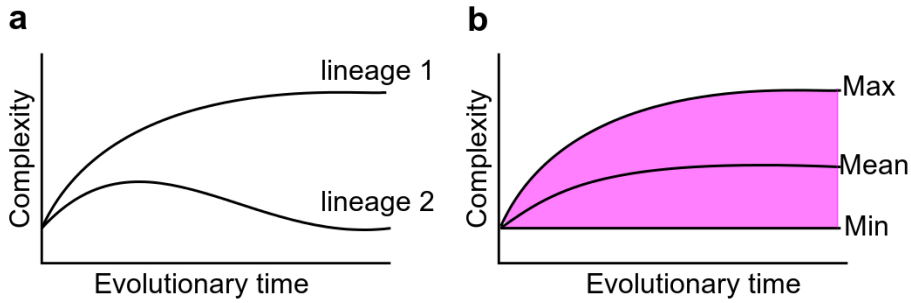

Figure 2.1: a) Two lineages with different complexity change through their evolutionary trajectories. b) Representation of the minimum, mean and maximum complexity of many lineages over evolutionary time in which the minimum stay constant while the mean and maximum increase. Redrawn from (Arthur, 2010).

the range of complexity has increased over time, with the a lower limit or minimum complexity that has stayed the same and with an increase in mean and maximun (or upper limit) of complexity (McShea, 1996; Arthur, 2010).

2.3.3 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 again as a proxy of morphological complexity, it can be said that during metazoan development, complexity increases as the zygote divides and differentiates into an adult with multiple cell types. This simple definition of complexity has its 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. For example, it could be that at the morphological level a cell seems to be undifferentiated, but when isolating it from its neighbour cells, it differentiates in an autonomous way into a specific cell type, suggesting that the cell fate was already determined without the necessity of further interactions with other cells.

Talk about differentiation and determination?

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.

Defining a pattern as a specific distribution of cell types in a specific temporal window of embryonic development (Salazar-Ciudad and Jernvall, 2004), development can be conceptualized as the continuous transformation of one pattern into another. The earliest pattern transformations usually establish the main axis or "compartments" of the embryo. For example, the anterior/posterior and dorsal/ventral axes in the fruit fly. Later pattern transformations define smaller compartments of the embryo, e.g, limbs, fingers or internal organs. As development proceeds, spatial compartments are progressively specified at an increasing finer resolution (Davidson, 2001). Thus, a great part of pattern transformation is the partition of specific embryo compartments into smaller sub-compartments.

The compartmentalization of the embryo can be considered then an intrinsic property of development, and, as it is going to be mention in later sections, can be used as a proxy of complexity.

2.3.4 Complexity at the molecular level

At the molecular level, the increasing compartmentalization of the embryo during development can be seen as the progressive spatial restriction of gene expression to subsequently smaller regions in the embryo. Sean Carroll conceptualizes 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).

If we consider again the number of cells as the complexity measure, we would expect 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).

For some, the combinatorial approach could also be seen as a solution for the "G-value paradox", as what really matters to be complex would not be how many genes an organism has, but how would these genes are differentially combined to produce more cell types in it. Another proposed solution is "gene co-option" in which a gene, usually after a duplication, evolves a new "function" different from its original one (REF Carrol 2003).

The differential gene expression in the various number of cell types are determined in great manner by the interplay of 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 sometimes referred as gene regulatory networks (GRNs) or "regulatory architecture" of the genome (Davidson, 2001). This approach to complexity relates to the "interaction among parts" definition of complexity mentioned at the beginning of this section.

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).

Other authors highlight the importance of modules in facilitating the evolution of complex forms (Carroll, 2001).

Different types of developmental genes

After acknowledging the importance of the spatio-temporal regulation of gene expression in development, it is useful to distinguish the type of genes

that are directly involved in it. 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 2.2). Due to the importance of both transcription factors as signalling pathway genes in cell differentiation, a brief description of each follows.

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

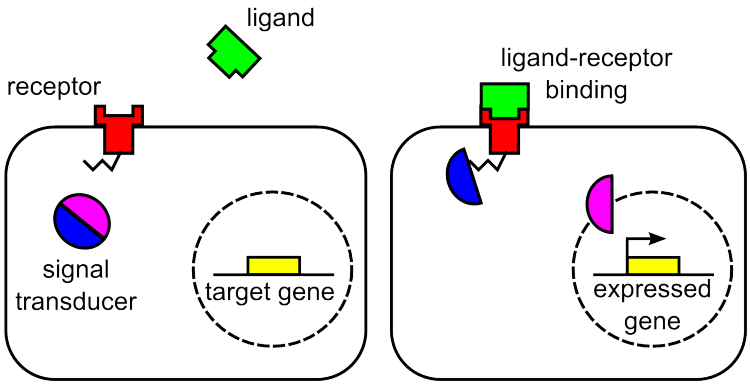

Figure 2.2: Scheme of an hypothetical signalling pathway. Left: The extracellular ligand (green) is not bound to the membrane receptor (red) so the signal transducer protein (blue/magenta) is inactivated. In the nucleus (dashed circle) the target gene (yellow box) is inactive. Right: As the ligand binds 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.

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).

2.3.5 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.

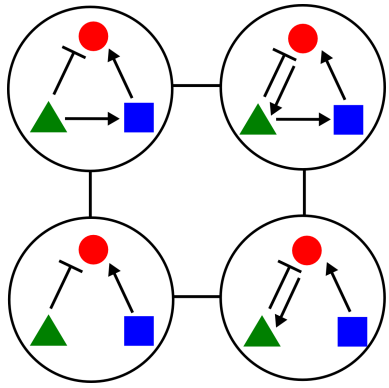

Figure 2.3: Some GRNs . Cotterel and Harpe calculated... (Cotterell and Sharpe, 2010)

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).

2.3.6 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 2.3). 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.

Make a figure?

2.4 Adaptation

Measuring adaptation is an important topic in evolutionary biology.

Since Darwin.. (REF Darwin)

In here, I refer to adaptation as a phenotypic character (or modification in a phenotypic character) that arise by natural selection in response to the environment, or other external factor. Measuring adaptation at the phenotypic level requires a clear understanding of the function of the phenotypic character under study, and how the modification of this trait would affect the fitness of the bearer organism (Emília Santos et al., 2015).

Importantly, all phenotypic changes (whether a new character or a modification of an existing character) 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. A proposed model for 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.

Some other remarkable examples of the genetic-developmental basis of adaptive change are:

-

-

It is evident at this point than many of the developmental changes leading to an adaptation are (at least partially) caused by mutations in gene regulatory or coding sequences. Therefore, 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.

2.4.1 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).

2.4.2 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.

2.4.3 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.

2.4.4 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.

2.4.5 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 2.4).

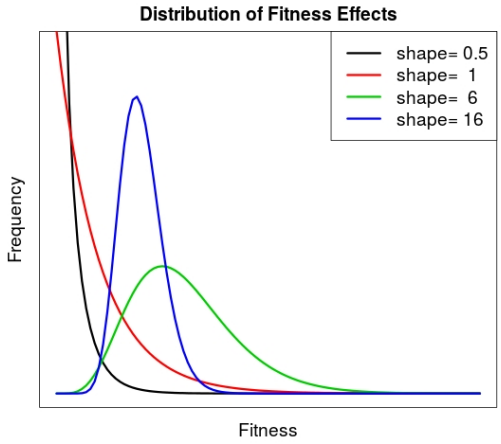

Figure 2.4: Example of different Distribution of Fitness Effects (DFE) represented by a gamma distribution. Many distributions can be represented by modifying the shape parameter of a gamma distribution, from a leptokurtic (shape parameter less than 1) to an exponential (shape parameter equal to 1) or a skewed normal distribution (shape greater than 1).

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 In Drosophila genes, we estimate that approximately 50% of amino acid substitutions and approximately 20% of substitutions in introns are adaptive (Eyre-Walker and Keightley, 2009).

Molecular evolution simulations have been performed 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 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). .

2.5 Drosophila

2.5.1 Subsection one

Subsubsection one

2.6 Ciona

2.6.1 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 (REF, Levine).

2.6.2 Current knowledge about Ciona development

The sequencing of the *C. intestinalis* genome (Dehal, 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, 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 (REF). 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, Nishida 1987), the cDNA spatial expression have been described at the single cell level up to the early gastrula stage (REF), making this database an invaluable resource to investigate the spatio-temporal dynamics of gene expression.

Start with some text.

Figure 2.5 shows an example on how figures in the png format can be

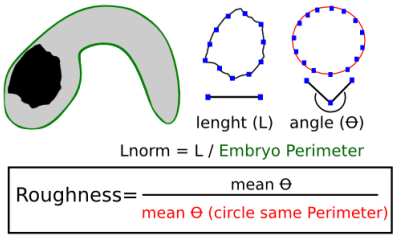

Figure 2.5: This is a figure caption

included into the text.

3 Aims of the study

4 Material and Methods

5 Results and Discussion

6 Concluding Remarks

List of Figures

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.

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.

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

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

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

Bonner, J. T. (1988). *The Evolution of Complexity by Means of Natural Selection*. Princeton University Press.

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.

References

Carroll, S. B. (2001). Chance and necessity: the evolution of morphological complexity and diversity. *Nature 409*(6823), 1102–9.

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.

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.

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.

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.

Emília Santos, M., C. S. Berger, P. N. Refki, and A. Khila (2015). Integrating evo-devo with ecology for a better understanding of phenotypic evolution. *Briefings in functional genomics 14*(6), 384–395.

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*. publisherNameCambridge University Press.

References

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*. Developmental Biology: A Comprehensive Synthesis. Boston, MA: Springer US.

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

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.

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.

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.

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.

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

References

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.

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

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. 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.

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.

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.

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

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.

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.

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

References

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.

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

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

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

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.

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.

References

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

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.

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.

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.

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.

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.

References

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

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

Waddington, C. H. (1962). How animals develop.

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.