1 What is Evo-Devo and Why is it Important?

What is Evo-Devo?

The two great creative processes of biology are evolution and development. You and I, as adult human beings, are products of both. Evolution took about four billion years to make the first human from a unicellular organism that emerged from the primordial soup. Development, in the form of embryogenesis together with its post-embryonic counterpart, takes less than 20 years to produce an adult human from a different unicellular organism – a fertilized egg or zygote. By this measure, development operates more than 200 million times faster than evolution. However, despite their very different timescales, the two great creative processes of biology are intrinsically interwoven. Evo-devo is the scientific study of this interweaving. Its full name is evolutionary developmental biology, but because this is an unwieldy phrase it is almost universally referred to by its nickname.

Fundamental to any field of science is the search for general, rather than piecemeal, explanations. However, we can only generalize as far as is consistent with the facts at hand. Biology is less fertile scientific ground for generalizations than physics, because there are nearly always exceptions to any proposed general rule (for example, there are exceptions to Mendel's 'laws' of inheritance). The solution to this problem for biologists is not – of course – to abandon the quest for general explanations, but rather to recognize how far any proposed generalization can go, and where its limits are set.

Against this background, we should consider the scope of evo-devo, and of any proposed general explanations that emerge from this relatively new scientific discipline. I said in the opening paragraph that evolution and

development are 'intrinsically interwoven'. However, while this is true for parts of the living world, it is not true of others. If we restrict 'development' to multicellular organisms in which changes occur in populations of self-adhering cells, such as embryos, larvae, or juveniles, then life forms that are unicellular throughout their entire life cycle do not have development as such. For example, a bacterium that lives for a certain period as a single metabolizing cell and then splits into two identical daughter cells by asexual reproduction cannot be said to have a developmental phase in its life cycle. In contrast, a human, a dolphin and a butterfly most certainly do have a developmental phase – indeed the butterfly has three of them (embryogenesis, larval growth, and metamorphosis).

Although I have contrasted bacteria with animals to make this point, the difference between the realms of life that are characterized by (a) occurrence of development and (b) absence of development is more complex than this introduction to the subject suggests. The realm to which development (and hence also the evolution of development) applies is that of multicellular organisms – or, to put it more precisely, the realm of organisms that go through at least one multicellular phase in their life cycles. This means that evo-devo concerns itself not just with animals but also with plants, and with some members of other kingdoms – for example the fungal and brown-algal kingdoms. It also potentially deals with certain 'microbes' (an undefined but generally useful term) – the ones that have a phase in their life cycle that is multicellular, albeit relatively simple, such as those cyanobacteria (previously called blue-green algae) that can form filaments or mats of attached cells.

At this point it becomes helpful to have some idea of the broad-scale structure of the living world, in terms of its hierarchical division into its three major groups (called domains) and the major subgroups within these (called kingdoms). Our view of broad-scale structure has changed considerably over the last few centuries. It will continue to change in the future, but probably much less than in the past, assuming that we are gradually homing in on a correct understanding of the course that evolution has taken. Figure 1.1 shows the broad structure of the living world, as currently perceived. Development and evolution of development characterize one of the eukaryote kingdoms in its entirety (animals), most of another (plants, defined to include both green algae and land plants), and parts of at least two others (fungi, which includes unicellular yeasts as well as multicellular toadstools, and what I think of as

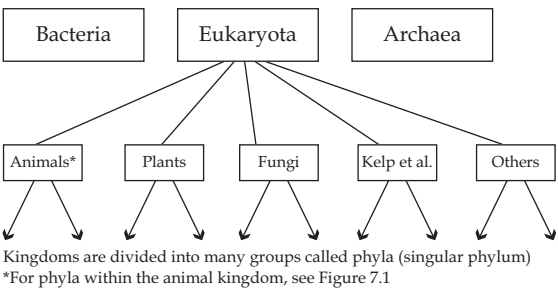

Figure 1.1 The broad structure of the living world: three domains, each divided into kingdoms. Here, kingdoms are only shown for the eukaryote domain. Evo-devo applies to animals, to most plants, and to some members of other kingdoms.

'the kelp kingdom', which includes unicellular diatoms as well as multicellular brown algae). Other eukaryote kingdoms, and the domains of Bacteria and Archaea, are not entirely without development, but its occurrence is very patchy, and evo-devo to date has largely omitted consideration of members of these groups.

Evo-devo began in the early 1980s with studies on animals, and consequently the evo-devo of animals is better known than that of other relevant kingdoms, with the evo-devo of plants coming second. Partly because of this, and partly because I am a zoologist and know the animal kingdom better than I know the others, animal examples will predominate in this book. I hope the reader will forgive this bias, and in mitigation I can at least say that the case studies discussed will range widely across both vertebrates and invertebrates.

Although evo-devo is inapplicable to some life forms on Earth, it may well in the future turn out to be applicable to many life forms elsewhere. At the outset of his 1992 book entitled *Natural Selection*, the American biologist George C. Williams states a philosophical position: he believes that natural selection will be seen to characterize all life forms in however many biospheres exist in the universe – probably a very large number. Similarly, I will state a philosophical position here: evo-devo will be seen to be relevant to all life forms everywhere that are multicellular in their construction. This is a slightly different sort of statement, since natural selection is a process while evo-devo is a field of study. However, I would predict that the most important

processes involved in the evolution of development here, as discussed in later chapters, will be relevant on other planets too.

We've now considered the realm within which evo-devo studies are meaningful. Or, in other words, we've clarified the realm in which evolution and development are interwoven. Having done that, we now turn to the *way* in which they are interwoven. From here on I will focus on the animal kingdom, unless specified otherwise. However, many of the points made will be equally applicable to plants, and to other multicellular organisms.

The development of any animal can be thought of as a trajectory from zygote to adult. Since I will be using 'developmental trajectory' often in this book, I should explain here what I'm thinking about when I use this term. Imagine the development of a human. Each of us starts our life as a single cell, which starts to multiply, producing a self-adhering cluster or ball of cells. As it continues to grow, this cluster begins to take a more definite shape, with elongation in one direction producing the anteroposterior (or head-to-tail) axis of the body. Internal changes, such as the separating out of different tissue layers, accompany the external changes in shape, and the overall growth. The embryo gradually elaborates its features, becoming more and more like a miniature human. After birth, development continues, but is more subtle. One important aspect of the post-embryonic development of a human is differential growth rates of different parts of the body, something that is referred to as allometric growth (to distinguish it from isometric growth, where different body parts grow at the same rate). For example, our heads grow more slowly, after birth, than our trunks and limbs. The combination of all these changes leading from zygote to adult constitutes the developmental trajectory of a human.

At any moment in evolutionary time – say halfway through the Jurassic period – the developmental trajectory found in a certain kind of animal – say a particular species of dinosaur – has been produced by the accumulated evolution of the past and is the starting point for the evolution of the lineage concerned in the future. Development is a quasi-predictable process. Its many repeat occurrences within a given species produce variants that are typically rather similar to each other – though not identical. In contrast to development, evolution is a very *un*predictable process. The fact that one dinosaur developmental trajectory gave rise to all of today's 10,000 species of birds while

the others left no descendants among today's fauna could not have been foretold. Evolution incorporates a major element of 'historical contingency' – chance events including asteroid impacts and volcanic eruptions – as emphasized by the American palaeontologist Stephen Jay Gould. Development is usually much less affected by such contingency.

One way of looking at the intertwining of evolution and development, then, is that development is a sort of raw material that gets moulded by natural selection in ways that adapt it to the prevailing environmental conditions in the relevant habitat. For example, most frogs, including all those living in temperate habitats, have a process of indirect development that includes a tadpole stage, but some species living in warm, moist, tropical conditions have dispensed with the tadpole stage of this ancestral life cycle and have evolved a process of 'direct development' (the embryo gives rise directly to a juvenile that's like a small adult). Among the many species of frogs living in temperate regions, evolution has again moulded the pattern of development to fit the environment, but in less dramatic ways. For example, a shorter tadpole phase of the life cycle would be expected to be favoured by selection in regions where the water bodies inhabited by the tadpoles are more transient than they are in others.

But the interweaving of evolution and development is not a one-way street. Evolution alters the developmental process, to be sure. But the evolutionary process is also altered by development. Or, to put it another way, evolution is effectively channelled in terms of what it can do with a particular lineage in the future by the prevailing developmental trajectory of the species concerned in the present. Likewise, evolution at any point in the past – say the mid-Jurassic again – was channelled by the developmental trajectories that were available at that time. This channelling is often referred to as 'developmental constraint'. However, I think this phrase has an overly negative tone. If development in some sense channels the direction of evolution, then it can be thought to steer it towards some types of change just as much as it steers it away from others. We'll look at various examples of this steering effect, which can be called developmental bias, in subsequent chapters.

The practitioners of evo-devo are not a homogeneous bunch. Those who are above a certain age have migrated there from various disciplinary backgrounds, because when they were students the field did not exist – or perhaps

had just begun but was not yet widely studied or well funded (regrettably the last of these remains true, though the situation is a little better than it used to be). Some practitioners have come from developmental biology, some from evolutionary biology, and some from elsewhere. Partly because of their heterogeneous origins, these practitioners are also rather heterogeneous in their views of the nature of evo-devo. With regard to the two sides of the interaction between evolution and development noted above, some emphasize one, some the other, and some both. This heterogeneity of views is of interest in relation to the wider philosophy of evo-devo. However, before considering the philosophical aspects of the new discipline, we need to know a little more about it – and that includes understanding its origins.

Origins of Evo-Devo: the Homeobox

If you were to ask me, 'what was the single most important discovery in the origin of evo-devo?' I'd reply, with little hesitation, 'the homeobox.' So that's a good place to start. This was a discovery made at the same time (in 1983, with publications following in 1984) by researchers in two laboratories – that of Thomas Kaufman in Bloomington, Indiana, and that of Walter Gehring in Basel, Switzerland. The lead authors of the papers concerned were Matthew Scott and William McGinnis.

The key question at this point is: what is a homeobox? To answer this we need to start with the related question: what is a gene? A reasonable working definition is that a gene is a stretch of DNA that's thousands of nitrogenous bases long, and that makes a particular product (typically a protein). Each different gene makes a different protein, because each gene is a unique sequence of the four bases that we're familiar with by their initial letters of A, C, G, and T (full names adenine, cytosine, guanine, and thymine). Recall that the genetic code works in triplets, so three bases in a DNA strand give rise to one amino acid in the resultant protein. For example, the sequence AAA in a gene corresponds to the amino acid lysine in the protein. If we say that a typical protein is 333 amino acids long (just a rough guess), then the gene making it must be at least 999 bases long. In fact, it is normally much longer than this for various reasons, principally that the genes of organisms from the kingdoms that have development (animals, plants, etc.) typically contain stretches of DNA (called introns) whose RNA counterparts are chopped out

before the protein is made. So the typical animal or plant gene is in fact thousands of bases long, rather than hundreds.

Now we return to the homeobox. A 'box', in genetics, is simply a rectangle drawn around a certain stretch of DNA to highlight it, for whatever reason. For example, if I wanted to highlight the AAA stretch in the following sequence to show you which bit of a longer sequence coded for the amino acid lysine, I'd draw a box around the central three bases (here I'm using bold text to do the same thing): TAT**AAA**GGG. The homeobox is a much longer sequence of bases than this. It is typically 180 bases in length, thus corresponding to a sequence of 60 amino acids in the corresponding protein – which in turn is called the homeodomain. In a homeobox-containing gene whose overall coding sequence is 1800 bases long, the homeobox is 10% of the gene (Figure 1.2), whereas in a gene that is 18,000 bases long (perhaps due to multiple long introns), the homeobox is just 1% of the gene's complete DNA sequence. In a homeodomain-containing protein that is 300 amino acids long, the homeodomain itself makes up 20% of this overall length.

So far, we recognize a homeobox as a sequence of a particular length that can be found within certain genes. But what *is* the sequence, which genes is it found in, and why is its inclusion in these genes significant?

It's not possible to specify the 180-base homeobox sequence *exactly*, because there are many variant versions of it. It's a recognizable pattern, or 'motif', rather than a precise sequence. Typically, one variant will be the same as another for most of the 180 bases, but will differ in a minority of them. Part of this variation is due to the redundancy of the genetic code. For example, it's not just AAA that codes for the amino acid lysine, AAG does so too. Thus it's possible to get a homeodomain with lysine in a particular position, by having either of these triplets in the corresponding stretch of the homeobox that codes for it.

But there is variation in the exact amino-acid sequence of homeodomains too. One variant homeodomain will typically be the same as another for most but not all of its amino acid sequence. The most important thing is that regardless of some variation in the structure of the homeodomain at this level, at a higher level its overall 3D structure is maintained. This 3D structure is much more complex than shown in Figure 1.2, which is schematic. We don't need to know this structure in detail, but its key feature is that it has three helical

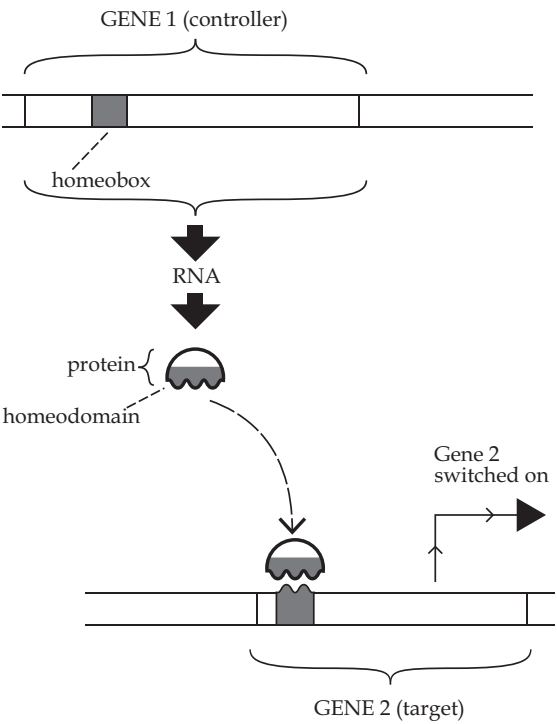

Figure 1.2 Diagram of a homeobox gene and the homeodomain protein that it makes. The homeobox represents only a small fraction of the overall length of the gene. Likewise, the homeodomain represents only a fraction of the protein. The important point to note here is that the protein's homeodomain enables it to bind to the DNA of other genes and to switch them on, thus causing them to make their own protein product. This provides a basis for developmental patterns of gene activity in the embryo.

regions that are vital to its DNA-binding function, and that are conserved in all variants.

Now let's turn to the genes in which a homeobox sequence is found. Genes can be roughly divided, in terms of their function, into three main categories: developmental (crucially important here), cell-type specific (less important),

and housekeeping (least important in evo-devo). As the names of these rough-and-ready categories imply, the first make proteins that contribute to the developmental process, the second make proteins that are restricted to certain cell types (such as the haemoglobin of our red blood cells), and the third make other proteins (often enzymes) that contribute to the housekeeping tasks that occur within almost all cells, such as the metabolic activities that keep a cell supplied with energy. As usual in biology, categories are never clear-cut, but this way of thinking about gene function is helpful nevertheless – and we'll delve further into it in Chapter 4.

Many developmental genes contain a homeobox; in contrast, other genes generally lack this sequence. There is a good reason for this difference. A major part of the causality of development is cascades of gene activity in which the product of gene A switches on gene B, whose product switches on gene C, and so on. To switch on a gene, a protein must bind to the DNA of that gene. And it turns out that the homeodomain is a DNA-binding region. So if we discover a new gene about which we initially know nothing, sequence it, and discover that it contains a homeobox, we are pretty sure that it plays a role in the development of the animal concerned.

The animal in which the homeobox was discovered was the fruit-fly *Drosophila melanogaster*. The genes in which it was discovered were already known from the fact that mutations of them produce bizarre adult flies that have legs growing out of their heads or two pairs of wings instead of the normal single pair. Back in the late nineteenth century, these mutations were called homeotic, and the phenomenon they produce was called homeosis – the right thing in the wrong place. This contrasts with the wrong thing in the right place, which is a much commoner type of mutation. An example of the latter in *Drosophila* is the vestigial-winged mutant, where the wings are in the right place but are small and malformed. The person who coined this terminology (homeosis/homeotic) was the British geneticist William Bateson, whose book *Materials for the Study of Variation* was published in 1894. We'll come back to him in Chapter 2, when we look at the antecedents of evo-devo. For now, it's just important to know that it was from his homeosis that the homeobox sequence got its name.

The huge significance of the homeobox arises from the fact that it was found to characterize multiple developmental genes and, as research continued in

the 1980s, it was found to characterize such genes not just in flies but in animals generally. This suggested that the genes that contributed to the developmental process were similar even when the end result of the process – the adult animal – was very different. In other words, it suggested that we could generalize about some aspects of the causality of development right across the animal kingdom, and perhaps even beyond. To say that this was exciting stuff would be an understatement. However, one discovery does not by itself make a new scientific discipline. So let's now look at some important events that preceded and followed the discovery of the homeobox.

Origins of Evo-Devo: Other Factors

In 1977, Stephen Jay Gould rekindled interest in the relationship between evolution and development with the publication of his book *Ontogeny and Phylogeny* (which, with a few ifs and buts, simply means Development and Evolution). Two years later, he and fellow Harvard biologist Richard Lewontin wrote an influential – and controversial – paper championing the role of various forms of constraint in evolution, including developmental constraint, and downplaying the role of natural selection. We'll discuss this paper, and the concept of constraint, in Chapter 4.

In 1978, the American geneticist Ed Lewis published an important paper on the structure of a gene complex in *Drosophila* that contained multiple genes that were subject to homeotic mutation and that would, a few years later, be found to contain homeoboxes. In 1980, the Heidelberg-based biologists Christiane Nüsslein-Volhard and Eric Wieschaus published an equally important paper on other genes that contribute to the development of *Drosophila*, including the now-famous *hedgehog* gene that gives mutant larvae a prickly appearance – hence the name. The genes studied by these three biologists were similar in that they all affected the development of segments – those sections of the body along the head-to-tail axis of which an insect is made. However, in another way they were different. The genes studied by Lewis were involved in the determination of segment identity (e.g. thoracic vs. abdominal). In contrast, those studied by Nüsslein-Volhard and Wieschaus were involved in the determination of segment number and segment polarity (which end of a segment is anterior and which posterior); these authors showed that such determination took place in stages of