13. The major transitions in evolution

Until the mid of the 19th century, natural history was dominated by a view that postulates a progressive development that started with non-living material and progressed from plants over simple to more elaborate animals. Lamarck postulated that all living beings would have an inherent tendency to climb a ladder of progression. Charles Darwin disagreed on this notion. Rather than having an intrinsic tendency for progression, the processes of mutation and selection lead to organisms that are well adapted to their environment. Today evolutionary biologists are aware of the fact that bacteria are not less successful in their environment than e.g. humans, nor are they less well adapted. However, without doubt some organisms are more complex than others. Again, there is neither any theoretical reason to expect, nor any true empirical evidence, that evolutionary lineages increase in complexity. Yet some taxa are more complex than others, e.g. eukaryotes are obviously more complex than prokaryotes (Figure 13.1). John Maynard Smith and Eörs Szathmary developed the fruitful idea that increases in complexity may have achieved as a result of a series of major evolutionary transitions, which involved changes in the way information is stored and transmitted. 41

13.1. What is a major transition?

**Complexity as information**

As said above, an increase in complexity is not a universal or inevitable evolutionary trend in all taxa. Some taxa may even show decreasing complexity if such reduction leads to adaptation to their environment, a trend that is often observed in parasites. However, the prevailing impression is that lineages have increased in complexity. This raises the question how to measure complexity. For sure, this task has not been convincingly solved as yet, but one way to deal with the problem is to focus on biological information. Since information is stored as DNA in most organisms, the amount of DNA may serve as an estimator of the degree of complexity (Table 13.1). In general, there is indeed a trend for increased genome size over evolutionary history. This estimate becomes more convincing when the proportion of the genome that encodes for proteins is taken into account, which can be quite low in some taxa, such as newts. In general, the larger a genome, the more “space” will there be for non-coding regions.

---

Three ways to increase complexity: Duplication, symbiosis and epigenesis

The most important ways to increase complexity may be based on the following three genetic processes (Figure 13.2): First, a gene may be duplicated. One of the resulting genes normally retains its original function, while the other one is so to speak free for evolution towards new function. Second, two or more originally independently replicating genetic units (“entities”) may come together in a symbiotic relationship. If this is followed by compartmentation, a new unit of replication may result, which is more complex than the original, independent entities. A famous example for such a process is endosymbioses whereby ß-proteobacteria gave rise to mitochondria 1500-2000 million years ago. Third, epigenesis is a process by which the pattern of gene expression rather than the genetic sequence itself is changed. The relevance of such processes for the increase of complexity will be discussed in more detail later (chapter 13.3).

Table 13.1 Genome size and DNA content.

<table>
<thead>
<tr>
<th></th>
<th>Genome size (base pairs x10 ^6</th>
<th>Coding DNA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterium (E. coli)</td>
<td>0.004</td>
<td>100</td>
</tr>
<tr>
<td>Yeast (Saccharomyces)</td>
<td>0.009</td>
<td>70</td>
</tr>
<tr>
<td>Nematode (Caenorhabditis)</td>
<td>0.09</td>
<td>25</td>
</tr>
<tr>
<td>Fruitfly (Drosophila)</td>
<td>0.18</td>
<td>33</td>
</tr>
<tr>
<td>Newt (Friturus)</td>
<td>19.0</td>
<td>1.5-4.5</td>
</tr>
<tr>
<td>Human</td>
<td>3.5</td>
<td>9-27</td>
</tr>
<tr>
<td>Lungfish (Protopterus)</td>
<td>140.0</td>
<td>0.4-1.2</td>
</tr>
<tr>
<td>Flowering plant (Arabidopsis)</td>
<td>0.2</td>
<td>31</td>
</tr>
<tr>
<td>Flowering plant (Fritillaria)</td>
<td>130.0</td>
<td>0.02</td>
</tr>
</tbody>
</table>
**Major transitions – a definition**

Based on what was said about complexity, we may now try to define what major transitions actually are (based on Maynard Smith and Szathmary).

"Major transitions are major stages in the evolution of complexity that involve a change in the level of organization, and hence the level of selection."

Such major transitions often have the following three key characteristics:

1. Entities that were capable of independent replication before the transition can replicate only as a part of a larger unit after it.
2. A major transition is often associated with division of labour and task specialization.
3. A major transition often implies a change in the way in which information is transmitted between generations.

According to Maynard Smith and Szathmary, increased complexity depends on a rather small number of such major transitions. Examples for major transitions are listed in Table 13.2.

**Table 13.2 The major transitions**

<table>
<thead>
<tr>
<th>Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicating molecules to populations of molecules in compartments</td>
</tr>
<tr>
<td>Unlinked replicators to chromosomes</td>
</tr>
<tr>
<td>RNA as gene and enzyme to DNA and protein (genetic code)</td>
</tr>
<tr>
<td>Prokaryotes to eukaryotes</td>
</tr>
<tr>
<td>Asexual clones to sexual populations</td>
</tr>
<tr>
<td>Protists to animals, plants and fungi (cell differentiation)</td>
</tr>
<tr>
<td>Solitary individuals to colonies (non-reproductive castes)</td>
</tr>
<tr>
<td>Primate societies to human societies (language)</td>
</tr>
</tbody>
</table>

**Major transitions may lead to conflict**

One aspect of the definition of major transitions needs special interest. It was said that “Major transitions are major stages in the evolution of complexity that involve a change in the level of organization, and hence the level of selection.”

This change in the level of selection bears the risk that selection on entities at the lower level (e.g. mitochondria, or a non-reproductive worker) may disrupt integration at the higher level (the eukaryotic cell, or the insect colony, respectively). The problem is not an imaginary one: there is a real danger for such conflicts (see examples in Table 13.3). The following chapters, where we will focus on two examples of major transitions (leading to the genetic code and to multicellular organisms) will show under what conditions the conflict exists, and what mechanisms reduce potential conflict.

**Table 13.3 Examples of conflict between selection at different levels**

<table>
<thead>
<tr>
<th>Form of cooperation</th>
<th>Exceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A fair meiosis</td>
<td>Meiotic drive, transposition</td>
</tr>
<tr>
<td>Sexual reproduction</td>
<td>Parthenogenesis</td>
</tr>
<tr>
<td>Differentiation of somatic cells</td>
<td>Escape from growth control</td>
</tr>
<tr>
<td>Non-reproductive castes of social insects</td>
<td>Egg-laying worker bees</td>
</tr>
</tbody>
</table>
13.2. The genetic code

Evolution of the genetic code implies the replacement of RNA (serving as both gene and enzyme) by DNA and protein. This split of function between nucleic acids and proteins implies translation between two different languages: a nucleotide sequence has to be converted into an amino acid sequence (Figure 13.4). For this conversion, a code is necessary – the genetic code. It is difficult to understand how such a rather complex system could evolve. We will approach the problem by splitting it into three main questions:

1. Why was information storage separated from enzymatic function?
2. How did nucleotide – amino acid associations originate?
3. Why is DNA instead of RNA used for information storage?

**Figure 13.3** Translation as it occurs today

**Why was information storage separated from enzymatic function?**

Today’s genetic machinery is based on high fidelity copying devices such as polymerases with proof reading functions. With amazing precision, large amounts of genetic data are duplicated and thereby stable over many generations of replication. When we want to imagine the evolutionary starting point that might have led to such high-precision machinery, we run into a problem: to enable sufficiently accurate replication of a large genome, accurate (and therefore large) enzymes are necessary. However, such large enzyme can only be coded for with a large genome. But what was first? This version of the chicken-and-egg problem, which has been addressed mathematically, is known as Eigen’s paradox: “No enzymes without a large genome, and no large genome without enzymes.” Although this does not resolve the problem completely (for a more thorough approach see the next lecture, chapter 14.6), part of the solution results from the likely fact that, originally, genetic information storage was not separate from enzymatic function. Rather, both functions were performed by so-called ribozymes (Figure 13.4).

**Figure 13.4** A ribozyme (left), RNA (middle) and an enzyme (triosephosphat isomerase, right)
What was the reason for separating information storage from enzymatic function during evolution? Most likely, this division of labour reduced mutational load during replication and, at the same time, enabled increased catalytic efficiency. Why so? Consider a simple analogue: Imagine you have to build a machine consisting of 1000 pieces that are lying in front of you. In the first case, there are 20 different types of pieces, whereas there are only 4 different types in the second case. You have to be quick, so you will occasionally grab the wrong type of piece. Which machine will contain fewer wrong pieces? Quite obviously, you will have made fewer mistakes when there were only 4 different types of pieces. For the same reason, a code that is based on only four letters (such as nucleic acids that are based on four different nucleotides) will show a lower mutational load during replication than a code that would consist of 20 amino acids. In fact, the decrease in copying fidelity with the number of monomer types has been shown to be faster than exponential, and models show that exactly 4 monomer types might be the optimal number.

Now imagine you have to build many different machines serving different purposes. The active centre of the machine consists of four monomers selected again from either 4 or 20 different types. In the first case, there are only $4^4 = 256$ different active sites, whereas in the second case, there are $20^4 = 16,000$ possibilities. Thus, with regard to the versatility of catalytic sites, enzymes made of 20 different amino acids are clearly superior to nucleic acids.

**How did nucleotide – amino acid associations originate?**

The above chapter shows that division of labour between nucleic acids and proteins makes sense. However, this does not give an explanation how this process could have occurred during evolution. Importantly, a useful end-product of an evolutionary process does not imply that such a process has occurred, since we need to find benefits resulting from all intermediate steps, otherwise they would not have happened. Different mechanisms may be involved in the origin and the maintenance of the genetic code. We may also use knowledge about today’s ribozymes and protein enzymes to bring light into the process that might have lead to the division of labour between nucleic acids and proteins.

The following scenario, although difficult to test, might explain how protein enzymes evolved from ribozymes. Oligopeptides that were originally used by ribozymes as coenzymes might have gradually been extended, until the long peptide chain took most of the enzymatic activity, with just some nucleotides remaining as coenzymes (Figure 13.5).

**Figure 13.5** Evolution of coenzymes from polynucleotides. The emphasized region of the RNA molecule represents part of the active site in the ancestral ribozyme, which evolves into a cofactor of a protein enzyme.

However, the above scenario does not explain how associations between certain amino acids and certain nucleotide triplets (i.e. the assignments of the genetic code) came into existence. One possibility is that amino acids were originally used as coenzymes of ribozymes (Figure 13.6). To enable binding of certain amino acids to ribozymes that made use of them as coenzymes, the amino acids might have been equipped with trinucleotide ‘handles’ as identifiers. Other ribozymes might have specialised on linking these ‘handles’ to the amino acid, becoming ancestral assignment enzymes. This scenario is plausible, but not easily testable.
The question remains whether the associations between one particular amino acid and a specific trinucleotide (i.e. the codon assignment) is random or not. To address this question, it may be helpful to analyze the genetic code, as it exists today (Figure 13.7). This is a so-called top-down approach, as opposed to the bottom-up scenarios given above.

Several observations suggest that codon assignments are not completely random. Firstly, chemically similar amino acids seem to have similar codons. Secondly, the first codon position seems associated with biosynthetic kinship of amino acids. Thirdly, the second codon position is associated with polarity of the amino acid it codes for. These relationships make it likely that the genetic code with its codon assignments is not simply a ‘frozen accident’. Rather, it likely has evolved to minimize mutational load, i.e. the negative effects that a
mistake during translation has (it is e.g. less harmful if one polar amino acid is substituted with another amino acid that is also polar). Evolution of codon assignment implies that an original assignment must have been changed by re-assignment and, potentially, transient or original ambiguity (i.e. one codon codes for several amino acids or *vice versa*). How universal is the genetic code? Although there are deviations from the ‘normal’ genetic code in some taxa, these are fairly minor deviations from a robust theme (Table 13.4). Such changes in the universal code may give valuable hints on mechanisms that might have been underlying evolution of the code.

**Table 13.4** Changes in the universal code

<table>
<thead>
<tr>
<th>Codon in universal code</th>
<th>Changes to</th>
<th>System</th>
<th>Nuclear</th>
<th>Mitochondrial</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGA, Stop</td>
<td>Trp</td>
<td>Mycoplasma</td>
<td>All except plants</td>
<td>Yeast</td>
</tr>
<tr>
<td>AUA, Ile</td>
<td>Met</td>
<td>Mycoplasma</td>
<td>Yeast</td>
<td>Metazoans, except echinoderms</td>
</tr>
<tr>
<td>Met to Ile</td>
<td></td>
<td></td>
<td></td>
<td>Echinoderms</td>
</tr>
<tr>
<td>AGR, Arg, Ser, Stop</td>
<td></td>
<td></td>
<td></td>
<td>Metazoans, except vertebrates</td>
</tr>
<tr>
<td>AAA, Lys, Asn</td>
<td>Acetabularia</td>
<td>Ciliated protozoans except Euplotes</td>
<td>Vertebrates</td>
<td></td>
</tr>
<tr>
<td>UAR, Stop, Gin</td>
<td></td>
<td></td>
<td></td>
<td>Flatworms, echinoderms</td>
</tr>
<tr>
<td>CUN, Leu, Thr</td>
<td>Candida cylindracea</td>
<td></td>
<td>Yeast</td>
<td></td>
</tr>
<tr>
<td>CGU, Leu, Ser</td>
<td>Vertebrates, eubacteria (in special enzymes)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UGA, Stop, SeCys</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*From Jukes & Osawa, 1991.*

**Why is DNA instead of RNA used for information storage?**

We have seen why the division of labour between nucleic acids and proteins makes sense, and how this transition, as well as certain codon assignments (i.e. the genetic code), might potentially have evolved. However, we did so far not separate between two different types of nucleic acids, RNA and DNA.

![Figure 13.8 Differences between RNA and DNA.](image)

The vast majority of organisms makes use of DNA instead of RNA for information storage, whereas the process of translation involves RNA intermediates. RNA was likely used as the original nucleic acid. Why was then DNA instead of RNA used for the information storage?
Most likely, the increased chemical stability of DNA makes it more suitable for this purpose, because it reduces mutational load. Both the use of deoxyribose instead of ribose and the use of thymine instead of uracil increase chemical stability. As above, the fact that DNA is now more useful does not give any reason as to why (and how) RNA might have been replaced by DNA in the beginning.

Some important general principles, which are relevant for the evolution of the genetic code, but also for many other major transitions, became clear in this chapter:

- Different mechanisms may be involved in the origin and the maintenance of higher-level organization (e.g. the origin and the maintenance of the genetic code).
- Contingent irreversibility: If an entity has replicated as part of a larger whole for a long time, it may have lost the capacity for independent replication that it once had (e.g. ribozymes can no longer replicate on their own).
- Division of labour leads to a non-additive, synergistic benefit (e.g. DNA for information storage, proteins for enzymatic function).

### 13.3. Multicellularity

**The evolution of multicellular organisms**

Early life consisted of single-celled organisms. The earliest definitive fossils of multicellular animals are found in the Ediacaran fauna, an Australian deposit dating to 670-550 million years ago. Since this early metazoan fauna was dominated by soft-bodied animals fossils are rare (Figure 13.9).

The Ediacaran fauna was followed by the impressive ‘Cambrian explosion’ (Figure 13.10), during which almost all metazoan groups appeared. Especially before knowledge of the Ediacaran fauna, it had been argued that this observation of a rapid increase in biodiversity might have been an artificial impression resulting from the fact that these animals, for the first time, possessed hard body structures and shells, so that fossils remained. More likely, the ‘invention’ of one or several key adaptations might have led to radiation that was indeed extremely fast. Multicellularity might have played an important role here. The transition to multicellularity likely occurred three times independently, leading to plants, animals and fungi (Figure 13.11).

![Figure 13.9 Examples of the Ediacaran fauna. (a) an attached cnidarian, (b) a jellyfish, (c, d) worms, (e) a possible pro-echinoderm](image-url)
Figure 13.10 The fossil record of most major animal groups begins in the Cambrian. The horizontal lines to the right of each taxon show the times when the group is represented in the fossil record.

Figure 13.11 The transition to multicellularity likely occurred three times, in the plants, animals and fungi. Pictures show devonian fossils of a plant (*Sphenopteris*, left), a fungus (*Paleomyces*, middle) and an animal (*Trilobite*, right). These examples are already highly derived, while early representative of multicellular organisms were far simpler but left no fossils (http://www.ucmp.berkeley.edu).

**Simple metazoans as ‘models’ for early multicellular organisms**

How might early multicellular organisms have looked like? An extant example of a very simple metazoan is *Trichoplax adhaerens* (Placozoa) (Figure 13.12). Superficially resembling an amoeba (which is, however, unicellular!), this organism is more simply organized than any other living metazoan. Even though, it already shows a high degree of cellular differentiation.
The major transitions in evolution

Figure 13.12 *Trichoplax adhaerens* (Placozoa, left) is more simply organized than any other living metazoan taxon, but it already shows a high degree of cell differentiation (middle). Its movement is amoeboid (right). (Pictures from Schierwater, B. (2005) Bioessays 27, 1294-1302).

Potentially even simpler with regard to cellular differentiation are sponges, which have, however, a more complex body plan. Interestingly, a certain cell type of sponges, the choanocytes, resembles very much unicellular choanoflagellates (Figure 13.13). Haeckel (1834-1919) already proposed that metazoans evolved form a colonial choanoflagellate, such as *Proterospongia haeckeli*. Only a simple central jellied matrix is found in *P. haeckeli*, whereas sponges have a more complex matrix, the mesohyl.

Figure 13.13 Choanoflagellates, such as *Proterospongia haeckeli*, are potential ancestors of the metazoa.

Why did multicellularity evolve?

The impressive radiation of metazoans suggests that multicellularity was an important evolutionary step towards complex organisms. However, it is important to keep in mind that we cannot explain transitions (here from univellularity to multicellularity) in terms of the ultimate benefits they conferred, because these cannot be the reason why the change occurred in the first place. The transitions must be explained in terms of immediate selective advantage to individual replicators (here: cells). So what were immediate benefits of multicellularity?
One possibility is that multicellularity might have helped against predation. Experimental studies that use facultative colonial unicellular organisms seem to support this hypothesis. For example, the green alga *Chlorella vulgaris* tends to form colonies when exposed to predation by the predatory flagellate *Ochromonas vallescia* (Figure 13.14). In response to predation pressure, large colonies dominated in the beginning of a long-term experiment, but were substituted by 8-cell colonies later on (Figure 13.15). Colonies of this size were virtually immune to predation by the flagellate, but still small enough that each *Chlorella* cell was exposed directly to the nutrient medium.

**Figure 13.14** (a) The predatory flagellate *Ochromonas vallescia* (Oc), with long (F1) and short (Fs) flagella, a *Chlorella* colony (CC) and a single *Chlorella* cell (FC) sampled together from a culture 240 days after inoculation.

**Figure 13.15** Time course of *Chlorella* colony size distributions in a mixed-species continuous culture of *Chlorella* and *Ochromonas*. The graph shows the *Chlorella* colony size vs. time. The majority of *Chlorella* cells were in colonies with more than 24 cells per colony for the first month of culture. The large colonies then disappeared from the culture and the number of cells per colony stabilized at eight.

Another benefit that might have played a role in an early stage of the evolution of multicellularity is that it enables division of labour between different cell types. One of the most important divisions is between somatic cells and specialised reproductive cells. *Volvox weismannia* is a simple organism with non-reproductive and reproductive cells and may thus

---

serve as an extant example that shows how an early multicellular organism with specialised germ cells could have looked like (Figure 13.16). An interesting precondition for the division of labour in the Volvocales is that motile cells cannot divide, and mitosing cells cannot move, because the same organelles are used either as basal bodies or as centrioles. Such a constraint that prevents simultaneous flagellation (i.e. movement) and cell division is probably more general within the metazoa, and may thus have been an important benefit of multicellular organisms (which thus can move and divide at the same time).

![Image](image1.png)

**Figure 13.16** Reproduction in *Volvox*.

*A conflict over access to the germ line?*

The distinction between germ line and soma goes back to August Weismann (1834-1914; Figure 13.17), who thereby destroyed Lamarck’s logic of the inheritance of acquired characteristics. The germ line consists of the gonads and their egg and sperm cells. These are normally sequestered early in animal development; cells from the rest of the body, the soma, do not later move into the gonads (Figure 13.18).

![Image](image2.png)

**Figure 13.17** The concept of the germ line.

---


42 Interestingly, even C. Darwin resorted to the idea of inheritance of acquired characteristics in later editions of *The Origin of Species*, in reaction to criticism of his model of blending inheritance.

Only cells that enter the germ line will be represented in the next generation. So, is there a conflict among individual cells of a multicellular organism over access to the germ line? Normally, such genomic conflict is unlikely, since all cells of an organism stem from a single cell and are thus genetically identical, except for (rare) mutations occurring during divisions form the single cell stage. However, there are systems where colonies were formed by aggregation of potentially unrelated cells (Figure 13.19). In this case, conflict over access to the germ line is possible, and work is still in progress to elucidate how such conflict is resolved. One option is that cells aggregating to form a colony are closely related. Avoidance of conflict over germ line access might also explain why development of multicellular organisms usually starts from a single cell, rather than aggregation of multiple cells. Likewise, cell aggregation is a rather unlikely way for the evolution of multicellularity (Figure 13.20), and extant organisms such as *Dictyostelium* are possibly not the best models for understanding the evolution of multicellularity.

Figure 13.18 Migration of GFP labelled primordial germ cells (PGC) in wild-type and odysseus (*ody*) chemokine receptor mutants of zebrafish (*Danio rerio*) embryos. Wild-type PGC migrate into the gonads (a-c), while *ody* mutant PGC go astray (e-g).

Figure 13.19 In the slime mold *Dictyostelium discoideum*, cells aggregate to form a fruiting body. Only part of the cells reproduces, while others end up in the stalk.
Cell differentiation is based on a ‘dual inheritance system’

We have seen in the previous section that multicellular organisms normally consist of genetically identical cells (except for somatic mutations), which prevents genomic conflict. Yet these genetically identical cells need to differentiate into different cell types to enable development of a multicellular organism. When a differentiated cell divides, daughter cells normally inherit the differentiated state, e.g. a liver cell will normally give rise to daughter cells that will again be liver cells. This requires regulation of gene expression and a certain form of cellular memory. Maynard Smith thus suggested that there is a ‘dual inheritance system’: The primary system is based on the DNA sequence, whereas the secondary system is based on states of gene activity. The latter system establishes cell heredity, which is the somatic inheritance of the differentiated state of the cell through cycles of cell division (i.e. epigenetic inheritance). A mechanism that is mainly responsible for epigenetic inheritance in many metazoa is DNA methylation of cytosin in CpG motifs (Figure 10.26).
Epigenetic inheritance seems to have evolved already in the protists. It might have enabled adaptation to environmental changes at a timescale too large to adapt within a generation, but too short for evolutionary change. It might also have been a system to recognize foreign genetic material (i.e. parasitism). Epigenetic inheritance would then have been a key pre-adaptation for the evolution of multicellularity. Cell heredity has important medical implications. For example, the use of adult stem cells, as an alternative to embryonic stem cells (Figure 13.21), largely depends on the ability of these cells to develop into diverse cell types, i.e. the absence of cell memory that is already too fixed. Finally, cancer cells have been regarded as cells that do not follow their epigenetic program any more (Figure 13.22). Whether or not this can be seen as resulting from the conflict between the interests of a mutated cell clone and the higher unit of the individual remains a matter of debate.

Figure 13.22 Cancer cells. Left: Normal fibroblasts show contact inhibition. Right: Transformed fibroblasts show uncontrolled growth.

13.4. **Take home messages**

- **Major transitions are stages in the evolution of complexity that involve a change in the level of selection.**
- **Different mechanisms may be involved in the origin and maintenance of a new level of complexity.**
- **Major transitions often involve a change in the way in which information is transmitted.**
- **Division of labour and task specialization are potential benefits resulting from major transitions.**
- **Conflicts may arise between selection on the original level and the higher level, especially when the sub-units are genetically not identical.**

13.5. **Literature**

*Recommended book for this chapter:*


*Further reading:*