INTRODUCTION

Evolutionary biology is based on several concepts. Learning the basic concepts of any field of science is inherently difficult due to a vicious circle: they are necessary to comprehend the relevant information but, before this information is absorbed, it is hard to see what they really mean. With evolutionary biology, there is an extra obstacle: its basic concepts are very heterogeneous. Some of them are common to all natural sciences, exact, and formal, while others are confined to biology and intuitive, because the enormous complexity of life cannot be, at least now, studied at the level of rigor expected in physics. Here I introduce the concepts of evolution, anagenesis, cladogenesis, phase space, determinism, space of genotypes, levels of organization, phenotype, trait, fitness, adaptation, mutation, variation, population, selection, allele replacement, Lamarckian and Darwinian evolution, fitness landscape, similarity, relatedness, compatibility, connectedness, clade, species, complexity, optimality, evolvability, designability, stochasticity, reproducibility, random drift, Microevolution, Macroevolution, and some others as simply as I can. Many additional details will appear later, when these concepts will be used and elaborated. Still, this Introduction is one of the most difficult parts of the book. You may wish to return here occasionally, while studying the later Chapters.

1. What is evolutionary biology?

Above all, evolutionary biology studies natural origin of living beings with new qualities. We will see that organisms of kinds which were previously absent originate in the course of many generations, as a result of mostly gradual, slow modification. While time flows, organisms which currently represent a lineage, a succession of organisms connected to each other by ancestor-descendent relationships, keep changing and deviate more and more from the initial ancestors. This phenomenon is called evolution. Because extant life is a product of evolution, the two key tasks of evolutionary biology are to understand the two key properties of life - its staggering diversity and complexity.

In addition to evolution in this narrow sense, the dynamics of genetic variation within a lineage, the processes leading to multiplication or extinction of evolving lineages, the origin of life from non-living matter, and controlled evolution in captivity also belong to the domain of evolutionary biology. The same is true for designing and

creating novel macromolecules, metabolic pathways, and, in the future, even cells and organisms. In short, evolutionary biology investigates any substantial changes of life.

Modern evolutionary biology can be conveniently subdivided into three distinct, although interconnected, branches: 1) study of past evolution, 2) study of Microevolution, and 3) study of Macroevolution. The following reasons justify this structure, which is reflected in the structure of this book.

1) Unfortunately, natural evolution is so slow that we usually cannot directly observe Macroevolution, *i. e*. profound changes of organisms. Still worse, we do not yet have a theory of Macroevolution, capable of explaining, to say nothing of predicting, diverse facts on the basis of few general principles. Fortunately, life on Earth evolves for over three billion years, and there are a lot of data on the history of life. Thus, since XIX century, studies of past evolution remain our main source of knowledge of Macroevolution.

2) A remarkable feature of biodiversity is that no organism is a loner - instead, everyone belongs to a population of similar organisms. Evolution occurs due to action of Darwinian natural selection on Mendelian genetic variation within a population. The dynamics of this variation are called Microevolution. Although the range of within-population variation is very narrow in comparison to differences between distinct forms of life, Microevolution drives even the most profound Macroevolution and is essential for understanding the origin of biodiversity. Studies of Microevolution, both theoretical and experimental, were very successful in the XX century.

3) Unfortunately, knowing the course of past evolution and the mechanisms of Microevolution is still not enough to really understand Macroevolution. How can an enormously complex object, such as human genome, hemoglobin, or eye, evolve gradually from something simple in such a way that all the intermediate states were functional? To what extent the evolutionary origin of modern organisms restricts their adaptedness? Why do many organisms reproduce sexually and experience senescence? Despite the ongoing progress of biology, these old questions remain unanswered. Hopefully, studies of Macroevolution will be successful in the XXI century.

Evolutionary biology is often referred to as "theory of evolution". This is not the most suitable choice of words. On the one hand, evolutionary biology as a whole did not yet earn the status of a comprehensive theory - at best, we have specific theories for only

some, relatively simple, aspects of evolution. On the other hand, if "theory" is used colloquially, as "hypothesis", evolutionary biology is more than that, because past evolution of life on Earth has been established as firmly as any other fact.

2. The Weak and the Strong claims about life in the past

Despite the ongoing mass extinction, we are still surrounded by hundreds of thousands of diverse, complex, and beautiful species, each adapted to its environment (Fig. I1). For a while, I will use the word species simplistically, to refer to a more or less distinct kind of organisms, representing a particular lineage at some moment of time. An essential task of biology is to explain how these species came into being. The general explanation, that modern life is the result of long evolution, is provided by evolutionary biology in its historical aspect. This striking and by no means obvious claim can be subdivided into two parts, the Weak Claim and the Strong Claim.

Fig. I1. Three charismatic species - female and male agamas from Kafarnahum (Israel), an orchid from Valaam (Russia), and a caterpillar from Ethiopia.

Everybody knows that all contemporary organisms have parents. Those who appear as the result of gamete fusion in the course of sexual reproduction may have two parents, while all others have just one. Indeed, no natural process can precisely assemble even the simplest genotype from scratch outside a living cell, since even a single deviation from the correct sequence of ~10^5 or more nucleotides may be lethal. Besides, even if a meaningful, but naked, genotype appears somehow, it would not be viable

without the enormously complex cell around it. Thus, modern cells originate only from other cells, and the two daughter cells inherit their genotypes, produced by DNA replication, from the mother cell.

Remarkably, not too long ago so little was known about the world we live in that this thesis was not at all obvious. Only in 1861, Louis Pasteur have finally shown that bacteria cannot originate from non-living mixtures of organic compounds. Spontaneous generation of mice from dirt was rejected a little earlier, in the XVIII century.

Everybody also knows that offspring are similar to their parents. Of course, this similarity is not exact, especially when reproduction is sexual. Still, elephants never beget bears (Fig. I2). Thus, while tracing backward the lineage of ancestry of an elephant (*i. e.*, considering its parents, grandparents, *etc*.), one may expect to encounter only elephants, all the way down to the First Elephant, but no bears, fruit flies, or apple trees.

Fig. I2. Mother and father elephants stare with disgust at their baby bears.

However, this expectation is wrong. Instead, according to the Weak Claim, the more and more remote ancestors of a contemporary species were more and more different from it. Thus, the lineage of descent which eventually led to modern elephants (*e. g.*, the Asian elephant, *Elephas maximus*) evolved, in the sense that it consisted of very different organisms at different moments in the past, although bears, flies, or trees were not among them (Fig. I3). The process of evolution of a lineage, asserted by the Weak Claim, is sometimes called descent with modification, or anagenesis. Because cosmology informs us that the Universe is only ~13.7 billion years old, every lineage must have had a beginning, but for now we will ignore the crucial issue of the origin of life. The Weak Claim may seem strange, although we must admit that it cannot be dismissed outright:

although differences between parents and offspring must be slight, their slow accumulation may eventually lead to profound evolution.

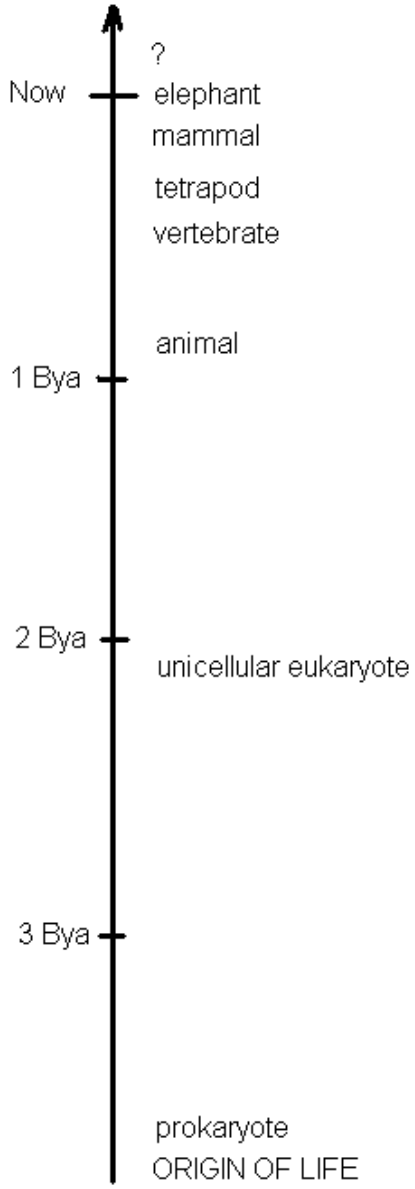


Fig. I3. A sketch of the history of a lineage which is currently represented by elephants.

Still, even if the Weak Claim is true, lineages which are currently represented by, say, Asian elephants, African elephants (*Loxodonta africana*), humans (*Homo sapiens*), and common fruit flies (*Drosophila melanogaster*) could be of independent origin and, thus, could have been distinct during all the time they existed in the past. This was the

view of Jean Baptiste Lamarck, who in 1809 published "Philosophie Zoologique", the first major work on evolution of life. In other words, Lamarck thought that modern species are products of evolution but are not related to each other.

However, this opinion is wrong, according to the Strong Claim. Instead, if we consider a set of contemporary species and trace the lineages which led to each of them backward, we will see these lineages merging, usually at different moments in the past, until eventually all of them become just one lineage, which represents the common ancestor of the set. The process of splitting of a lineage, asserted by the Strong Claim, is called cladogenesis. Graphical description of this branching process is called phylogenetic tree (Fig. I4). Compelling arguments in favor of the Strong Claim were presented in 1859 by Charles Robert Darwin in a treatise titled "On the Origin of Species by Means of Natural Selection", one of the most brilliant books ever written.

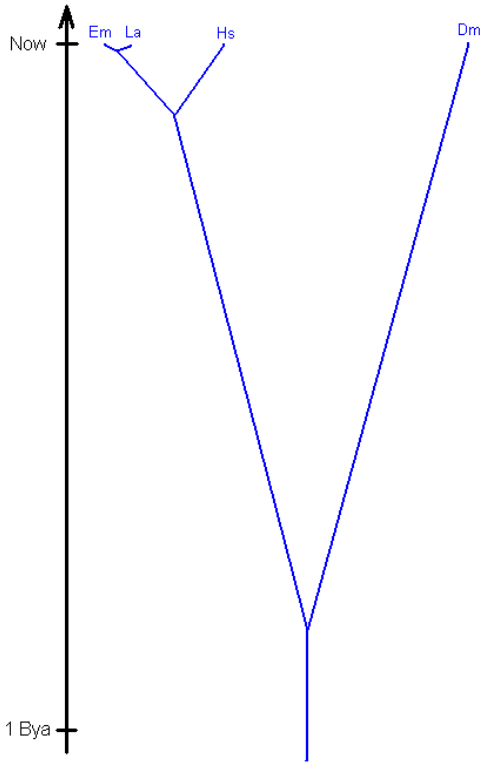


Fig. I4. Phylogenetic tree representing the process of the origin of the two elephants, humans, and fruit flies from the common ancestor. The first branching event probably occurred 800-550 Mya, the second – 120-90 Mya, and the third - 8-5 Mya.

Because organisms which belong to successive generations within a lineage must be very similar to each other, a striking implication of two contemporary species having the common ancestor is that the gap between them can be bridged by a continuous series of intermediate forms, with interbreeding possible between the adjacent sexual forms (Fig. I5). Thus, the Strong Claim may seem even stranger than the Weak Claim. Indeed, extant species are often clearly distinct: the ranges of variation within Asian elephants, African elephants, humans, or common fruit flies are much narrower than gaps between them.

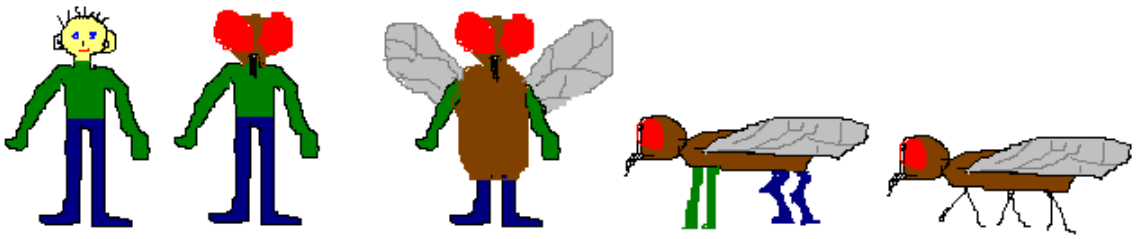


Fig. I5. *Homo sapiens*, *Drosophila melanogaster*, and wacky intermediate forms.

Still, we must admit that the Strong Claim is not impossible. Although changes of an evolving lineage during one generation are much smaller even than the range of its variation, long, independent evolution of two lineages since the moment of their split may eventually produce species as different as humans and flies. Moreover, intermediate forms do not need to be monsters shown in Fig. I5; instead, our last common ancestor with flies probably was a primitive worm. If adjacent intermediate forms lived at different times in the past, as they mostly did, they cannot provide stepping stones for any genetic exchange between modern humans and flies.

Obviously, if the Strong Claim is true for a set of contemporary species, the Weak Claim must also be true for each of them, with a possible exception of just one species: the common ancestor of a set of species can, at most, coincide with only one of them and probably was different from all. The Weak and the Strong Claim are hypotheses about events in the past. In Part 1 we will examine the data on the past life on Earth and see that both these Claims are true. The last universal common ancestor (LUCA) of all modern species lived over 3,500,000,000 years ago and was a rather advanced prokaryote.

3. Phase spaces and determinism

Natural sciences usually deal with objects which can exist in more than one state. A powerful way of thinking about such an object is to imagine a space which consists of all the possible states of the object, called its phase (state, configurational) space. The nature of the phase space depends, of course, on how the object is described. Let us consider a few examples.

1) The phase (related meaning!) of a chemical compound - solid, liquid, or gaseous - is determined by its two parameters, temperature and pressure. For each compound, there are three regions in its two-dimensional temperature x pressure phase space, corresponding to these phases. We can say that the phase of a compound is a function within its phase space, with only 3 possible values. Obviously, there must be a triple point, the combination of temperature and pressure where the three regions meet, and all the three phases coexist (Fig. I6).

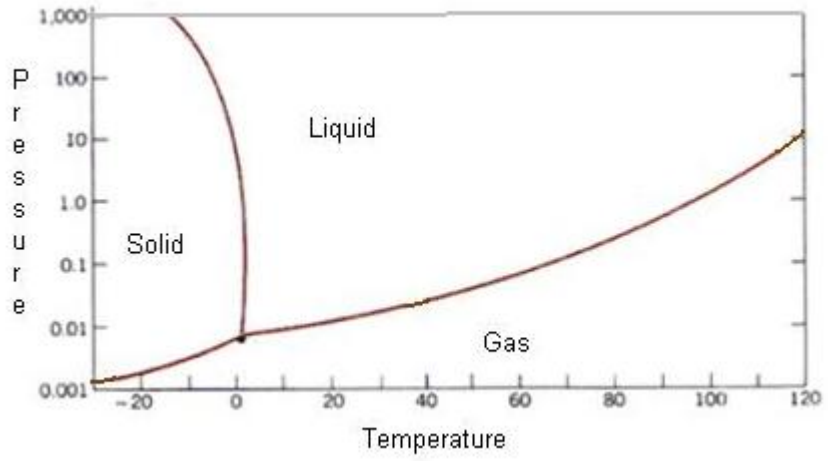


Fig. I6. Phases of water in the two-dimensional Euclidian temperature-pressure phase space.

2) A taxpayer can be described by the amounts of money paid, during a particular year, as federal, state, city, local, and property taxes. Thus, the corresponding phase space is 5-dimensional and a person is just a point in it. Evidently, a phase space of more

than 3 dimensions cannot be easily represented by a picture, but a table may be used to list descriptions of the available objects.

3) Some objects cannot be fully described by any finite number of parameters. For example, if we consider shapes of wings of different flies, our objects are the whole contours of wings (Fig. I7). A generic curve is an infinite-dimensional object, and the corresponding infinite-dimensional phase spaces are widely used in natural sciences.

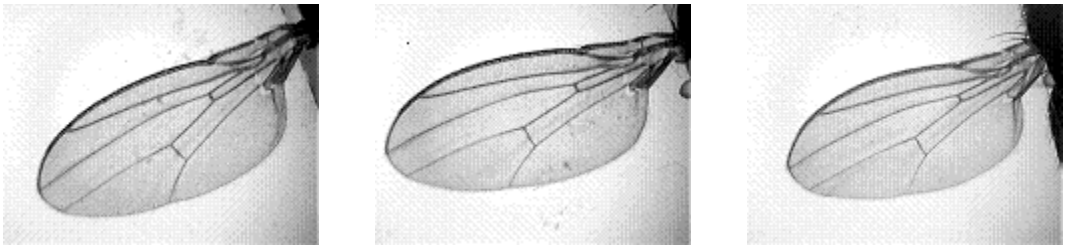


Fig. I7. Shapes of wings of several species of fruit flies (David Houle).

Of course, it is always desirable to keep the description of an object as simple as possible and, thus, to consider only essential parameters. What is essential depends on a particular problem. A tax auditor needs fewer parameters to describe a human being than a cardiosurgeon. For many purposes, an infinite-dimensional wing can be described just by its length, width, and one or two more parameters characterizing its shape.

The concept of phase space is especially valuable if we are interested in the dynamics of an object, *i. e.* in how it changes with time. A changing object accepts different states at different moments, and its phase space allows us to think of all the possible states simultaneously, although only one of them is present at a time, and many states will probably never be realized. Changes are then represented by movements of the point which describes the current state of the object. Parameters of a changing object are called variables. A description of a changing object must include all its variables which can help to predict its states in the future. Let us consider several examples.

4) If a very large number N of radioactive atoms is observed, a fixed fraction λ of them will decay in the course of a year (for example, $\lambda = 5.34*10^{-10}$ for ^{40}K). Thus, macroscopic dynamics of radioactive decay are described by a differential equation $dN/dt = -\lambda N$, so that N declines exponentially:

$$N(t) = N_0 e^{-\lambda t} \qquad \text{(I-1)}$$

where N_0 is the initial number of atoms (Fig. I8, see Chapter 1.2). Here N is the only variable and λ is the only fixed parameter of the process.

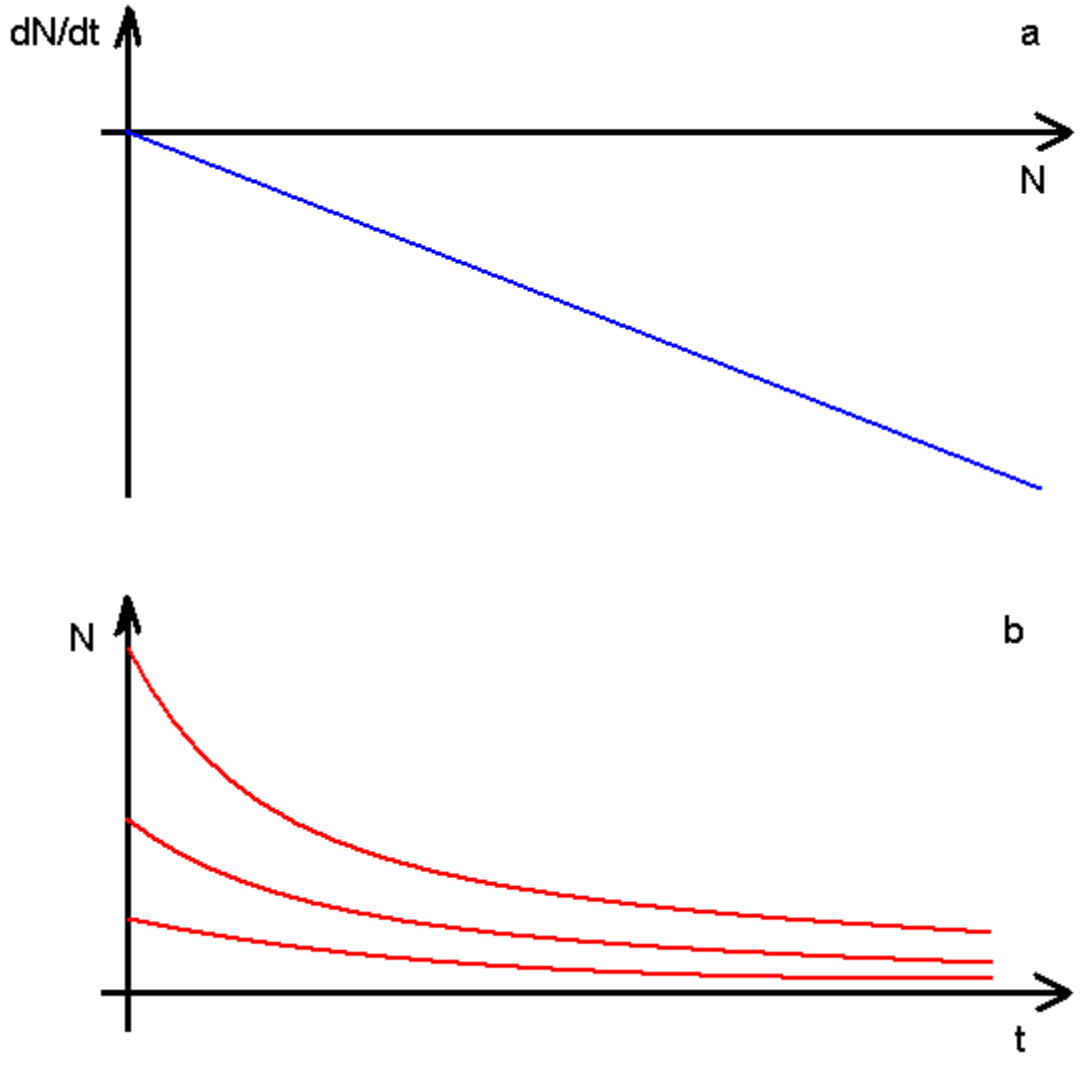


Fig. I8. Velocity of N, the number of ^{40}K atoms, as a function in the phase space of N (a) and the corresponding dynamics of N as movements within the same phase space (b).

5) Consider selection acting on a locus A with alleles A and a (Chapters 2.1 and 2.2). If the population is big enough, the frequency of allele A, [A], can be treated as a continuous variable. In the simplest case, the velocity of [A] is determined by the current

value of [A] in the following way: $d[A]/dt = s[A](1-[A])$, where s is the invariant selective advantage of A over a. This leads to the following dynamics of [A]:

$$[A](t) = 1/(1+((1-[A]_0)/[A]_0)e^{-st}) \quad \text{(I-2)}$$

where $[A]_0$ is the initial frequency of A (Fig. I9).

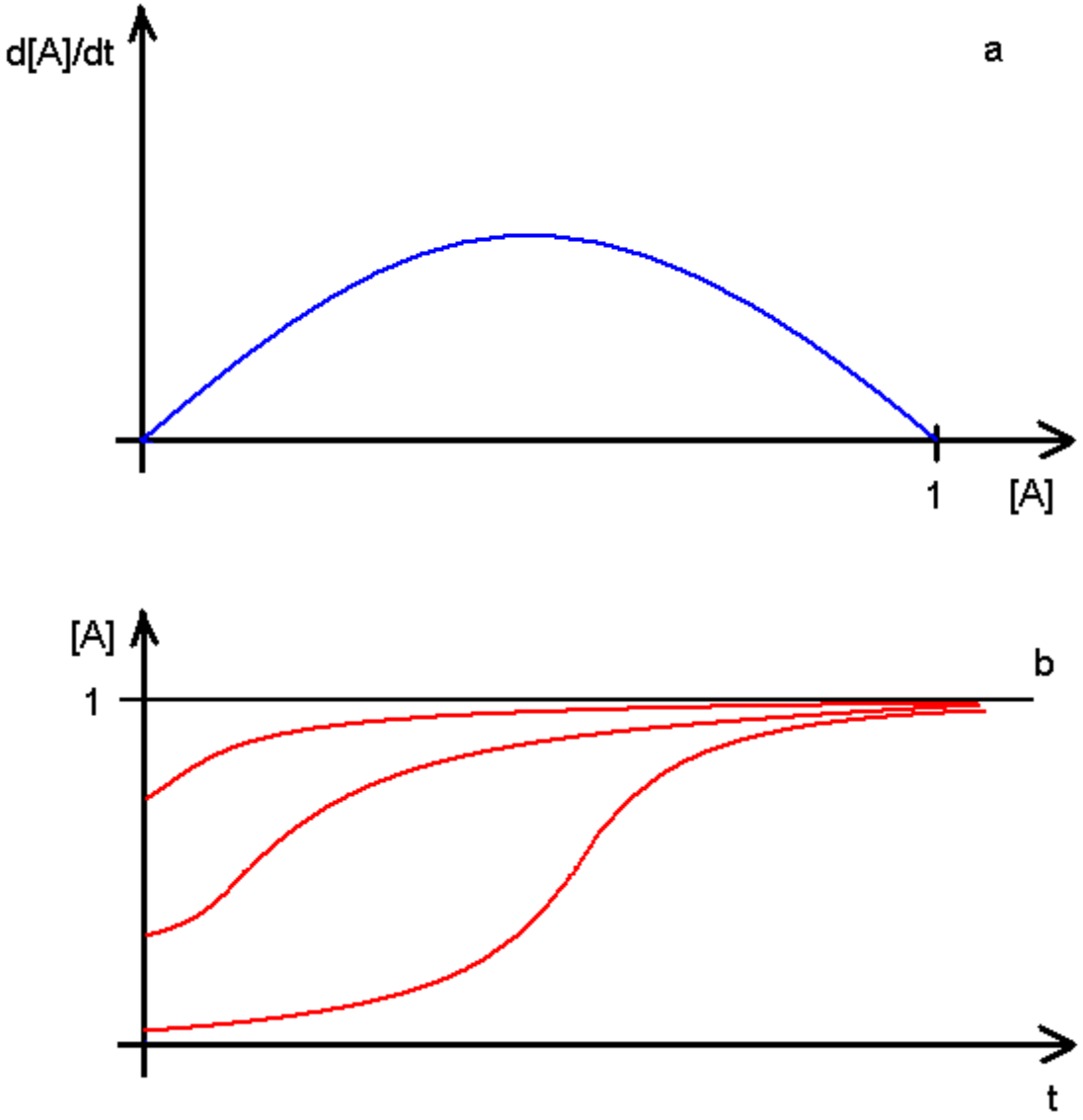


Fig. I9. Velocity of [A] as a function in its phase space with $s > 0$ (a) and the corresponding dynamics of [A] for $s > 0$ as movements within the same phase space (b).

6) Hydrogen reacts with oxygen to form water: $2H_2 + O_2 \rightarrow 2H_2O$. Within a closed volume, if the reactants are gaseous and well-mixed, they are described by two variables, $[H_2]$ and $[O_2]$. The rate of this reaction generally increases with the two parameters, temperature and pressure, although the exact pattern is rather complex. In contrast to λ or s in the previous two examples, here both the parameters change in the course of a reaction, so the distinction between variables and parameters is blurred. Still, stoichiometry dictates that $[H_2]$ always declines two times faster than $[O_2]$ (Fig. I10).

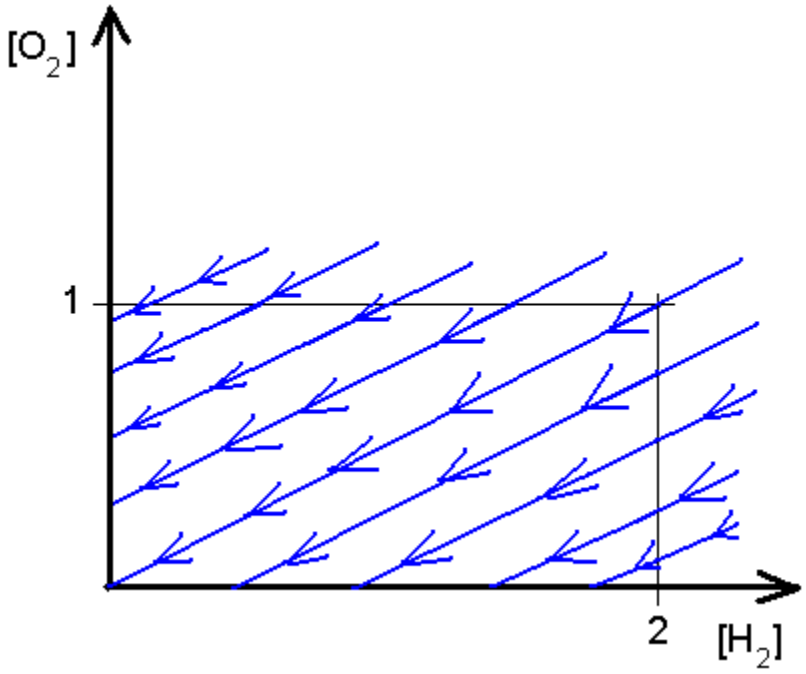


Fig. I10. Dynamics of reaction $2H_2 + O_2 \rightarrow 2H_2O$. Joint changes of $[H_2]$ and $[O_2]$ (in moles) are shown by arrows in the $[H_2]$x$[O_2]$ phase space.

7) The discrete phase space of chess consists of all possible positions of 16 or less White pieces and of 16 or less Black pieces on the 8x8 chessboard. This space can be thought of as 64-dimensional (for each square, we specify which piece, if any, occupies it) and the number of possible positions is astonishingly high. A particular game is the succession of simple jumps, each corresponding to a legal move, within this phase space. Usually, each side can chose from many possible moves. Some positions, however, can never emerge in any game (Fig. I11).

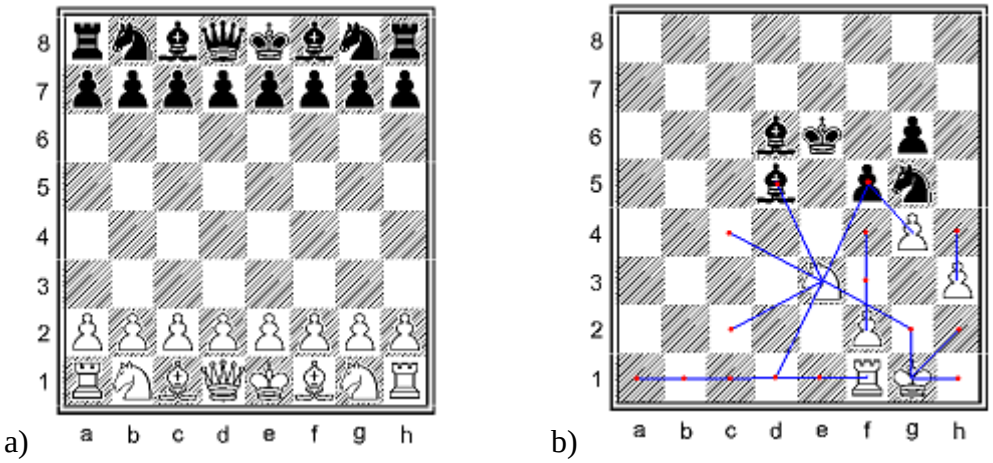


Fig. I11. The initial chess position (a). All legal moves by White, in a particular position (b). An impossible position - a King or a Queen cannot move before a pawn on the same side moved (c).

8) An English text is the sequence of symbols from the Latin alphabet. Obviously, the number of all such sequences of any substantial length exceeds even the number of chess positions. However, only a tiny fraction of all sequences are gems (*e. g.*, Hamlet), and a vast majority are nonsense. Being creative, one can invent a meaningful text "from scratch", *i. e.*, jump into an arbitrary point of the phase space of texts. Editing a text, however, usually involves only small-scale movements within this space: substitutions of individual letters, adding, removing, or reshuffling individual words, etc. (Fig. I12).

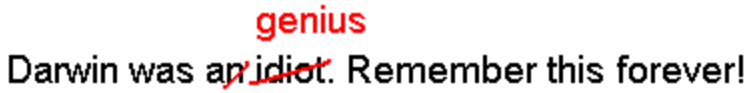


Fig. I12. A mistyped English text with the necessary corrections.

In examples 4), 5), and 6) precise knowledge of the current state of a changing object makes it possible to exactly predict its future. In contrast, we generally cannot predict the next move in chess, unless a computer is playing and we know the program (7), or the next editorial change from the current text (8), because the they are affected by unpredictable decision-making of a player or an editor. The notion that we can predict the future from good enough data on the present is called determinism, and theories, models, and analyses based on this notion are called deterministic. Deterministic models are much simpler than those which allow for a variety of outcomes, and for a while we will simplistically treat evolution as a deterministic process.

4. Space of genotypes

The right approach to thinking about evolution of life is to visualize it as movements of evolving biological objects within suitable phase spaces. Biological objects can be described in a variety of ways, depending on what problem is studied. However, the fundamental phase space of evolution of life is the space of genotypes. Although the genotype of an organism does not determine all its features exactly, no environment will produce a bear from a zygote with the genotype of an elephant. Continuity of lineages of descent is ensured primarily by inheritance of genotypes. Thus, evolution of life consists, above all, of movements of genotypes representing current states of evolving lineages within the space of genotypes.

Physically, the genotype of a cell may consist of one or many double-stranded DNA molecules, either linear or circular. However, for now we can ignore these details and think of a genotype as one sequence of nucleotides. Genotypes are very long, so that the space of genotypes has very many dimensions - as many as the length L of the longest genotype we need to consider. If you find it hard to visualize L-dimensional spaces when $L > 3$, you are not alone (Fig. I13).

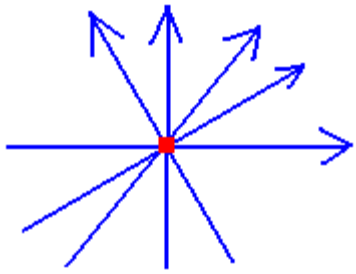


Fig. I13. Five axes, each perpendicular to the other four, pass through every point of a 5-dimensional Euclidian space.

In the simplest case, the i-th dimension tells us what nucleotide is present in the i-th position within the genotype. There are only five possibilities (5-letter alphabet): A, T, G, C, and, say, "-" (if a genotype is shorter than i and so does not have the i-th nucleotide). Thus, the space of genotypes is discrete. Numerical values may be assigned to these letters arbitrarily (*e. g.*, A - 1, T - 2, G - 3, C - 4, and "-" - 0), because there is no natural order within the set nucleotides.

Lengths and other properties of genomes of several species are presented in Table 1 (some other genomes are even longer but consist mostly of junk DNA). "Genotype" and "genome" have similar meanings, but genotype is the property of an organism, and genome is a property of a species as the whole. Thus, Alex Kondrashov has a genotype (in fact, two haploid genotypes) but *Homo sapiens* has a genome, the averaged (consensus) genotype of all humans. Most of the complete genomes known so far are, in fact, individual genotypes. Since within-species variation is crucial for Darwinian evolution, we need to consider the space of genotypes, and not of genomes.

Table 1. Parameters of some well-studied genomes.

	Length	# of protein-coding genes	% coding
Mycoplasma genitalium	0.65	650	90
Haemophilus influenzae	1.8	1800	85
Bacillus subtilis	4.3	4000	85
Saccharomyces cerevisiae	14	7500	70
Caenorhabditis elegans	100	20000	30
Drosophila melanogaster	160	14000	30
Arabidopsis thaliana	70	30000	30
Fugu rubripens	400	20000	10
Gallus gallus	2000	20000	2

Homo sapiens	3200	20000	1

Words cannot describe the volume of the space of genotypes, *i. e.* the total number of possible genotypes. Just for genotypes of length L = 1,000,000, their total number is $4^{1,000,000}$: there are 4 genotypes of length 1 (A, T, G, and C), $4^2 = 16$ genotypes of length 2 (AA, AT, AG, AC, TA, TT, TG, TC, GA, GT, GG, GC, CA, CC, CG, and CC - we can choose the first and the second nucleotides independently, each in 4 ways), $4^3 = 64$ genotypes of length 3, etc.

To call the number $4^{1,000,000} \approx 10^{600,000}$ astronomical would be an understatement, because it is incomparably larger than, say, the total number of electrons in the visible Universe ("only" $\sim 10^{80}$). As it is the case for English texts, only a tiny fraction of all possible DNA sequences can serve as genotypes of viable organisms, and the vast majority of sequences are garbage. Still, even the number of possible viable genotypes (and of meaningful books) must be much larger than the number of species (and of writers or even readers) which have actually existed or will ever exist - on Earth or elsewhere in the Universe as we know it.

In contrast to its enormous volume, diameter of the space of genotypes, the maximal distance within it, is rather moderate. Indeed, it takes replacing no more than L nucleotides to transform any genotype of length $\leq$ L into any other such genotype. In other words, the shortest path connecting even the most dissimilar pairs from $>4^{1,000,000}$ genotypes with $L \leq 10^6$ consists of no more than 10^6 smallest steps. A modest diameter of the space of genotype is one of the properties of nature which make evolution feasible: at the very least, it must be possible to arrive to a human genotype from something simple after only a sane number of small steps (Fig. I14).

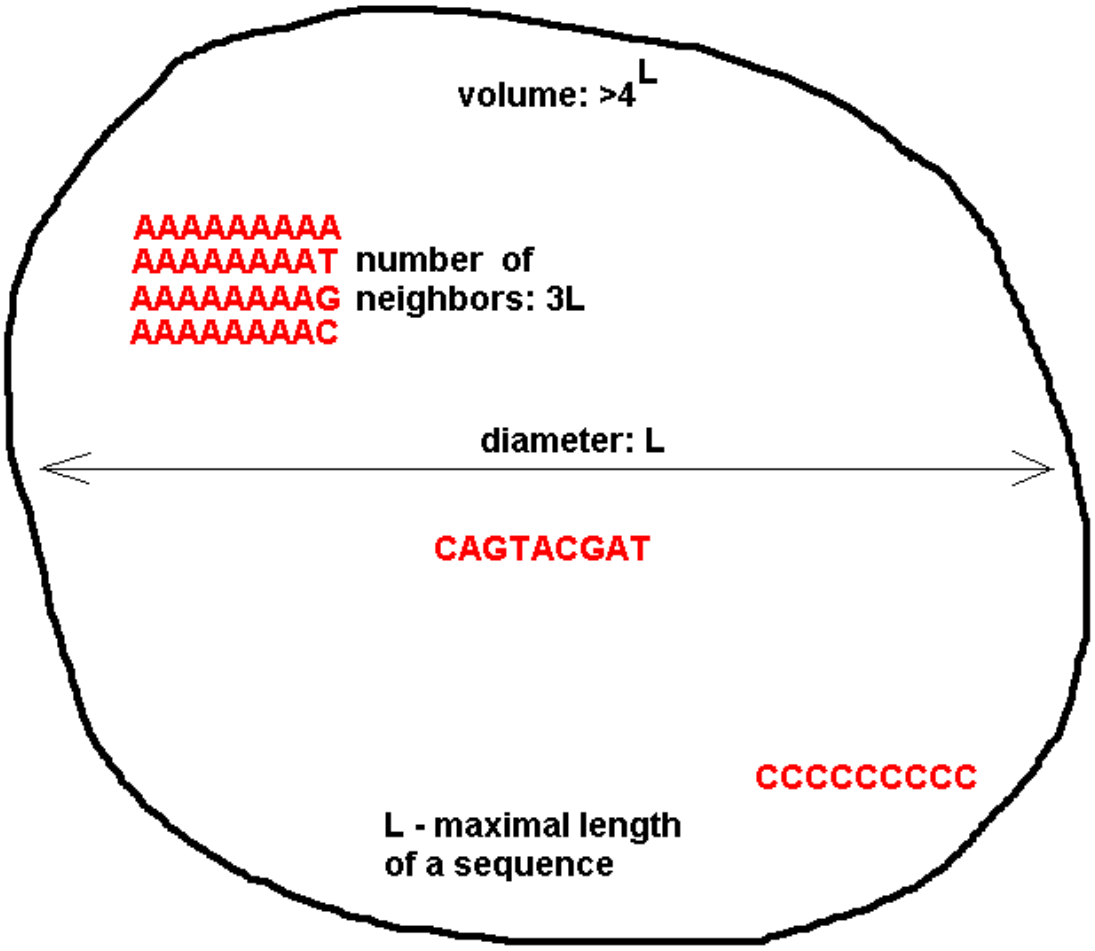


Fig. I14. The space of genotypes.

This huge volume/diameter ratio is due to a high number of dimensions of the space of genotypes. Even within the 3-dimensional space we call home, a lot of sand grains can be packet into a cube with the side of 1m, much more than what fits into a 1m long line. High number of dimensions also causes each genotype to have many neighbors: a single-nucleotide substitution can convert a genotype of length L into any of its 3L nearest neighbors.

In practice, evolution is rarely studied at the level of whole genotypes. Even when such studies are performed, for example, when the evolution of the genotype length is considered, the degree of detail does not reach to individual nucleotides. Thus, the enormous space of whole genotypes has almost no direct applications. Instead, we will encounter a variety of smaller, although still huge, subspaces of this space. Still, keep in

mind that evolution of life is, ultimately, the movement of evolving lineages within the space of all possible genotypes.

5. Levels of organization

Biology studies objects and processes at a staggering range of space and time scales, although physics has to deal with even wider ranges. A tiny protein molecule can catalyze thousands of reactions per second, while a vast tundra ecosystem changes rather slowly. If information on every constituent molecule were necessary to understand an ecosystem, ecology would be impossible. Fortunately, this is not so because life has a hierarchical structure, consisting of several levels of organization. Sequences, molecules, cells, (multicellular) organisms, populations, and ecosystems are the basic levels, but sometimes it is convenient to add extra intermediate levels such as primary, secondary, and tertiary structures of macromolecules, organelles within cells, tissues and organs within multicellular organisms, and societies of organisms within some populations.

The concept of levels of organization reflects one fundamental fact: interactions between some biological objects of similar nature are much stronger than between others. Some of the molecules which constitute a human body - those which belong to the same cell - interact so tightly that together they behave as one unit. In contrast, interactions between molecules from different cells are much weaker. Thus, it is productive to think of an organism as consisting of cells, instead of molecules. Similarly, a population can be viewed as consisting of organisms, instead of cells, because cells of different organisms interact with each other only indirectly, as parts of interacting organisms (Fig. I15). In fact, when populations are considered, organisms are often referred to as individuals, because we are not concerned with their internal "organization" and, instead, treat each one as an "indivisible" entity. In the same way, an ecosystem can be understood as a set of interacting populations.

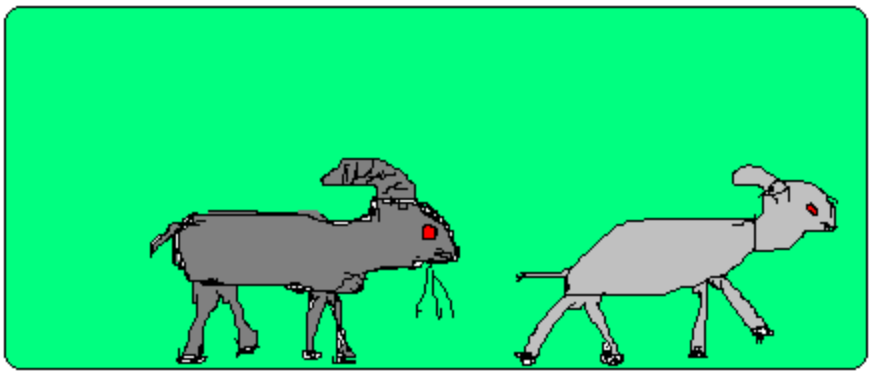


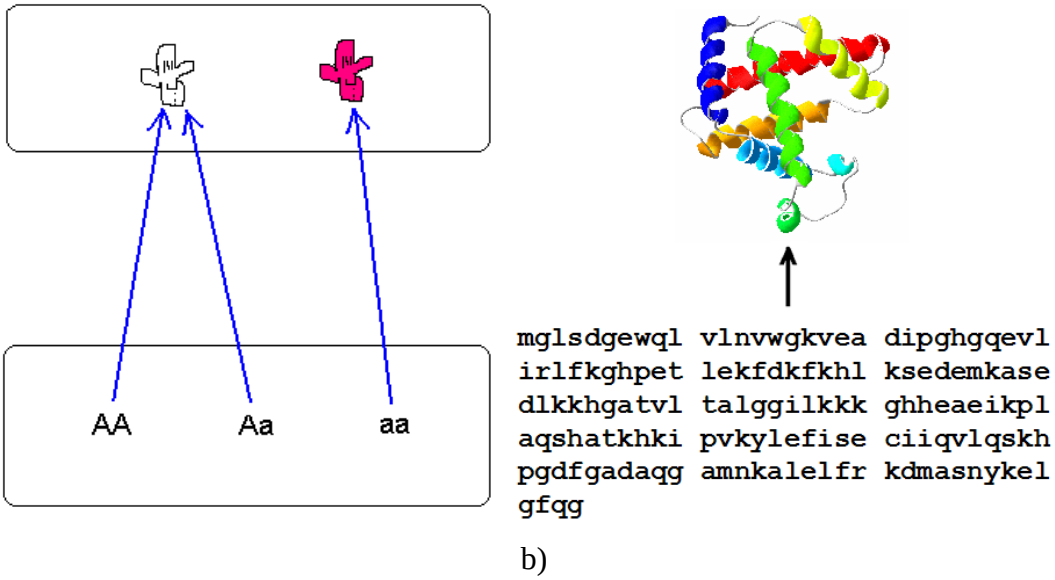
Fig. I15. Interaction of molecules and cells which constitute a dominant goat with those constituting a subordinate goat.

Boundaries between objects at a particular level of organization are to some extent arbitrary: there may be occasional covalent bonds between "different" molecules, cells can merge into syncytia, organisms can form colonies, populations can be parts of metapopulations, and ecosystems are often separated by gradual ecotones. Still, all these objects are real, in the sense that considering them helps to understand how life works.

Strictly speaking, only cells and multicellular organisms are "life". Study of macromolecules and their sequences became a part of biology in XX century, after their importance for understanding cells and organisms had been recognized. The superorganismal levels, populations and ecosystems, were incorporated into biology at approximately the same time, as a result of Darwinian revolution. We will see that, in order to understand evolution by natural selection, populations must be viewed not just as bunches of individuals but as objects with their own unique properties. Indeed, organisms (as well as molecules and cells) are short-lived and evolve only in the sense that organisms which live at different times are not the same. Only potentially immortal lineages, represented at each moment by populations, as well as ecosystems, can evolve in the literal sense of the word.

Because almost all biological information is passed from generation to generation in the form of DNA sequences, life constantly has to reconstitute itself from bottom to top: proteins fold into 3D structures mostly determined by their sequences, cellular organelles self-assemble from molecules, a child develops from one cell, etc. Consequently, states of higher-level objects can be viewed as functions (in the

mathematical sense!) of states of lower-level objects. Such functions can be visualized by their graphs, maps from phase spaces of lower-level objects into those of higher-level objects. Since the work of Gregor Mendel, published in 1866, studying such maps is one of the central tasks of genetics (Fig. I16).



a) b)

Fig. I16. Genotype -> phenotype maps. a) A map from the space of partial genotypes into the space of flower colors in the case of complete dominance. b) The sequence and the 3D structure of human myoglobin, a part of the map from the space of protein sequences into the space of their conformations.

Maps which connect different levels of organization also depend on the external conditions: even identical twins can be substantially different from each other. Thus, we have to deal with families of maps, with each map describing what happens under a particular environment. Even under a given environment, different outcomes at the higher levels are still possible due to unpredictable, random factors, referred to as developmental noise. Thus, the outcome which corresponds to a particular genotype under a particular environment is, in fact, the average of many possible outcomes.

If all the properties of any higher-level object could be deduced from the sufficiently complete information about the underlying levels, the whole biology would eventually be reduced to studies of DNA sequences. There is no agreement on whether this can be done in principle. For example, some people think that, ultimately, our

behavior is determined by laws of physics, while others believe that we have free will. At the moment, however, the reductionist program is certainly not very successful, even where it must eventually be. We still cannot deduce the 3D structure of a protein from its sequence, and do not know which of the differences between very similar genomes of humans and chimpanzees are responsible for us being much smarter than them. Thus, evolution at each of the higher levels will not be reduced to the evolution of genotypes any time soon, and different levels often have to be studied more or less separately.

6. Phenotypes and traits

The word phenotype can be applied to any description of any biological object at any level of organization. However, DNA sequence-level phenotypes are traditionally referred to as genotypes, and descriptions of populations and ecosystems are rarely called phenotypes. We will mostly follow this tradition.

Since complete phenotypes are very complex, usually we have to consider only partial phenotypes: the sequence of a particular intron, the conformation of a particular protein, the network describing mutual influences of only a subset of all genes, the arrangement of bones within a wing, or the age at first mating. The word "partial" can be omitted when this should not lead to confusion. Often, it is convenient to further subdivide even partial phenotypes into traits (characters, features, properties). A trait can assume different states, *e. g.*, the color of a flower can be yellow, red, etc. Thus, every trait is associated with the phase space, consisting all its possible states, and the phase space of a phenotype is a combination (direct product, mathematically speaking) of the phase spaces of all its constituent traits.

A phenotype (and a genotype) can be subdivided into parts and traits in many ways, and it is important to chose the most convenient subdivision. Partial phenotypes and traits should be defined in such a way that they:

1) Work within the organism as independently as possible.

2) Vary between the organisms as independently as possible.

3) Provide the shortest description of differences between the available organisms.

4) Make sure that descriptions of parents and offspring are not too dissimilar.

The last requirement is particularly important for evolutionary biology. Indeed, evolutionary thinking depends on the assumption that parents and offspring are similar, *i. e*. that no long leaps are possible within the phase space of the evolving object. Of course, if anything could begot anything (Fig. I2), no definition of traits would prevent such leaps. However, this is not the case, and we only need to define the traits properly. Thus, we do not want two phenotypes to be distant from each other if one of them can be easily transformed into the other. These considerations may dictate the minimal number of dimensions of the space of phenotypes we want to consider.

To illustrate these requirements, let us consider one of the simplest possible phenotypes, the amino acid sequence of a protein. Fig. I17 presents the sequences of preproinsulin, precursor of insulin, encoded by human, chimp, dog, pig, and cow genomes.

```
Hs   1  MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED 60
Pt   1  MALWMRLLPLLVLLALWGPDPASAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED 60
Cf   1  MALWMRLLPLLALLALWAPAPTRAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVED 60
Ss   1  MALWTRLLPLLALLALWAPAPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREAEN 60
Bt   1  MALWTRLRPLLALLALWPPPPARAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEG 60

Hs  61 LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN 110
Pt  61 LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN 110
Cf  61 LQVRDVELAGAPGEGGLQPLALEGALQKRGIVEQCCTSICSLYQLENYCN 110
Ss  61 PQAGAVELGGGLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN-- 108
Ss  61 PQAGAVELGGG--LGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN 108
Bt  61 PQVGALELAGGPGAGGLEGPPQKRGIVEQCCASVCSLYQLENYCN----- 105
Bt  61 PQVGALELAGGPGAGGL-----EGPPQKRGIVEQCCASVCSLYQLENYCN 105
```

Fig. I17. Amino acid sequences of preproinsulin from *Homo sapiens*, *Pan troglodytes*, *Canis familiaris*, *Sus scrofa*, and *Bos taurus*. Amino acids in non-human sequences which differ from the corresponding human amino acids are in red. The sequences can be considered "as is" or after being aligned to each other (shown in italics).

The first three sequences each contain 110 amino acids. However, if we define, as we did before, 110 traits with 21 possible states which are "the amino acid which

occupies the i-th position of the protein or "-" if the protein is too short", we get a mess at the end of pig and cow sequences. In contrast to all the other sequence regions, the distal parts of pig and cow proteins appear to be totally dissimilar from each other and from the remaining sequences. The remedy is to align our sequences, by introducing some gaps, in a way which, roughly speaking, minimizes their differences. Now, the i-th trait is the content of the i-th column of the alignment, and gaps can appear anywhere, not only at the ends of sequences.

In fact, we are killing several birds with this stone. First, the aligned amino acids perform essentially the same tasks within their respective proteins. Second, all the sequences became very similar after alignment. Third, an insertion or a deletion of one or several successive amino acids can no longer produce a radically "new" sequence and create an illusion of a long jump within the sequence space. Insertions and deletions are less common than substitutions of individual nucleotides and amino acids, but they do happen - a deletion of 6 nucleotides has been fixed in the pig lineage, and another deletion of 15 nucleotides nearby has been fixed in the cow lineage.

Later, we will encounter much more complex, multidimensional higher-level phenotypes. Still, subdividing phenotypes into traits will always be based on simple considerations outlined here. Success of an evolutionary analysis often depends on the proper recognition of traits.

7. Fitness and adaptation

Life persists, and evolves along the way, only because individually short-lived organisms reproduce themselves. Thus, fitness, a trait of the whole phenotype of any organism (individual) which characterizes its reproductive success, is of key importance. Fitness describes the contribution of an individual to the next generation and reflects its viability, probability of finding a mate (if reproduction is sexual), fecundity (age-dependent), and longevity. Here we can simplistically think of fitness as the total number of offspring produces by the individual during its lifetime.

In fact, this number represents the absolute fitness. In the long run, the mean absolute fitness of individuals in any lineage must be very close to 1 (without sex) or 2 (with sex). Indeed, the lineage will go extinct if the mean absolute fitness is lower, and would multiply without a limit if it is higher, which, of course, is impossible.

Temporarily, the mean absolute fitness within a lineage can increase, for example, when a new ecological niche becomes available, or decline, when the lineage experiences a drop in its number. Still, density-dependent regulation of the number of individuals causes their absolute fitnesses to decline when the number of competitors increases, and thus makes the long-term (geometric) mean absolute fitness within a lineage very close to 1 (or 2). Thus, propagation of a lineage depends on absolute fitness.

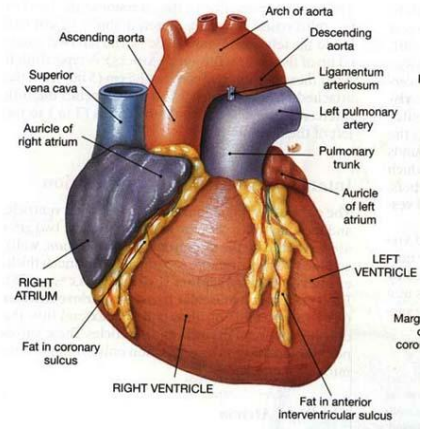
In contrast, evolution of a lineage depends only on relative fitnesses of individuals. Indeed, if every individual starts producing five times more (or seven times less) offspring, changes of frequencies of individuals of different kinds within the lineage will remain unaffected. Thus, we can normalize absolute fitnesses of all individuals by the same quantity, for example, by the mean absolute fitness. The result is relative fitnesses: instead of saying that an individual produces 32 offspring, we say that it produces, say, 103% of the mean number of offspring.

As it is the case for any trait, fitnesses of even identical twins are not necessarily identical. Still, we can think of the expected, under a particular environment, fitness of an organism with a particular genotype (or phenotype), and call it the fitness of this genotype (phenotype). At the very least, an evolving lineage must always be represented by genotypes and phenotypes with non-zero fitnesses, which is a stringent requirement, since the vast majority of possible genotypes would be lethal.

Fitness reflects the efficiency of just one process, reproduction, performed by an organism as the whole. Processes performed by a part of the organism which facilitate reproduction are called its functions, in the biological sense of the word. For example, the function of heart is to pump blood, and not too produce subtle noises (although a physician can put them to good use), and the function of hemoglobin is to carry oxygen, and not to preferentially reflect red light (except when it colors skin and thus produces social signals). If the process is, indeed, a function, disrupting it will adversely affects fitness. Still, some parasitic processes may be unavoidable due to physical constraints: a noiseless heart is impossible.

States of traits and phenotypes which are involved in performing functions are called adaptations. The degree of adaptation of a particular phenotype characterizes its ability to perform its functions and, thus, its impact on fitness. There exists an infinite variety of specific functions and of the corresponding adaptations. Some adaptations are

fairly general - a lipid bilayer membrane which separates "inside" from "outside" is necessary for every cell, and a heart, although not necessary for all life, is useful in a wide variety of environments. In contrast, other adaptations facilitate reproduction (and, thus, deserve the name adaptation) only under very specific environments, and, thus, can be considered as adaptations only to these environments (Fig. I18). The rayed pattern is adaptive only in one locality, lower Huallaga, when it has been adopted by many unpalatable species, causing predators to rapidly learn to avoid all rayed butterflies and moths.



a)

b)

Fig. I18. a) Heart, an example of a rather general adaptation. b) An example of a very specific adaptation, provided by a case of Mullerian mimicry, mutualistic resemblance

between several unpalatable species. On the top line are *Heliconius melpomene* (left) and *H. erato* (right) from the Rios Maya and upper Huallaga. Their pattern switches to join a "rayed" mimicry ring in the lower Huallaga: left - *H. melpomene*, *H. elevatus*, *H. demeter*, centre - *Laparus doris*, *Neruda aoede*, *Eueides tales* and a pericopine moth; righ - *H. erato*, *H. burneyi*, and *H. xanthocles*. © James Mallet 1997 (http://www.ucl.ac.uk/taxome/jim/Mim/helicon.htm).

Reproduction can be viewed as the sole function of the organism as a whole, with all the trait states which facilitate reproduction being adaptations. It may be possible to define functions and adaptations of parts of the organism relative to other whole-organism processes, less inclusive that reproduction. For example, carrying oxygen can be considered the function of mule's hemoglobin since it facilitates the organism's survival, despite its sterility and zero fitness. However, fitness is of paramount importance for evolution, and we will mostly evaluate a processes by its impact on reproduction.

8. Mutation and variation

"Everything consisting of parts crumbles ..." were the last Buddha's words. This is fully applicable to rather fragile DNA molecules. Thus, it is not surprising that DNA sequences are routinely altered. First, errors occasionally occur during DNA replication, which takes place before cell divisions. Second, physical integrity of DNA can be violated at any moment, and occasionally is repaired incorrectly, producing physically whole DNA with a new sequence. In both cases, the outcome is called mutation.

Molecular physics informs us that to replicate or to repair DNA without errors is impossible - if the precision (fidelity) of these processes tends to infinity, so does their cost, in terms of both time and energy. Thus, mutations must happen, although their natural rates are amazingly low. The probability that a mutation involving a particular nucleotide will occur during a generation is usually only $\sim 10^{-8\text{-}9}$. For comparison, when a human types on a keyboard, an error is made, at the very least, once for $\sim 10^{4}$ letters.

A vast majority of mutations are minor and involve substitution, deletion, or insertion of one or few nucleotides (Chapter 2.2). Major mutations are rare and usually

deleterious: because viable genotypes are extremely sparse in the space of genotypes, a long leap almost always lands a genotype in the garbage. Still, deleting or inserting a long segment of useless, junk DNA may be more or less harmless, or even mildly advantageous. Adding a new DNA segment is more likely to be advantageous if it has already been subject to selection, which is the case for duplication of a pre-existing segment (perhaps long enough to contain one or more genes), polyploidization (genome-level duplication), or lateral gene transfer (acquisition of a DNA segment from a distant genome).

Even deletion or insertion of a substantial DNA segment does not lead to a very long leap within the space of genotypes, as long as distance in this space is defined intelligently, on the basis of sequence alignment and, thus, of proper recognition of sequence traits (Fig. I17). Simplistically, the distance D(a,b) between a pair of genotypes a and b can be defined as the sum of the number of mismatches plus the total number of all gaps (gap lengths may also be taken into account) in their alignment, chosen, among all their possible alignments, to minimize this sum (Fig I19).

```
AGAGGTCG--GTATCTCAGCTAGTCATACATTA
|||||  ||   ||||||  |||||   |||  |||||
AGAGGACGAAGTATCT-AGCTAAACATGCATTA
```

Fig. I19. An alignment of two DNA sequences, the distance between which is 6. Any attempt to align these sequences differently would lead to a larger distance (try this!).

Under this simple definition, the distance between human and chimpanzee genomes is ~$4x10^6$: these genomes are so similar that their alignment is unambiguous, and this alignment contains mismatches at 1.3% of ~$3x10^9$ nucleotide sites, together with a number of gaps, some of which are quite long. It can be shown that D(a,b) satisfies the three requirements necessary for a function to be a distance: 1) D(a,a) = 0 (the distance from an object to itself is 0), 2) D(a,b) = D(b,a) (the distance between two objects is the same in both directions), and 3) $D(a,c) \leq D(a,b) + D(b,c)$ (triangle inequality).

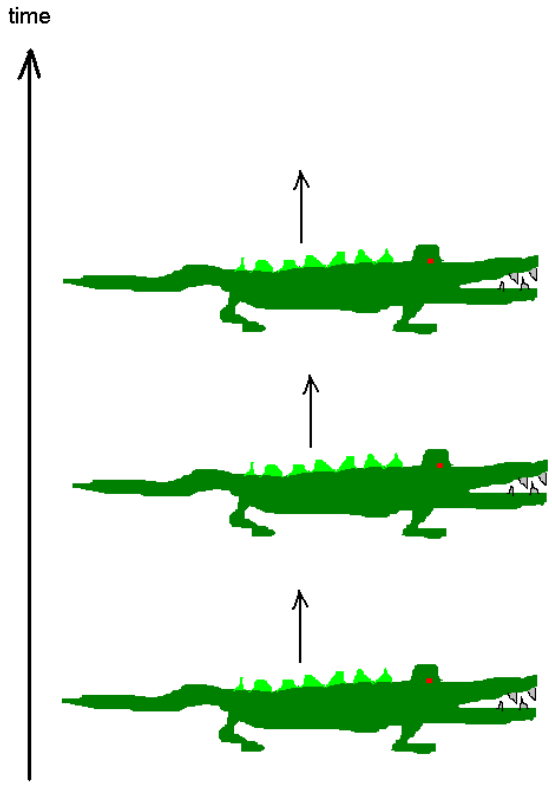
In fact, knowing the distance between a pair of sequences is not enough to tell exactly how often a mutation will convert one of them into the other one, because the

probability of mutations a -> b and b -> a may be very different. For example, if the only difference between a and b is that a has 10 extra nucleotides somewhere in the middle, the a -> b mutation is deletion of these 10 nucleotides. A deletion is generally much more probable than the reciprocal b -> a insertion, because there are 4^{10} = 1,048,576 ways of inserting 10 nucleotides into a particular place. However, this complication does not affect our main conclusion: non-lethal mutations can hardly produce very long leaps within the space of genotypes, of length comparable to distances between distinct species.

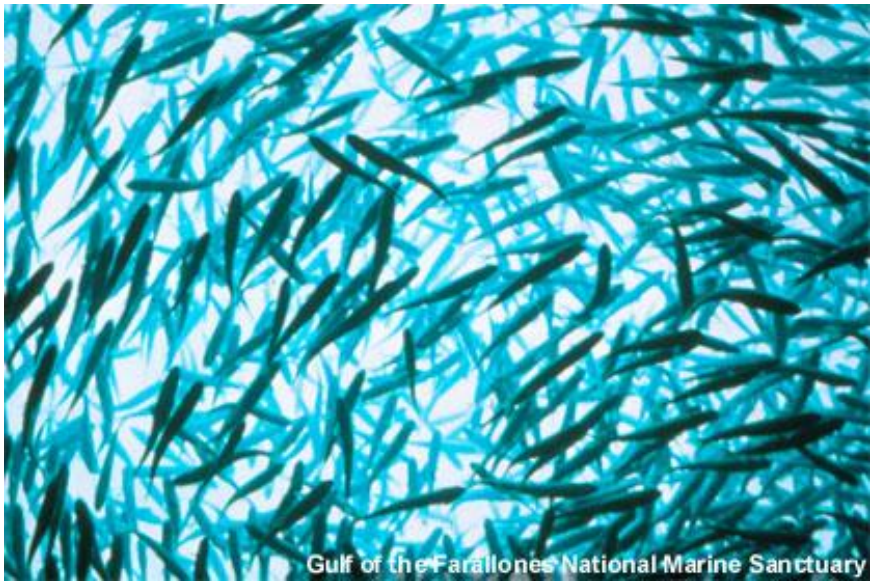
Ubiquity of mutation makes some genetic differences even between parents and offspring, as well as between siblings, unavoidable. Thus, variation is a key property of life. A vast majority of differences between similar genotypes are simple, so that alleles, the available variants of the genotype, mostly differ from each other by a single-nucleotide substitution or by presence or absence of one or several successive nucleotides. Sexual reproduction, when present, creates new combinations of alleles by reshuffling genotypes and, thus, may enhance variation. Still, mutation is the only ultimate source of new alleles, and patterns in standing genetic variation are to a large extent dictated by patterns in *de novo* mutation. Phenotypic variation, however, is also affected by the environment.

9. Population, selection, and allele replacements

We take it for granted that at any particular moment of time an evolving lineage consists of many individuals. Indeed, no real individuals are like Loch Ness monsters, with only one or very few of a kind living at a time. Instead, every individual has many peers, and a set of similar individuals living together, which compete for resources, and, if reproduction is sexual, interbreed with each other, is called population (Fig I20).



a)



b)

Fig. I20. a) Lonely, unhappy, and perhaps asexual, monsters. b) A school of jack mackerel, *Trachurus symmetricus*.

http://www.sanctuaries.nos.noaa.gov/pgallery/pgfarallones/living/living_10.html

Populations exist due to action of fundamental ecological forces (Chapter 2.1). Briefly, if a particular genotype is fit, an environment can support more than one individual of this or very similar genotypes, all occupying the same ecological niche, and density-dependent factors keep the number of such individuals approximately constant. These individuals compete for resources, so that the success of one means the defeat of the other. As the result, most of individual lineages are short-lived, but some lucky lineages proliferate. Thus, competition exerts a strong homogenizing influence, ensuring that all the competitors share relatively recent common ancestors.

Conversely, individuals which are substantially different from each other usually utilize different resources. Bats and mice, even when living side-by-side, do not compete directly: multiplication of mice will not immediately lead to elimination of bats, so that bat and mouse lineages may coexist for a long time, and accumulate a lot of extra differences. As a result, different local populations living alongside each other usually consists of individuals of distinct kinds.

By the definition of fitness, if different individuals have different fitnesses, offspring of those with higher fitnesses will be overrepresented in the next generation. This process of differential reproduction within a population ("survival of the fittest") is called natural selection, or simply selection. The outcome of selection is determined by relative fitnesses of different individuals. Even if all the individuals were genetically identical, or all their genetic differences were functionally irrelevant, different individuals would still leave different numbers of offspring, just "by chance". However, it is usually convenient to regard such differential reproduction as a separate phenomenon and to restrict the notion of selection only to differential reproduction of genotypes, *i. e*. to those situations where differences between fitnesses of individuals are caused by their different genotypes. Many genetic differences, indeed, affect fitness, but many others, such as a nucleotide replacement within a pseudogene, may be selectively neutral.

If we adopt population as the object of consideration, it immediately becomes clear that at least some kinds of natural selection are unavoidable, as long as there are any functionally important differences between the present genotypes. Indeed, the winner in a heat will be, by definition, an athlete running at the highest average speed, and the

winner in a race to reproduce itself is the genotype which does this faster than the others. Population is the arena which ensured that this ruthless race never ends (Fig. I21).

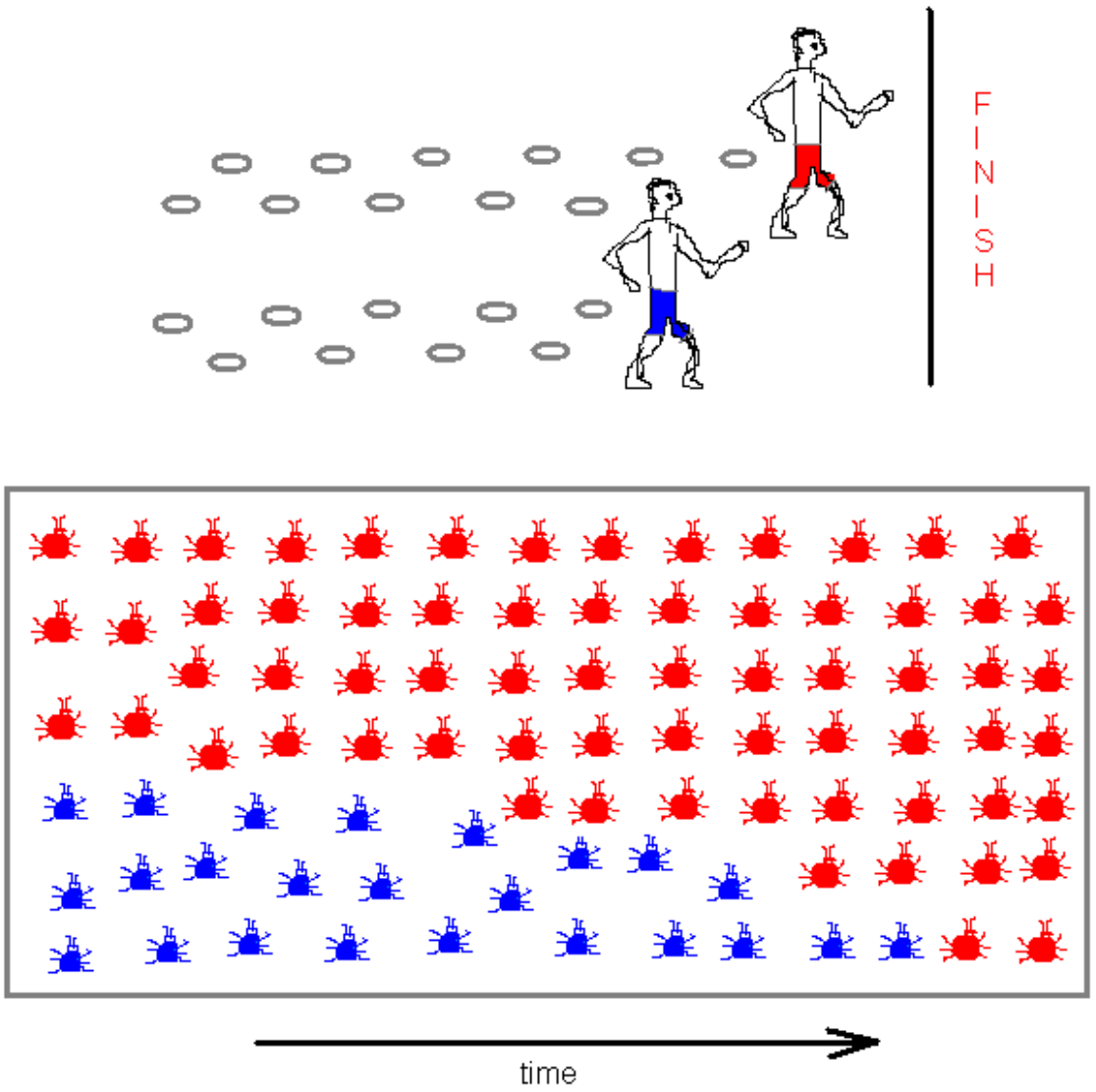


Fig. I21. Red bugs reproduce faster than blue bugs, so that natural selection favors the allele which makes the bugs red, which will eventually replace the allele which makes the bugs blue.

Genetic investigation of any population, including human, always reveals a lot of selectively important genetic variation. Individually rare unconditionally deleterious alleles, constantly eliminated by negative (purifying) selection which favors the currently

common alleles and thus maintains the *status quo*, are particularly common. New advantageous alleles are much less common but are essential for evolution. Selection favoring such, initially rare, alleles changes the *status quo* and is called positive (Darwinian) selection. The key process responsible for evolution at the level of genotypes is called allele replacement and consists of simple replacement, within the population, of an old allele with a new, superior allele. Eventually, individuals carrying even a slightly less fit allele will be completely replaced by ecologically equivalent individuals carrying a more fit allele (Fig. I9). Positive selection-driven allele replacements occasionally occur within every population and may be used to define long-term boundaries of populations.

10. Evolution according to Lamarck and to Darwin

We have already seen that substantial changes of living beings can occur only through slow, gradual evolution. Genetically, an organism cannot supply its offspring with a genotype too different from its own. Physiologically, there must be continuity between parents and offspring - a bear's nucleus planted into an elephant's egg would die. Thus, movements of the small spot which represents a variable population within the space of genotypes at a particular moment can hardly involve any long instant leaps.

However, the total number of possible mutations, even only minor, of a genotype is still huge, because every nucleotide is mutable. What determines which changes will occur in an evolving lineage? Remember that modern organisms possess amazingly complex adaptations, often fine-tuned to match sublime features of their environments (Fig. I18). If evolution, indeed, produced these organisms, it must be adaptive, *i. e.* must be able to choose, among the innumerable variety of mostly unfit genotypes, those genotypes which correspond to phenotypes possessing such adaptations. Thus, only two really different answers to our question are possible:

1) Mutation proceeds in such a way as to increase fitness and, thus, sets the course of adaptive evolution. An organism somehow knows what changes of its genotype will be good and mutates it accordingly, before passing it to its offspring. The desirable direction of mutation is either determined by some intrinsic process or is dictated by the environment, if an organism acquires traits as direct response to external challenges (*e.*

g., strong muscles as the result of exercise) and then changes its genotype so that its offspring will have these traits from the very beginning. Thus, the course of evolution is determined between generations and selection is irrelevant and may even be absent. This is the modern interpretation of the mechanism of evolution proposed by Lamarck.

2) Mutation proceeds regardless of its impact on fitness and does not set the course of adaptive evolution. Instead, this course is determined by selection, which compares fitnesses of different genotypes, although extreme rarity of some mutations may restrict what selection can actually accomplish. Thus, the course of adaptive evolution is determined during a generation. A mutation affecting fitness is usually deleterious: changing an already fit genotype regardless of the consequences is very risky, because most genotypes are unfit (think of fixing a car by hitting it indiscriminately with a hammer). Indeed, mutations, like editorial corrections (Fig. I12), are mostly minor, but unlike such corrections, they are also mostly senseless. Still, rare beneficial mutations are retained and multiplied by selection. This is the modern interpretation of the mechanism of evolution proposed by Darwin and Alfred Russell Wallace.

Obviously, the Lamarckian mechanism of evolution would be more efficient. Instead of producing a lot of variants, almost all doomed, as the Darwinian mechanism requires, it pushes a genotype straight into the right direction (Fig. I22). However, organisms are unable to deliberately modify their genotypes and the Lamarckian mechanism is never (or almost never) responsible for adaptive evolution. In contrast, evolution of functionless DNA regions and of other selectively neutral phenotypes is mostly mutation-directed: it cannot be selection-directed because selection does not care.

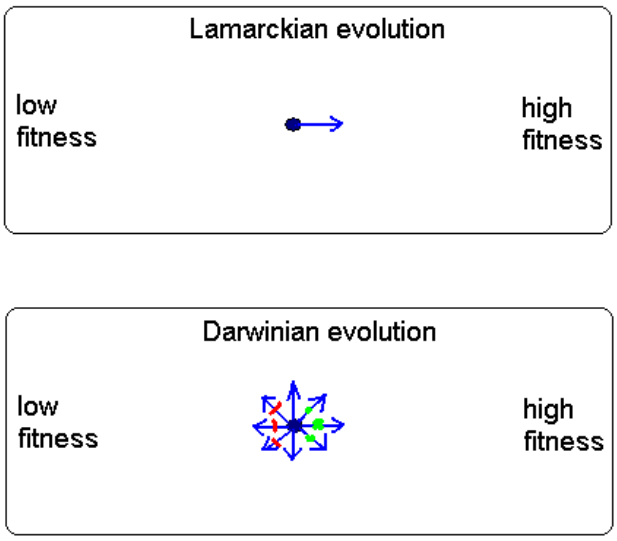


Fig. I22. Lamarckian and Darwinian evolution. Mutation is shown by blue arrows, negative selection by red lines, and positive selection - by green dots.

Lamarckian evolution by adaptive mutation would be an organism-level process, and could improve even lone monsters (Fig. I20). In contrast, Darwinian evolution by natural selection is a population-level process which can occur only within a genetically variable population. Naturally, the direction of Darwinian evolution is determined only by fitnesses of the genotypes that are currently present in the population, as selection knows nothing about fitnesses of absent genotypes. Because the range of within-population variation is relatively narrow, Darwinian evolution is not only gradual and slow, but also greedy, using a term from computer science: it has no foresight and just maximizes the immediate gain in fitness.

One may wonder how this greedy Darwinian evolution can, nevertheless, produce good results. We do not know the exact answer to this question, but we do know that some greedy algorithms work surprisingly well. For example, democracy is inherently greedy, as it is hard for a politician to care about problems which will emerge only after the next elections. Still, democracy works better than autocracy, although a king is supposed to care about long term-interests of his dynasty and subjects (Fig. I23).

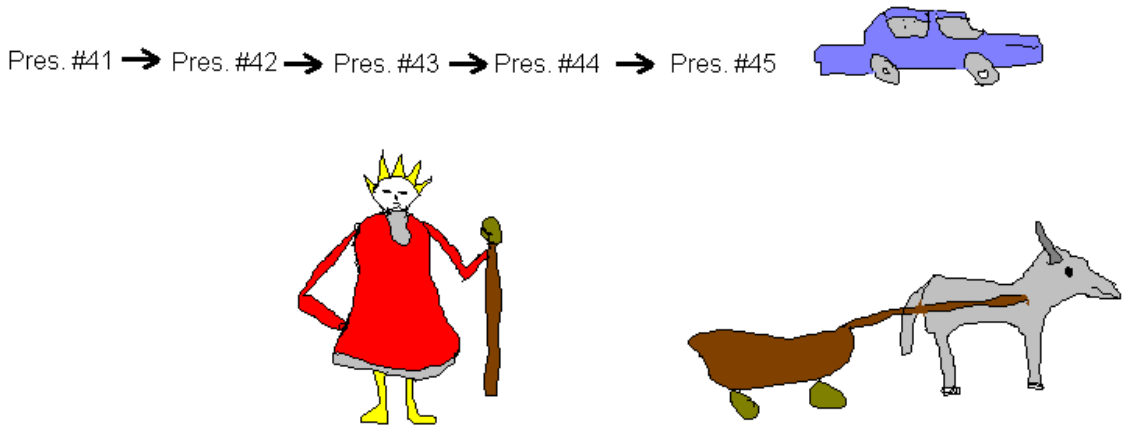


Fig. I23. Some greedy algorithms are reasonably efficient.

11. Fitness landscapes

In the long run, the course of evolution of a lineage may depend on fitnesses of all possible genotypes and phenotypes, because any genotype may eventually appear in the lineage. Thus, to understand evolution, we need to consider fitness as a function in a suitable phase space. The graph of such a function is called fitness landscape (adaptive topography, adaptive landscape, phenotype → fitness map). Naturally, fitness is always non-negative. However, sometimes it is more convenient to consider the logarithm of fitness, which can accept any values (Fig. I24). Together, the fitness landscape and the currently present variation determine how selection acts within the population.

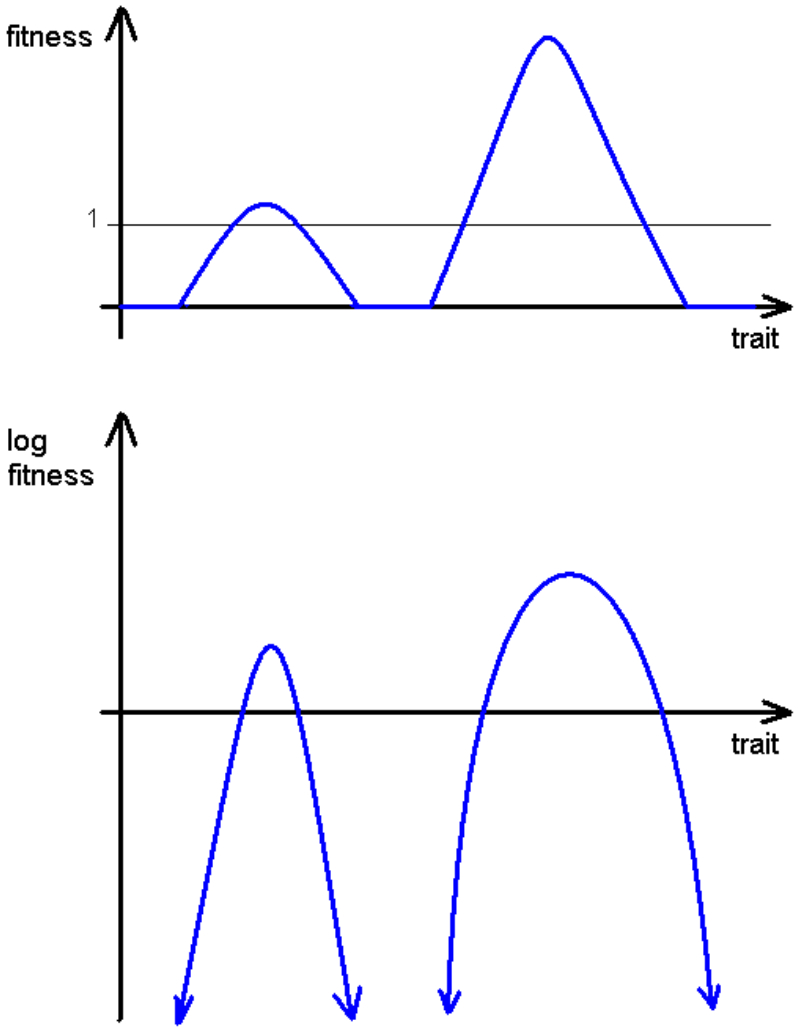


Fig. I24. Fitness (top) and its logarithm (bottom) as a function in 1-dimensional space of phenotypes.

Let us first consider the basic properties of a fitness landscape in a continuous phase space. Although the space of genotypes is discrete, the does not matter much, as long as similar genotypes usually have similar fitnesses. Any generic "good" function of one variable has a number of local maxima (peaks), separated by local minima (Fig. I25). The first derivative of the function is zero at either a maximum or a minimum, so that the tangent to the fitness landscape is horizontal at all such points, while the second derivative is negative at a maximum and positive at a minimum. A local maximum (peak) on a fitness landscape is surrounded by its slopes, which constitute its domain of attraction, the portion of the phase space from which a population will reach the peak by

always climbing upwards, along the gradient of the fitness landscape. Thus, a lineage will climb the peak in whose domain of attraction it was located initially, and stay on it forever, even if many other peaks are higher. In other words, a local fitness maximum is a stable equilibrium of evolution. Local minima are also equilibria, since a population which is located exactly at such a point will not evolve. However, these equilibria are unstable, as an evolving population will not approach a minimum from its near vicinity, but will move away and, eventually, reach a maximum.

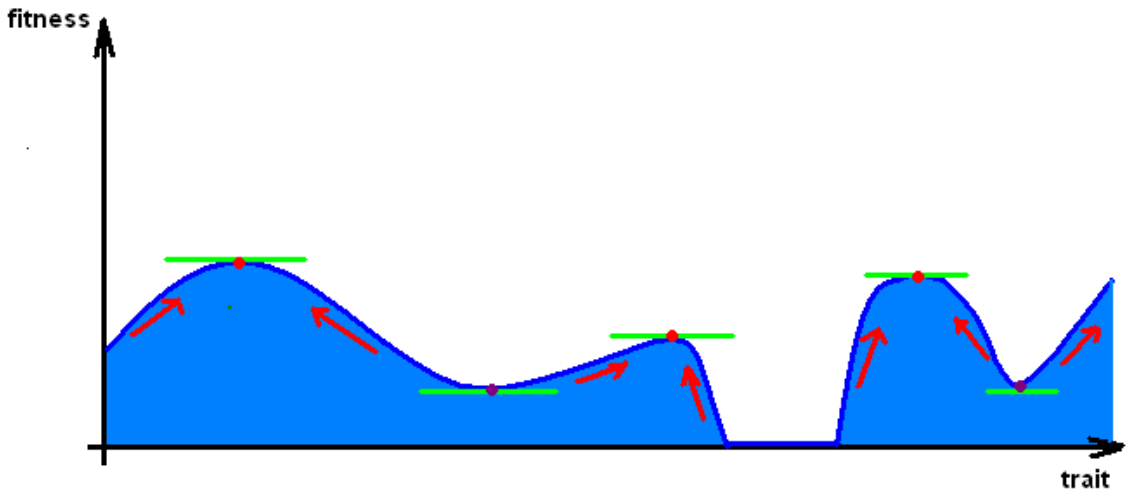


Fig. I25. A generic fitness landscape over 1-dimensional phase spaces, with three maxima (red dots) and two minima (purple dots).

In two dimensions, we also encounter fitness peaks and their domains of attraction, now separated by 1-dimensional boundaries, instead of minima (Fig. I26). On such boundaries there may be unstable equilibria, which can now be of two kinds: local minima (a population runs away from the minimum in all directions and saddle points (a population approaches the saddle point from one direction but runs away from the other). Nothing qualitatively new happens when we consider phase spaces of more than two dimensions, although drawing fitness landscapes over them becomes impossible.

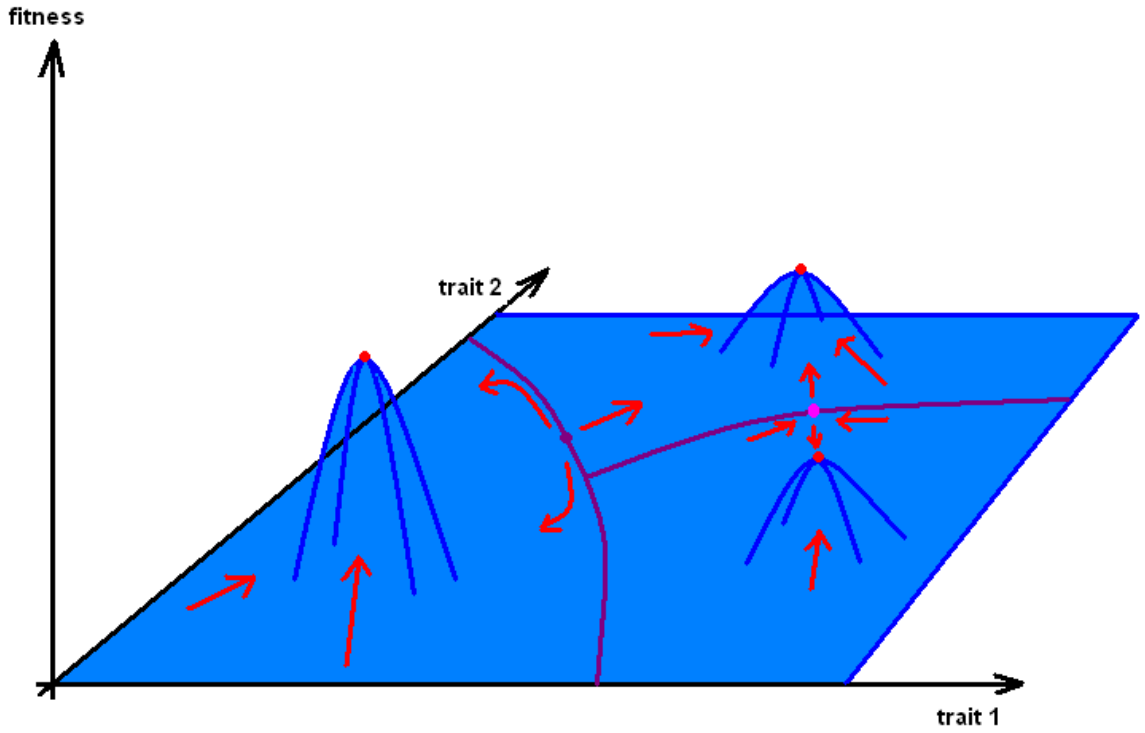


Fig. I26. A generic fitness landscape over 2-dimensional phase spaces, separated into domains of attraction of three maxima (red), with a minimum (purple) and a saddle point (pink).

Thus, the basic properties of fitness landscapes predict that greedy Darwinian evolution must be subject to historical constraints. Indeed, such constraints, imposed by initial conditions, are ubiquitous in the evolution of life. For example, sharks, extinct ichthyosaurs, and toothed whales are much more similar to their closer relatives, rays, lizards, and cows, respectively, than to each other, despite their external similarity, forced on them by their common adaptation to fast swimming (Fig. I27). In contrast to sharks, ichthyosaurs and whales depend on air for breezing, a property inherited from their terrestrial ancestors, which is probably suboptimal for a marine animal.

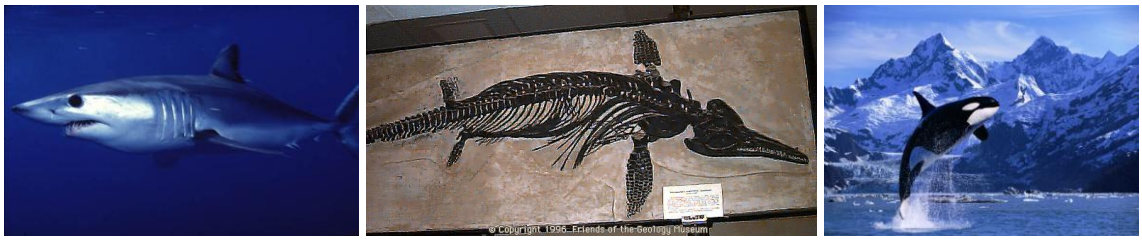


Fig. I27. Shortfin mako shark *Isurus oxyrinchus*, a cartilaginous fish (http://www.flmnh.ufl.edu/fish/Sharks/mako/mako.htm), top; Jurassic ichthyosaur *Stenopterygius quadricissus*, a reptile (http://www.geology.wisc.edu/~museum/old/ichthyosaur.html), middle; killer whale *Orcinus orca*, a mammal (http://www.wildthingsphotography.com/adm/photo/27_D00090.jpg), bottom.

Still, something must be missing from the simple reasoning outlined above. First, fitness peaks on a generic fitness landscape are all of different heights, but the mean absolute fitness of any lineage must be very close to 1 (or 2). Second, according to the Weak Claim, evolution keeps going even now, although 3.5 billion years may seem to be enough for every lineage to reach some fitness peak. Third, according to the Strong Claim, a lineage sometimes splits into branches, instead of always climbing up as a coherent unit on the near-by fitness peak.

All these inconsistencies can be resolved if we take into account an obvious fact that fitness landscapes are not invariant, but depend on the environmental conditions, both biotic and abiotic. In other words, we have to consider not just one fitness landscape, but a family of them, each corresponding to the one of many possible sets of environmental conditions. Then, if the height of an occupied fitness peak exceeds one, the size of the population which occupies it increases, eventually causing the peak to become lower, until it approaches the equilibrium height of one. One can say that the fitness landscape bends locally under the weight of the population. In contrast, unoccupied peaks can be of any height, if we define the height of an unoccupied peak as the absolute fitness of individuals with the corresponding genotype, when their number is low (Fig. I28). If a peak has the height below 1 even when unoccupied, it may be unable to support the population, because the height of the peak generally diminishes when the size of the population on it increases. In any case, it makes no sense to compare, on an invariant fitness landscape, those organisms whose densities are regulated independently. Thus, slowly reproducing elephants can evolve, by natural selection, from rapidly reproducing prokaryotes.

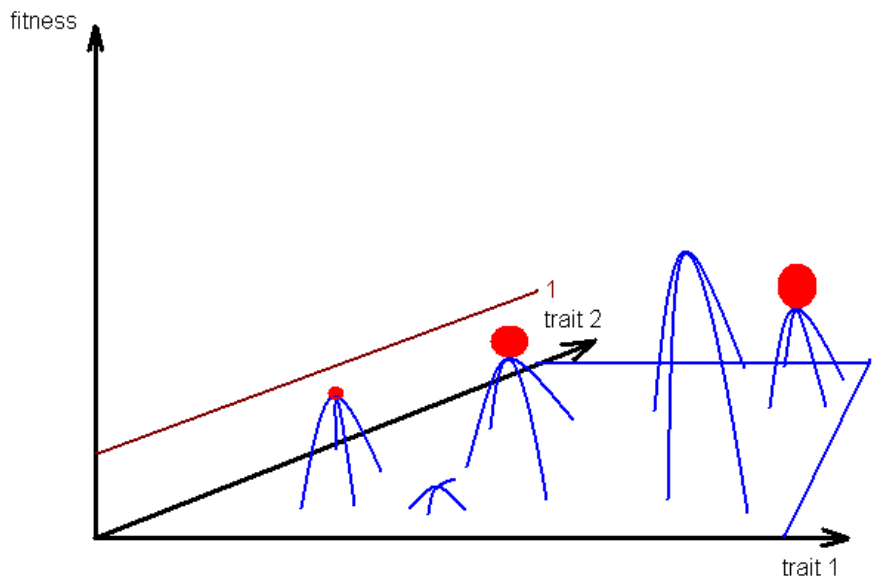


Fig. I28. All the occupied peaks (populations on them are shown by red balls) are of the absolute height 1, and unoccupied peaks can be of any height.

Moreover, the environment may affect not only the heights, but also the locations of peaks. This can lead to never-ending evolution if, in the simplest case, the peak which is occupied by a population is constantly moving slowly, due to incessant changes of the environment. Then, the population will follow, with some lag (Fig. I29).

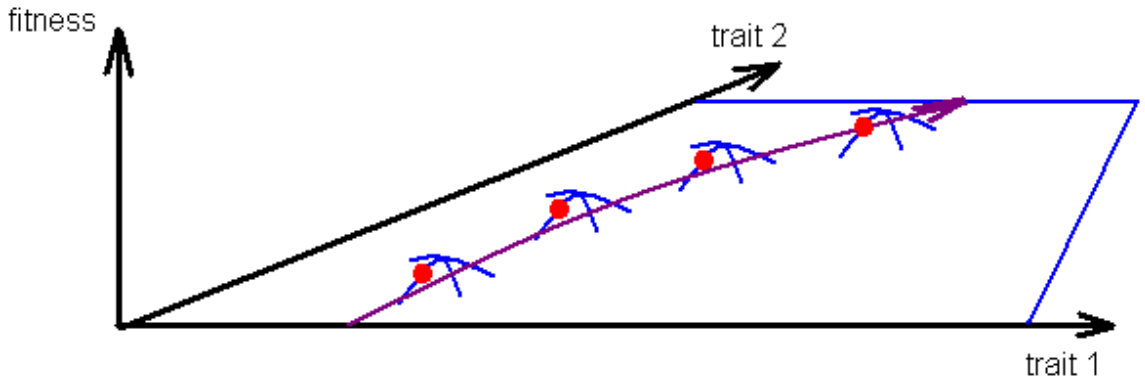


Fig. I29. The population rides a continuously moving peak. The purple line shows the trajectory of the peak.

Finally, if two populations occupying the same position within the space of genotypes experience different fitness landscapes, their independent evolution can proceed along completely different paths, leading to cladogenesis. This can happen, for example, if the two populations live in different locations under different environments (Fig. I30).

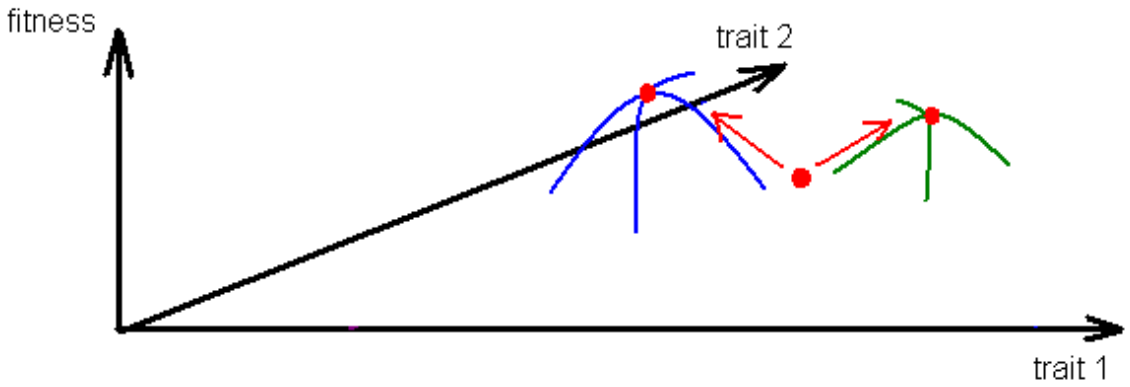


Fig. I30. The blue peak exists under one environment, and the green peak exists under another environment, so that the population can end up on either of them.

However, the family of all possible fitness landscapes - the paramount arena for evolution - is a rather complex object, and several of its simplified representations may be useful. First, we may consider only fitness peaks, and plot their locations under all the possible environments. When the environment, and the fitness landscape, changes slightly, a peak usually responds by moving a little. However, occasionally a peak can disappear altogether or appear out of nothing. Such abrupt responses to small changes are called bifurcations, and birth or death of a peak is a relatively common bifurcation - mathematically speaking, it has codimension 1. Still rarer (due to a codimension 2 bifurcation), one peak splits into two, or two peaks merge into one. Thus, a simplified representation of a family of similar fitness landscapes which shows only locations of peaks consists of finite segments of curves, with rare branching points (Fig. I31). Of course, a fitness landscape, in contrast to an evolving lineage, may also experience a drastic change, with a lot of peaks appearing and disappearing at once. This happens

when the environment changes suddenly, and leads to rapid extinction of lineages which occupied lost peaks, followed by slow colonization of some of the new peaks.

Fig. I31. Locations of peaks in a family of similar fitness landscapes which depend on two or more parameters of the environment.

Alternatively, we can ignore all quantitative differences between fitnesses, and be concerned only with whether a genotype or a phenotype has a positive or zero fitness. This approach makes sense, since the vast majority of possible genotypes are lethal. Then, we can show either the set of all fit genotypes for one particular fitness landscape or the set of all genotypes which are fit under at least one of all possible environments (Fig. I32).

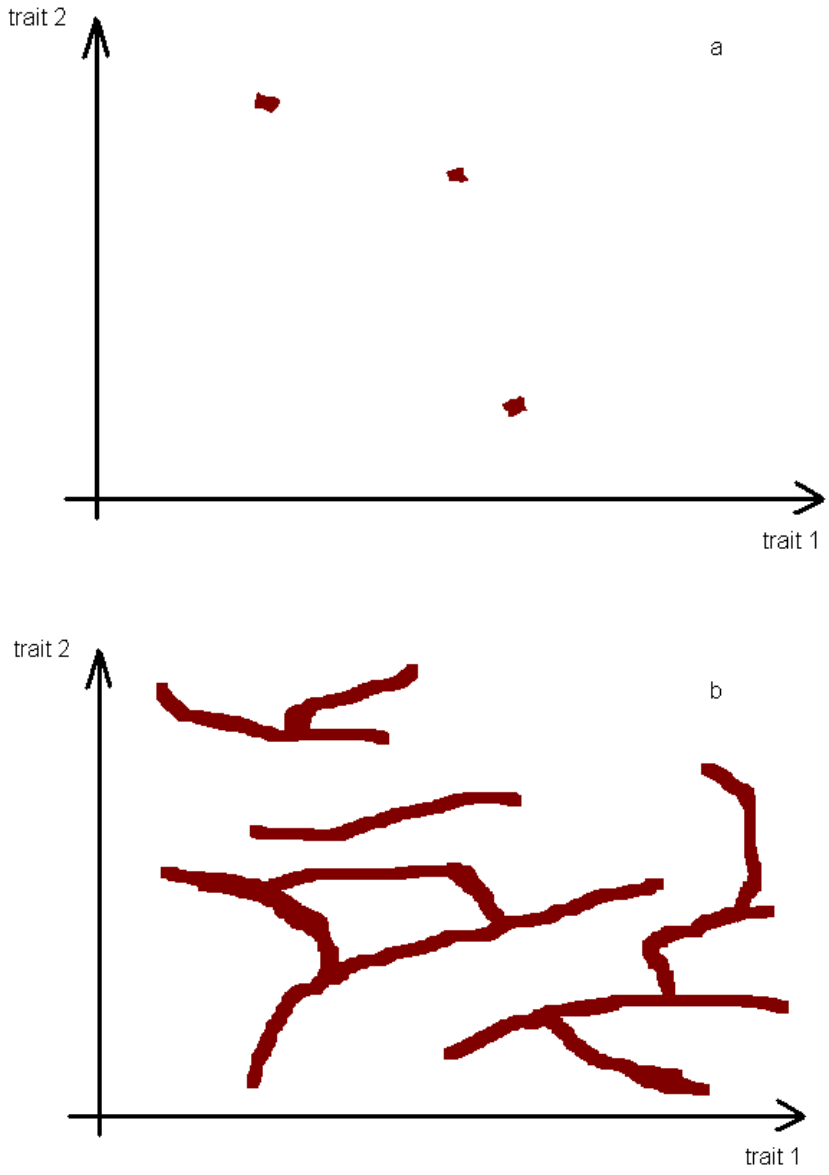


Fig. I32. Genotypes which have non-zero fitness under a particular environment (a) or under at least one of many possible environments (b).

Naturally, the former set is always a subset of the later one. In contrast to the locations of peaks, which are arranged in 1-dimensional curves (Fig. I30), fit genotypes generally occupy areas with the full number of dimensions; however, such areas must be mostly thin and rare. Finally, we may associate with each genotype its highest fitness, attained under the most favorable environment. This obviously makes sense if we want to simultaneously consider birds and fishes, which require different environments.

Because complete genotypes or phenotypes of individuals are almost never known, in practice we mostly have to consider fitness as a function of a partial phenotype. Such limited analysis can still be meaningful, as long as the partial phenotype affects fitness more or less independently of the rest of the phenotype. For example, a child homozygous for a loss-of-function allele at a locus encoding enzyme beta-hexosaminidase A always dies from Tay-Sachs disease before the age of 10. Thus, if we collectively denote all such alleles by a, and normal alleles by A, the corresponding fitness landscape is, approximately, AA -> 1, Aa -> 1, and aa -> 0.

Naturally, if two or more loci or phenotypic traits strongly interact with each other while affecting fitness, they must be considered together. For example, considering fitness as a function of only the length of a fish makes little sense, since other traits also determine whether it will be able to swim fast. Sometimes it is convenient to consider composite genotype -> phenotype -> fitness maps, which simultaneously take into account several levels of organization.

The chief obstacle to studying fitness landscapes is that a vast majority of genotypes do not exist, and their fitnesses remain unknown. One can think of immense, highly complex fitness landscapes mostly hidden in the darkness of non-existence, with only tiny spots on them visible, each illuminated by existence of the set of similar genotypes which constitute a population (Fig. I33). Fortunately, even these spots can tell us something. In particular, unviability of hybrids between fit but dissimilar genotypes demonstrate that fitness landscapes are strongly nonlinear, and mostly consist of regions of zero fitness, separating rare regions of high fitness. Still, we currently have very few means to study continuous paths of high fitness connecting distant genotypes.

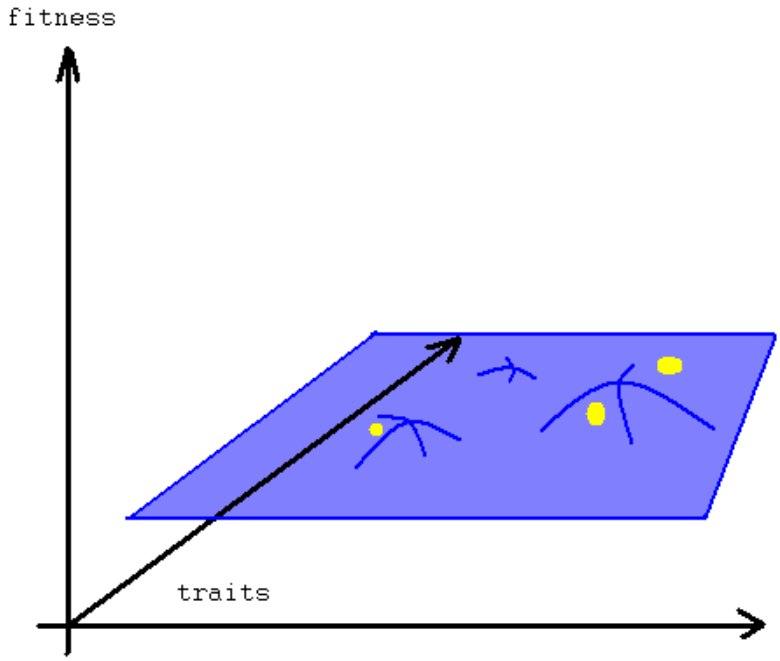


Fig. I33. Existing biodiversity illuminates only rare, small spots on the fitness landscape.

Consideration of fitness landscapes over properly chosen spaces of genotypes or phenotypes constitutes the conceptual foundation of evolutionary biology. Presentation of indirect evidence for evolution in Chapter 1.1 will be based on fitness landscapes. In Chapters 2.1 the small-scale properties of fitness landscapes, at the scale of within-population variation, will by described and applied to the analysis of natural selection. In Chapter 2.6, medium-scale properties of fitness landscapes will be considered in the context of speciation. Finally, Chapter 3.2 will treat large-scale properties of fitness landscapes as the conceptual foundation of the theory of Macroevolution. Here, we will rely on fitness landscapes to introduce several concept which provide the basis for dealing with the two key wonders of life - diversity and complexity.

12. Similarity, relatedness, compatibility, and connectedness

Diversity of life - today (Fig. I1) or at any moment during the last 3 billion years - is striking. To comprehend the structure of this diversity, we first need to introduce four concepts that are applicable to just pairs of individuals: similarity, relatedness, compatibility, and connectedness.

Similarity of two individuals is simply the inverse of the distance between their genotypes. This distance can be defined in a variety of ways, but, under any reasonable definition, two genotypes which can be transformed into each other by a small number of nucleotide substitutions, insertions, deletions, etc. are similar. Similarity of phenotypes can also be defined sensibly.

Relatedness of two individuals can be defined as the time which lapsed since their last common ancestor was living. In contrast to similarity, relatedness cannot be unequivocally deduced just from the genotypes or phenotypes. Instead, we need to know the evolutionary history of the individuals.

Compatibility of two individuals describes whether genotypes consisting of assortments of the corresponding segments from their genotypes would result in fit individuals. Indeed, two perfectly fit individuals can, nevertheless, be incompatible to each other: their crosses can either fail or produce unviable, weak, or sterile hybrids (Fig. I34). While any genotype can tolerate one or several foreign genes, as natural lateral gene transfer and artificial genetic modifications demonstrate, more or less equal mixtures of segments from two distant enough genotypes are always unfit. Almost all data on compatibility were obtained using an excellent natural tool for assaying it, sexual reproduction. Still, the notion of compatibility is equally applicable to organisms reproducing asexually, although studying it is difficult in this case. Incompatibilities can be hard, if hybrids are unfit under any environment, or soft, if they may be fit only under some suitable environments.

Fig. I34. Mules, the products of crosses between mares (*Equus caballus* females) and jackasses (*Equus asinus* males) are viable and vigorous but sterile http://www.lovelongears.com/mulepix.html

Inviability of hybrids between two species is not a reason to reject the Strong Claim for them: only if the space of genotypes were unidiminsional (Fig. I25), ancestors of modern horses or asses must have passed, in the course of their divergence from the common ancestor, through the mule stage, a clear impossibility. In contrast, in the space of two or more dimensions, twisty paths of high fitness which circumvent unfit hybrids may connect two fit but incompatible genotypes (Fig. I32).

Incompatibilities are pervasive even when we consider only partial genotypes or phenotypes. For example, in humans a protein alpha synuclein contains alanine at its 53 site, and a mutation which replaces this alanine with threonine causes early-onset Parkinson disease. However, normal alpha synucleins of New World monkeys, as well as of many other mammals, contain threonine at this site (Fig. I35). Thus, the contents of site 53 of New World monkeys alpha synuclein would cause harm if substituted into the human alpha synuclein, due to its incompatibility either to other sites of the human alpha synuclein or, more likely, to other human proteins.

```
Hs 1   MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKEGVVHGVATVAEKTK 60
Ag 1   MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKEGVVHGVTTVAEKTK 60

Hs 61  EQVTNVGGAVVTGVTAVAQKTVEGAGSIAAATGFVKKDQLGKNEEGAPQEGILEDMPVDP 120
Ag 61  EQVTSVGGAVVTGVTAVAQKTVEGAGNIAAATGFVKKDHSGKSEEGAPQEGILEDMPVDP 120

Hs 121 DNEAYEMPSEEGYQDYEPEA 140
Ag 121 DNEAYEMPSEEGYQDYEPEA 140
```

Fig. I35. Alignment of normal alpha synuclein sequences from human (top) and from black-handed spider monkey *Ateles geoffroyi* (bottom), with threonine which would cause harm in humans shown in red and other differences of the protein from this New World monkey from the human protein shown in blue.

Connectedness of two individuals describes whether their genotypes and phenotypes are connected, within the corresponding phase space, by a continuous chain of genotypes and phenotypes of other currently living individuals, such that adjacent individuals within this chain are very similar to each other. As long as the Strong Claim is correct for all life, every two individuals must be connected if we take into account both present and past intermediate individuals (ignoring occasional leaps due to instant polyploidization and symbiogenesis). Even if we consider only extant individuals, moderately dissimilar genotypes are occasionally connected (Fig. I36). However, very dissimilar genotypes are always disconnected and belong to distinct clusters of genotypes.

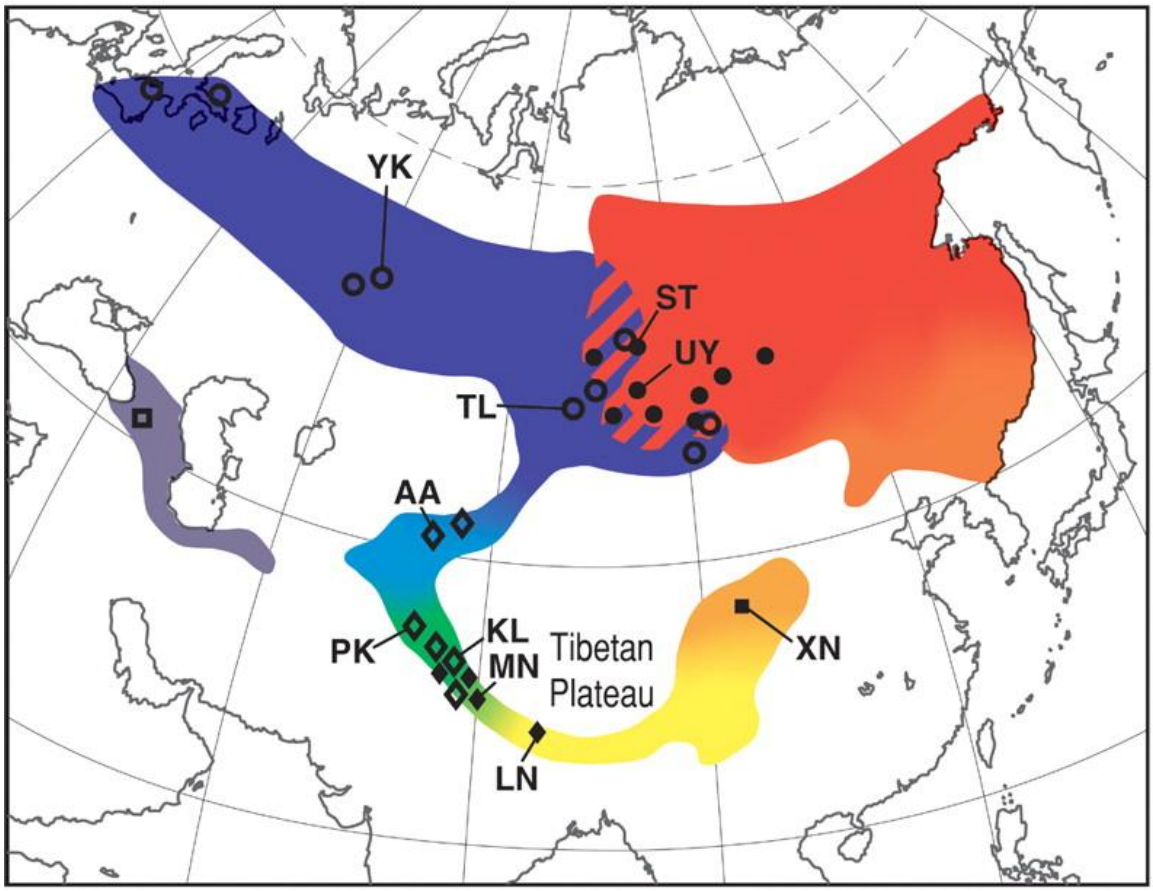


Fig. I36. A ring of forms of greenish warbler, *Phylloscopus trochiloides*. The two extreme forms, *P. trhochiloides viridans* (blue) and *P. trochiloides plumbeitarsus* (red) live together in Central Siberia without much interbreeding. Color denotes position of birds living at a particular location within the space of genotypes. Thus, these two forms

are connected, in this space, by a continuous succession of intermediate forms, whose ranges go around the Tibetan Plateau. A geographical break in this continuum is likely due to recent habitat destruction (Science 307, 414 – 416, 2005).

Two processes are responsible for the lack of connectedness between descendants of the same common ancestor. First, some intermediate genotypes may go extinct, without leaving any offspring. Second, and more importantly, intermediate genotypes may be simply abandoned by evolving lineages. For example, *Australopithecus afarensis*, in all probability our direct ancestor, who lived 3.5 Mya (Chapter 1.4) did not become extinct - it evolved into us - but its genotype has been abandoned and is no longer present in any extant organism.

How are these four concepts related to each other? Naturally, individuals with identical or very similar genotypes must share a recent common ancestor and are always compatible and connected to each other. However, things are not that simple for substantially dissimilar organisms.

First of all, the degrees of similarity can be very different at different levels of organization (Fig. I37). Salient differences between external phenotypes of different human populations are caused by only ~0.0002 interpopulation dissimilarity at the level of genotypes (defined as the fraction of mismatches in the sequence alignment). The genomes of humans and chimpanzees, despite their striking morphological differences, are only 0.013 dissimilar, while the genomes of *Drosophila melanogaster* and *D. simulans*, which are barely distinguishable phenotypically, are ~0.12 dissimilar.


01047139 © BRUCE DAVIDSON / naturepl.com

a)


Copyright 1998 Frantisek Staud

b)

c)

Fig. I37. a) Mbuti man (Ituri forest, Democratic Republic of Congo, adult man height is ~140cm, http://www.naturepl.com/photokeywords1.html) and Finns (Finland, http://www.centralfinland.net/gallery.htm, adult man height is ~175 cm). b) *Homo sapiens* (http://phototravels.net/japan/pcd2453/girl-kitano-79.html) and *Pan troglodytes* (http://www.primates.com/chimps/chimp.htm). c) *Drosophila melanogaster* (left) and *Drosophila simulans* (right) (myweb.uiowa.edu/bballard/Dsimulans.htm; the left individual is a female on both photos).

Because independent evolution of different lineages mostly leads to their divergence, less similar organisms must generally be less tightly related. However, the similarity-relatedness correlation is not absolute, because the rate of evolution is not uniform - indeed, there is no reason for its uniformity. Variability of the rate of evolution of phenotypes is particularly striking: for example, a hippopotamus is more tightly related to a dolphin than to a pig. Genomes generally evolve with more uniform rates. Still, for example, genomes of humans and dogs are more similar to each other than genomes of humans and mice, but the last common ancestor of humans and mice lived after the last common ancestor of humans and dogs. This discrepancy arose because the lineage which led to modern rodents evolved unusually fast, due to their short generation time, after its divergence from the lineage which led to modern primates (Fig. I38).

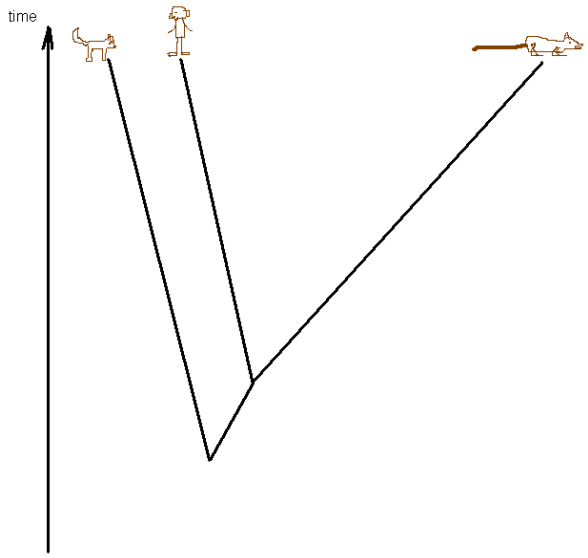


Fig. I38. Humans are more similar to dogs, at least at the level of genotypes, but are more tightly related to mice.

Correlation between similarity and relatedness could be improved if relatedness is measured by the number of generations since the last common ancestor (instead of the number of years), but it still would not be perfect, as the rate of divergence is not exactly constant even per generation (Fig. I39).

Fig. I39. A number of populations of geese are each others closest relatives. Most of these populations have been attributed to one species, *Branta canadensis* (Canada goose, left) www.chesapeakebay.net/images/COM_canada.jpg
However, one population, living in Hawaii, became so distinct morphologically that it has been described as a different species, *B. sandvicensis* (Hawaiian goose, or nene, center) www.hear.org/starr/hibirds/images/600max/starr_010809_0063.htm

Yet another population, from Northern Eurasia, has been described as one more species, *B. leucopsis* (Barnacle goose, right)
www.avesphoto.com/website/pictures/GOSBAR-1.jpg
Phylogenetically, both Hawaiian goose and Barnacle goose are nested within Canada goose, in the sense that some populations of Canada goose are more tightly related (share a more recent last common ancestor) to Hawaiian goose or Barnacle goose than to other populations attributed to Canada goose (PNAS 99, 1399, 2002).

Analogously, there is no perfect correlation between similarity and compatibility, or relatedness and compatibility. Crosses between genotypes with similarity above 99% usually produce viable and fertile offspring, and substantially more dissimilar genotypes are incompatible, but this pattern is rather fuzzy (Chapter 1.5). Some species which had the last common ancestor well over 10 Mya produce fertile hybrids (*e. g.*, many rhododendrons, Fig. I40), but *D. melanogaster* and *D. simulans*, which diverged <3Mya, are almost completely reproductively isolated from each other.

Fig. I40. *Rhododendron catawbiense* (Eastern North America, left), *R. fortunei* (China, center), and a complex hybrid "Bali" (right), which has both these species in its ancestry.
http://www.rhododendron.org/rhododendronA-Z.htm

The lack of a perfect correlation between similarity and compatibility is not surprising: the more thin and curved are ridges of potentially fit genotypes, the less is the genetic distance which can lead to incompatibility (Fig. I41). There is no reason for these properties of the fitness landscape to be the same in all its regions.

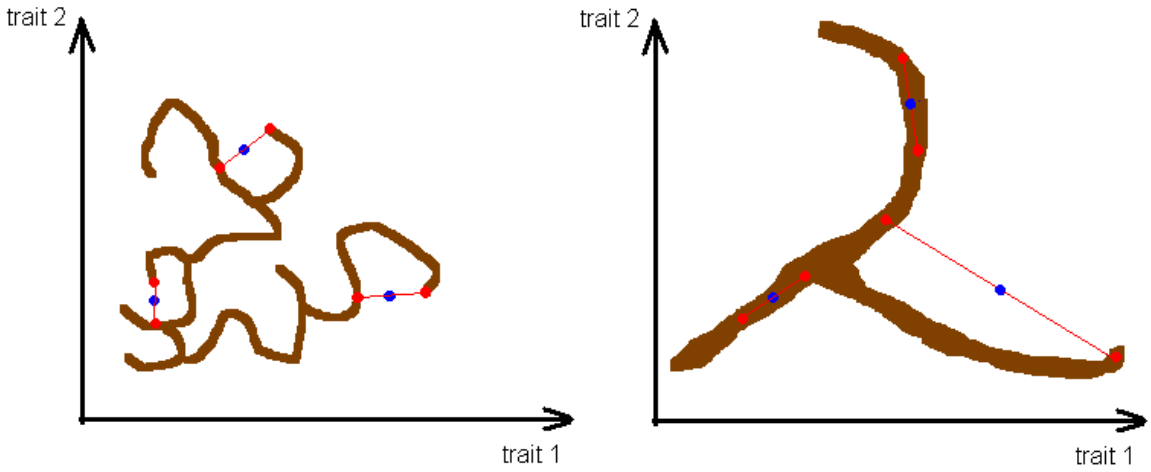


Fig. I41. If fitness ridges are thin and strongly curved, crosses (red) between even relatively similar genotypes can produce unfit hybrids (blue), whose genotypes reside outside the ridges (left). In contrast, with thick, nearly-straight fitness ridges, incompatibility occurs mostly between very dissimilar genotypes (right).

Genotypes differing by less than 0.01 are usually connected, and those differing by more than 0.05 are usually disconnected, but, again, the correlation between similarity (or relatedness, or compatibility) and connectedness is far from perfect. Genotypes that are incompatible can, nevertheless, be connected (Fig. I36) and, in contrast, compatible genotypes are often disconnected (Fig. I40).

13. Clades and species

Now we are equipped to consider the structure of the diversity of life. This structure can be understood only if we take into account its evolutionary origin and, conversely, it represents one of chief sources of our knowledge of evolution. At any moment of time, structure of the diversity of life can be thought of as a pattern of occupation, by the genotypes or phenotypes of actual organisms, of the mostly empty space of all possible genotypes or phenotypes. Here analogy with the distribution of stars in physical space (not in phase space) can be helpful (Fig. I42).

Fig. I42. An ancient galaxy, viewed by Hubble telescope through a gravitational lens (http://hubblesite.org/gallery/album/galaxy_collection/pr1996010a/).

The fundamental structure of biodiversity is two-scale. One the one hand, every individual belongs to a population (Fig. I20). On the other hand, there are many distinct kinds of individuals (Figs. I1), so far referred to as species. Generally, individuals from the same population are very similar, tightly related, fully compatible (with some exceptions, such as Rh-incompatibility in humans), and connected to each other. In contrast, individuals of distinct kinds are dissimilar, only distantly related, incompatible (with many exceptions), and not connected, because variation within each kind is much narrower than gaps between kinds. Indeed, no human can be mistaken for a chimp (Fig. I37) and, when taxonomists investigated ornithofaunas of remote lands, they often described the same sets of bird species as the aboriginal people recognized since times immemorial. This two-scale structure is to some extent analogous to stars (individuals), organized in galaxies (distinct kinds). In both cases, intermediate-level clustering can also be present.

Let us consider evolutionary processes which shape biodiversity, by making vacant genotypes occupied and occupied genotypes vacant. These processes are:

1) deviation of an offspring from its parent, due to mutation and, perhaps, recombination. As the result, a lineage conquers the new location in the space of genotype and abandons the old one.

2) extinction of individual lineages, which occurs when an individual leaves no offspring and vacates its genotype.

3) branching of an individual lineage, which occurs when a lineage splits into two (or, perhaps, more) independently evolving parts. Independent evolution of lineages generally leads to their progressive divergence. Occasional episodes of convergent evolution or lateral gene transfer cannot counterbalance this pervasive trend. Analogously, galaxies mostly move away from each other in physical space, due to its expansion.

4) interorganismal genetic exchanges, due to sexual reproduction, lateral gene transfer, and rare instances of symbiogenesis. Sexual reproduction includes routine cross-breeding between similar members of the same sexual population and rare hybridization between less similar individuals. Lateral gene transfer, common in prokaryotes but rare in multicellular eukaryotes, sometimes occurs between very different forms of life.

If all the lineages were more or less equally durable and did not exchange genes, their branching and extinction would produce phylogenetic trees with a simple structure (Fig. I43). In reality, however, phylogeny is more like a two-scale object, with few fat trunks. These trunks only rarely go extinct or multiply and consist of bunches of mostly ephemeric, tightly related individual lineages. The two-scale structure is due to very heterogeneous life spans of individual lineages: most of lineages are short-lived, but a few are very long-lived. Indeed, nothing protects individual ecologically equivalent lineages, which together constitute a trunk, from extinction, although their total number is more or less conserved, maintaining the trunk as a whole (Chapter 2.1). Rare additional durable lineages emerge if they manage to avoid competition with other similar lineages, due to geographical separation or ecological differentiation. Genetic exchanges between similar organisms make the trunks ever more cohesive, and abolish tree-like structure within each trunk. As the result, the current biodiversity, as well as biodiversity at any moment in the past, can be viewed as mostly consisting of clusters of similar individuals. Analogously, the Universe is mostly populated by galaxies, instead of individual stars.

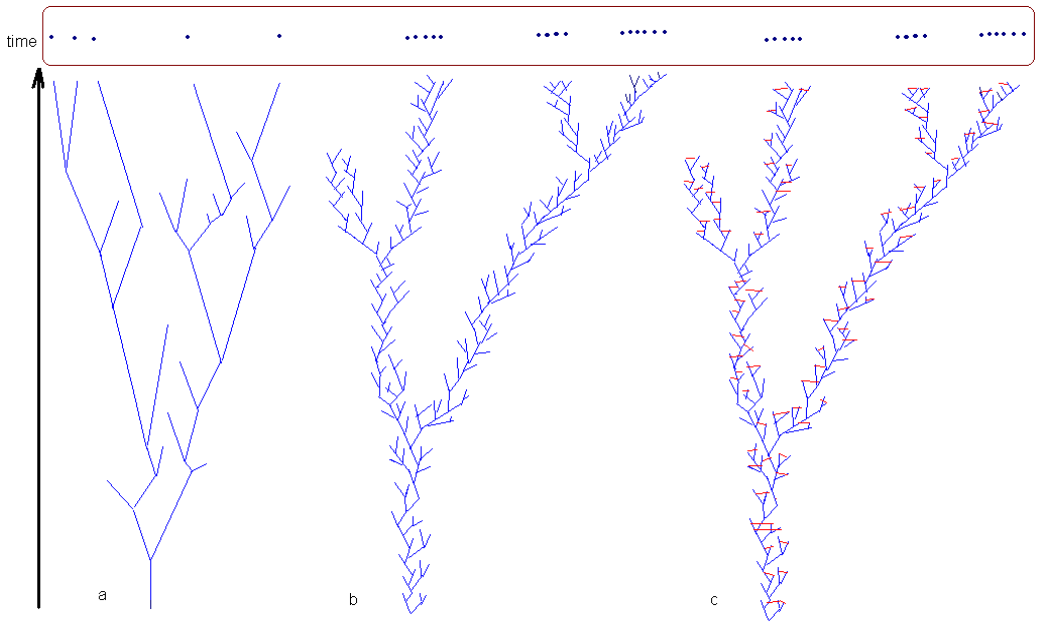


Fig. I43. Divergence of ancestor-descendant lineages. Divergent evolution with all lineages being more or less equally durable (a) and with most lineages ephemeric and a few lineages very durable, without (b) and with (c) genetic exchanges between tightly related individuals. Patterns of occupation of the space of genotypes by extant organisms is shown on top of each tree.

However, even the pattern shown in Fig. I42c is still an oversimplification because it consists structureless clusters that are neatly distinct from each other. However, structures at the intermediate levels of similarity are also important (Fig. I37a). In particular, large continuous objects can appear in the space of genotypes (Fig. I44), if new genotypes are conquered faster than old genotypes are vacated. Such objects can result from evolution within a wide continuous geographical range (Fig. I36) or from hybridization (Fig. I40). Analogously, a galaxy always has structure, instead of being just a disordered bunch of stars.

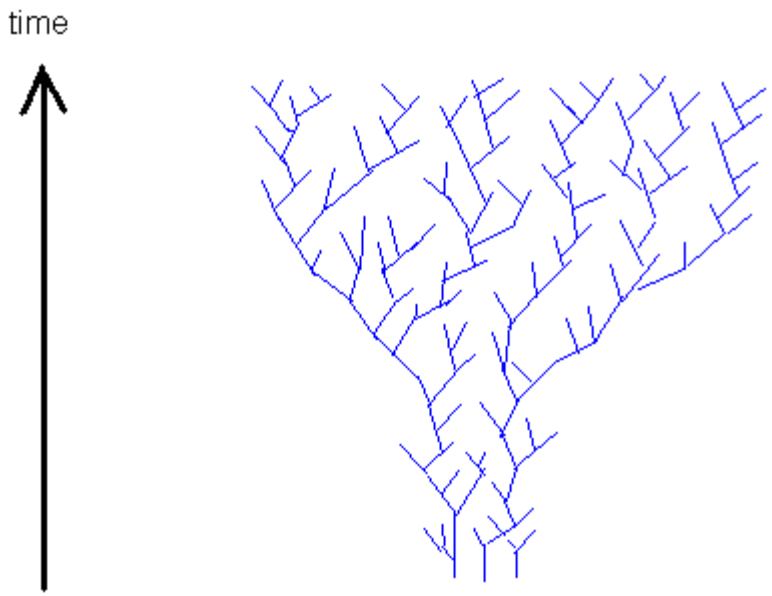


Fig. I44. Evolution of wide, continuous occupation of the space of genotypes by simultaneously living individuals.

Let us consider possible patterns of occupation of the space of genotypes from the point of view of one particular individual. The genotype of this individual is at the center of a sphere of some radius I, such that other genotypes located inside this sphere are compatible to it, and more distant genotypes are incompatible. The genotype of a chosen individual is always the part of a cluster which represents genotypes of other members of its population. At a larger scale, the space of genotypes is mostly void, since gaps between clusters of existing genotypes are much larger than these clusters. Analogously, intergalactic distances are much larger than diameters of galaxies.

In the simplest case, only one small cluster is located within the sphere of compatibility (Fig. I45). A more complex situation appears if the individual belongs to a larger but continuous cluster which, however, still fits completely into the sphere of compatibility, and there are no other objects within this sphere. The second situation is exemplified by modern humans. In both cases, we can say that the individual belongs to a "good species", all members of which are connected and compatible to each other, and distinct from and incompatible to the rest of biodiversity.

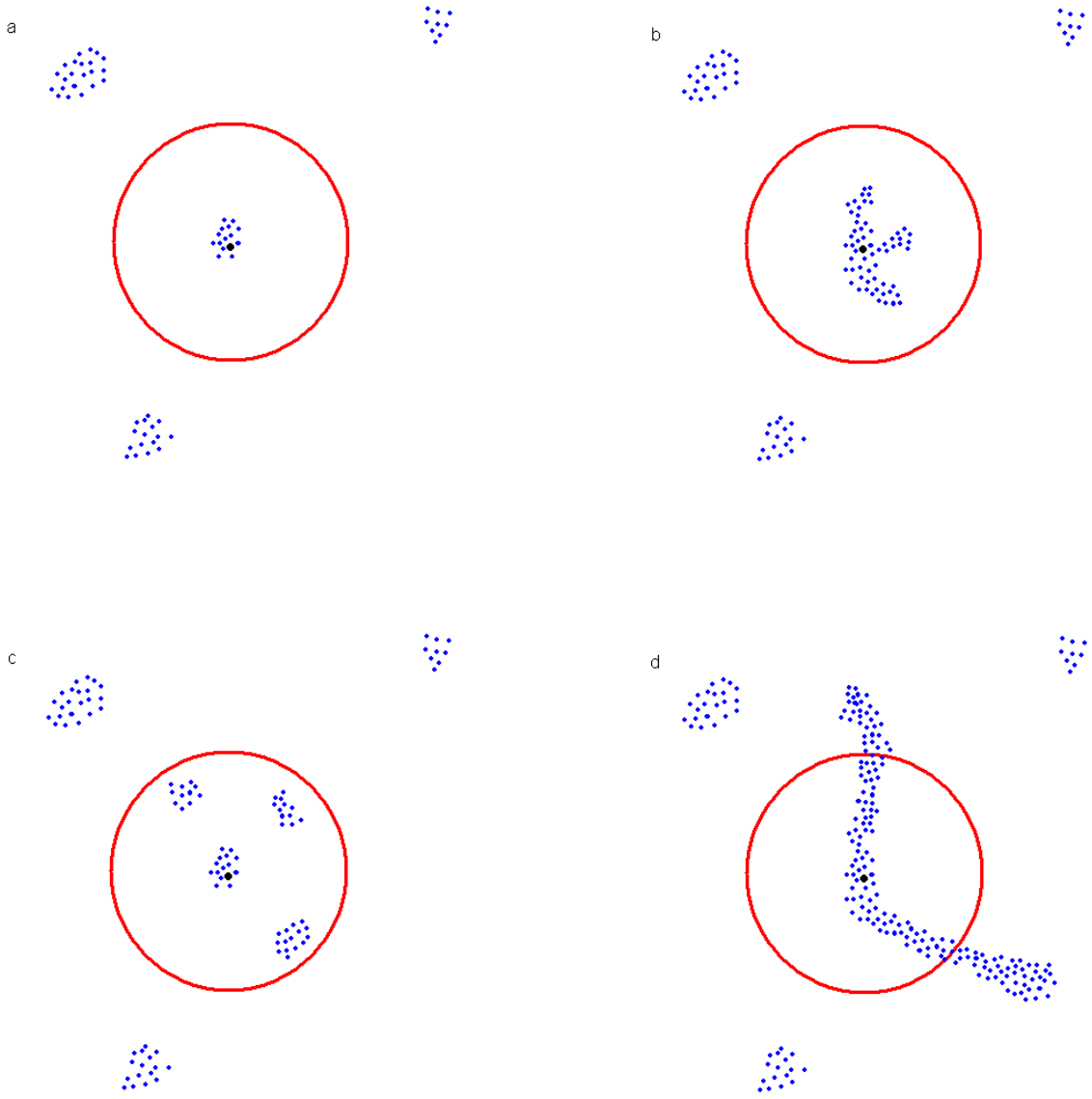


Fig. I45. Patterns of occupation of the space of genotypes, relative to one particular genotype, shown by black dot. a) Good, homogeneous species. b) Good, heterogeneous species. c) Distinct compatible species. d) Continuous incompatible species.

However, more complex situations, of the two opposite kinds, are also possible. First, there may be several compact separate clusters of genotypes within the sphere of compatibility, as it is the case with distinct but fully compatible rhododendrons (Fig. I40). Alternatively, there may be a continuous cluster which transcends the sphere of compatibility, as it is the case with greenish warblers (Fig. I36).

To comprehend all these situations, it is necessary to view the structure of biodiversity from all the four related perspectives introduced above. From the perspectives of similarity and connectedness, biodiversity simply consists of more or less distinct clusters of similar individuals, which may be referred to as forms of life. From the perspective of relatedness, biodiversity consists, as long as genetic exchanges between lineages can be ignored, from clades (branches in Latin; hence cladogenesis), a clade being the set of all the descendants of a particular common ancestor, so that a member of a clade is more tightly related to any other member than to any non-member (Section 1.1.4). Finally, from the perspective of compatibility, biodiversity consists of more or less distinct species (kind in Latin): two individuals belong to the same species if they are compatible, and to different species otherwise (Chapters 1.5 and 2.6).

A "good species", such as modern humans, is also a form of life and a clade, but more complex situations are also very common. Indeed, when a population splits, this sets in motion divergent evolution of the newly isolated parts, and it makes no sense to seek the exact moment when they become disconnected enough to earn them a status of separate forms of life or incompatible enough to be regarded as different species. Even the boundaries between tightly related clades are somehow blurred, as long as we consider phylogenies consisting of fat trunks (Section 1.1.3). This is just a manifestation of Zeno's heap paradox (what is the minimal number of grains which already constitutes a heap?). Populations are usually the least inclusive groups of individuals, because many similar populations can constitute a form of life, a clade, or a species (Chapter 2.1).

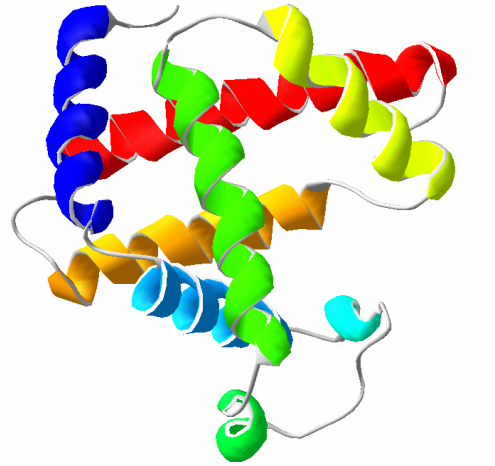
14. Complexity, optimality, evolvability, and designability

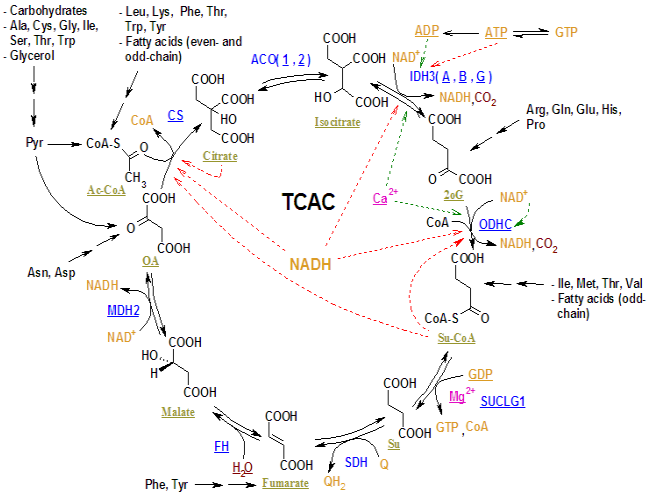
The second wonder of life is its complexity. Before Darwin, it was universally accepted that complex phenotypes cannot appear naturally and, thus, must be of supernatural origin. For example, in 1802 William Paley famously claimed in his book "Natural Theology" that complex adaptations testify to the existence of God, because, if you find a pocket watch in a field, it is reasonable to assume that is was made by a watchmaker, instead of natural forces, and was simply lost there.

Indeed, how could gradual evolution produce complex adaptations, somehow finding them among countless possible phenotypes? Darwin was fully aware of this

problem, and wrote: "To suppose that the eye, with all its inimitable contrivances for adjusting the focus to different distances, for admitting different amounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree" ("The Origin of Species", Chapter 6). Before we can address this central issue in depth, four concepts need to be introduced - complexity, optimality, evolvability, and designability.

There is a number of formal definitions of complexity, but here it is sufficient to say that a phenotype is complex if it 1) consists of many interacting parts, 2) requires a long description, so that its phase space has many dimensions, and 3) is fragile, in the sense that even a small change can alter its properties radically. At all levels, functional phenotypes are usually extremely complex (Fig. I46). Thus, it is important to remember that fitness landscapes over 1- and 2-dimensional phase spaces provide only a crude approximation to biological reality, and may sometimes be misleading.





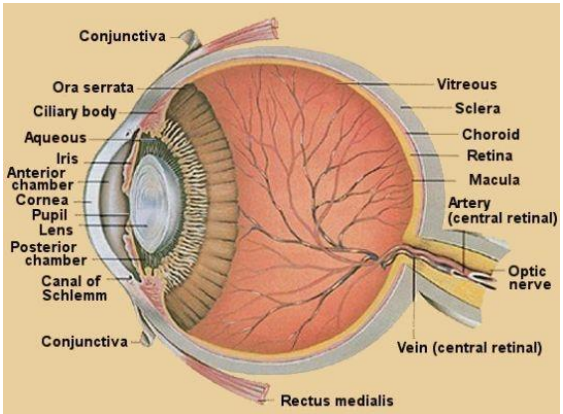


Fig. I46. Complex phenotypes at different levels. A representation of the 3D structure of myoglobin, a 153 amino acid single-chain heme containing protein, the primary oxygen-carrying protein of muscles
http://en.wikipedia.org/wiki/Image:Myoglobin.png
Krebs cycle
http://upload.wikimedia.org/wikipedia/en/9/97/TCA_reactions.png
A simplified scheme of the anatomy of human eye
http://www.lhup.edu/~dsimanek/scenario/labman3/eyeanat1.jpg
The anglerfish, *Melanocetus johnsonii*, dangles its lure, which glows due to presence of symbiotic bioluminescent bacteria, to attract prey.
http://www.newsday.com/media/photo/2003-12/10801670.jpg

A phenotype is optimal if it performs its function better than all other feasible phenotypes and, thus, corresponds to the highest fitness peak. Assaying optimality is difficult, since this requires knowledge of the whole fitness landscape. Still, we can claim that, for example, human blood vessels usually branch off at more or less optimal angles, minimizing the resistance of blood, since the laws of physics make it possible to calculate this resistance for any angle. In contrast, human eye, although fully capable of doing its job, cannot be optimal, since the blood vessels and nerves that attach to photoreceptor cells are positioned between the pupil and the surface of the retina. In fact, evolution is prone to produce functional complex phenotypes that are only suboptimal (Section 1.1.1).

The two remaining concepts, evolvability and designability, describe the possibility of evolution from the opposite perspectives. A phenotype is evolvable if it belongs to a continuous set of fit phenotypes (Fig. I32b) and, thus, can be substantially changed, and perhaps improved, without losing its function. This is not necessarily the case. Indeed, it is impossible to gradually transform (Fig. I12) Hamlet into Pride and Prejudice (or one computer program into another) in such a way that all the intermediate sequences of letters are also meaningful, to say nothing of them being great works of literature (or useful programs). In other words, sensible texts in English or C correspond to only isolated peaks in the corresponding spaces of texts. One can think of evolution of life as percolation of lineages, through the space of genotypes, along the paths within the set of potentially fit genotypes, similarly to how a liquid percolates through a porous

material. The Strong Claim implies that all existing organisms belong to one set of fit genotypes, which is continuous at the scale of possible parent-offspring differences. In contrast, parts of phenotypes, such as proteins, probably cannot be always transformed into each other gradually without losing their function because several primordial proteins could have appeared independently and because even today proteins occasionally appear from scratch, when a non-coding sequences starts being transcribed and translated.

Currently we have only a very limited direct data on fitness landscapes of complex phenotypes. For example, is has been shown experimentally that replacing five amino acids is enough for an enzyme isopropylmalate dehydrogenase to switch its coenzyme from nicotinamide adenine dinucleotide (NAD) to nicotinamide adenine dinucleotide phosphate (NADP), and that all the intermediate molecules are functional (Fig. I47). Naturally, obtaining such data for functioning phenotypes at the higher levels of organization would be even more difficult.

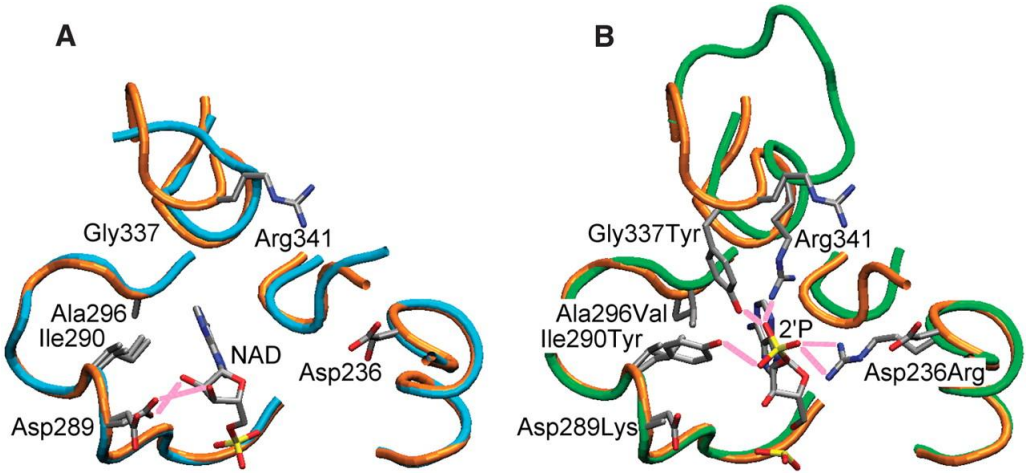


Fig. I47. A: Five amino acids of isopropylmalate dehydrogenase which are important for preferentially bounding NAD (ignore anomalous Arg341). B: Five amino acid replacements which convert NAD-binding enzyme into NADP-binding enzyme. Together, these replacements change the enzyme from the 100-fold preference for NAD to the 200-fold preference for NADP (Science 310, 499, 2005).

Continuity of the set of fit genotypes is particularly remarkable because fitness is very fragile, and a single nucleotide replacement can be lethal. Thus, fit phenotypes are

never far away from the overwhelming majority of unfit phenotypes. However, the rareness of potentially fit phenotypes may, as long as they are arranged continuously, also facilitate evolution. At least, arriving to the correct amino acid sequence of myoglobin cannot involve producing and evaluating all the $20^{153} > 10^{100}$ possible amino acids sequences of a protein of its length, most of which are useless, together with all their neighbors in the sequence space. Although the evolution of myoglobin probably involved more than the minimal number of 153 steps, sufficient to arrive to the correct protein from any corner of the space of 153-amino acid proteins, there is no reason for this number to be astronomical.

Designability describes the same issue from the perspective of a target phenotype, instead of the actually present phenotype. A phenotype is designable if it can be reached, by gradual adaptive evolution, from a large volume within the space of phenotypes, *i. e.* if it belongs to a large continuous set of fit phenotypes. Darwin pioneered using heuristic arguments for designability of even very complex phenotypes. Although this may seem to be absurd at the first sight, in reality there is no reason why evolution cannot arrive to human eye by gradually perfecting an initially very simple photoreceptor.

Still, there likely exist a lot of potentially fit phenotypes which, however, can never be reached by evolution departing from a simple initial phenotype, as a chess position shown in Fig. I11c cannot be reached from the initial position. This may be the reason of why there are no photosynthesizing cows. Also, the properties of the set of fit genotypes may restrict evolution due to unavailability of shortcuts. For example, it may be useful for a dinosaur to acquire wings in addition to its four legs. However, if no feasible beneficial mutation in a dinosaur genome will move it closer to the six-limb condition, the only way for it to get extra wings would be to go all the way back to a legless worm, and then grow four legs and two wings. Because this did not happen, we have two-legged birds.

We may soon see artificial creation of functional proteins which do not exist in nature only because they are not designable. Still, rigorous theoretical analysis of designability of complex phenotypes is not yet possible and, generally, both the theoretical (Chapter 3.2) and experimental (Chapter 3.3) studies of the evolution of complex phenotypes are in their infancy.

15. Stochasticity, reproducibility, and random drift

So far, we implicitly assumed that the outcome of evolution is always predetermined. Indeed, if a lineage simply climbs upwards on the fitness landscape, we can predict on which peak it will end up, given its initial state. This convenient assumption belongs to a venerable paradigm of determinism. Just 100 years ago, people thought that deep down Nature is deterministic - if we knew everything about the current moment, we could predict the future. For example, equations of classical mechanics possess a property known as Laplacian determinism: current positions and velocities of all the bodies predetermine their future dynamics.

However, in early XX century quantum mechanics informed us that Nature is inherently stochastic: no attainable knowledge of the present is enough to predict the future exactly. Instead, we can only talk about probability of each of many possible outcomes, although some constraints on them do exist, for example, due to conservation of energy. Einstein refused to accept this revolutionary conclusion, insisting that "God does not play dice" - but, apparently, he was wrong (Fig. I48). Roughly speaking, a process is deterministic if it is predictable, and stochastic if it is unpredictable, although there are some complications, such as deterministic chaos.

Fig. I48. It is now generally accepted that Nature is inherently stochastic.

Probability is a fundamental concept (like number or trait) and thus cannot be usefully defined through other concepts. Instead, let us consider its relationship with two other fundamental concepts - of symmetry and frequency (Fig. I49). If n outcomes of a

trial are symmetric (there is no reason why one is more likely than the other), we can say that each will occur with probability 1/n (think of rolling of a die). If we repeat the same trial more and more times, the frequency of a particular outcome will approach some limit, and we can regard this limit as the probability of this outcome.

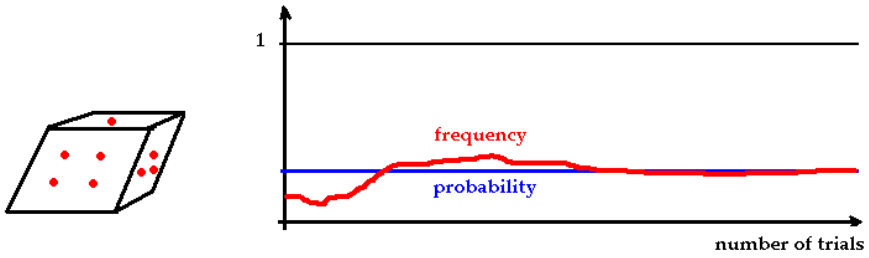


Fig. I49. The concept of probablity is related to the concepts of symmetry (left) and frequency (right).

A complementary aspect of determinism (stochasticity) is reproducibility (irreproducibility). Indeed, if a deterministic process is run many times from exactly the same initial conditions, we will always obtain the same outcome, which is, thus, reproducible. In contrast, outcomes of a stochastic process are irreproducible.

Evolution of life is stochastic, *i. e.*, unpredictable and irreproducible, due to a number of factors. First, elementary processes on which evolution is based are inherently stochastic. Mutations are molecular events, and, thus, fall into the domain of quantum mechanics, so that there is no way to predict whether a particular mutation will happen on a given occasion. Reproduction of an individual may be more remote from quantum effects, but it is still obvious that we cannot predict whether a newborn fly will be eaten in 10 minutes or will live on to lay 1000 eggs.

Second, populations are never large enough to conceal stochasticity of these elementary processes. Indeed, decay of a radioactive atom is inherently stochastic, but the decline of the number of such atoms in a macroscopic sample is deterministic for all practical purposes (Fig. I8), because there are $6x10^{23}$ molecules in a mole, and, according to the law of large numbers, probabilities essentially turn into frequencies. In contrast, an evolving population usually contains "only" 10^4 - 10^{10} individuals, and stochasticity of individual acts of mutation and reproduction is felt even at the population level. A term

random drift refers to random fluctuations of allele frequencies due to stochasticity in allele transmission by individuals (Fig. I50). Even two initially identical populations will soon become different, due to irreproducibility of this random drift. Finally, environments and, thus, fitness landscapes are themselves irreproducible, so that even two identical populations will never be subject to exactly the same selection.

Fig. I50. Possible changes of an allele frequency in the course of one generation, due to random drift. The population size is ~100 individuals, and the brightness of an arrow reflects the probability of the corresponding change.

Together, these factors are responsible for inherent unpredictability and irreproducibility of evolution, making it even more fascinating. Indeed, even when faced with essentially the same challenge, different populations, although tightly related to each other, often adapt differently (Fig. I50). Still, the assumption of determinism makes things so much simpler that it is always a good idea to use it for as long as possible, being fully aware of its limitations.

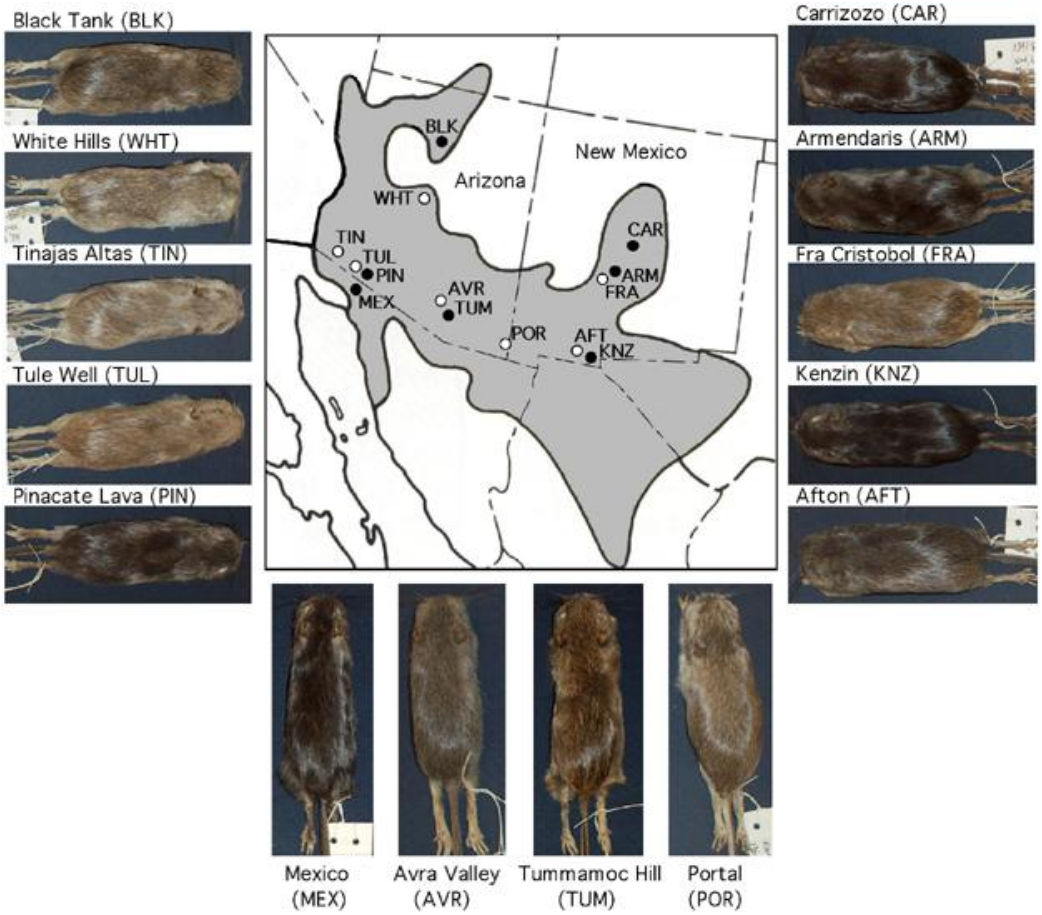


Fig. I50. Rock pocket mice, *Chaeotdipus intermedius*, are generally light-colored and live on light-colored rocks. However, some of their populations live on dark lava, and they acquired dark coat color which conceals them from predators. Pocket mice from Pinacate and from Armendaris (and, perhaps, from other lava flows, shown by black dots on the map) evolved their dark coat colors, from the ancestral light color, due to completely different genetic mechanisms.
(http://www.nature.com/hdy/journal/v94/n2/full/6800600a.html)

16. Microevolution and Macroevolution

The main goal of evolutionary biology is to explain diversity and complexity of life or, in Darwin's words, "the origin of species". According to Weak and Strong Claims, these phenomena are results of long, profound changes, Macroevolution. If the Lamarckian mechanism were at work, Macroevolution would be the whole story: a relatively small amount of variation within large-scale lineages could be safely ignored.

However, within-lineage variation is essential for the real, Darwinian mechanism of evolution. Thus, we must take into account that any evolving large-scale lineage is a spot, however tiny, but not just a point, within the space of genotypes. Macroevolution is always driven by changes of the genetic variation of populations, Microevolution. Microevolution proceeds much faster than Macroevolution and can often be directly observed during a relatively short period of time (Fig. I51).

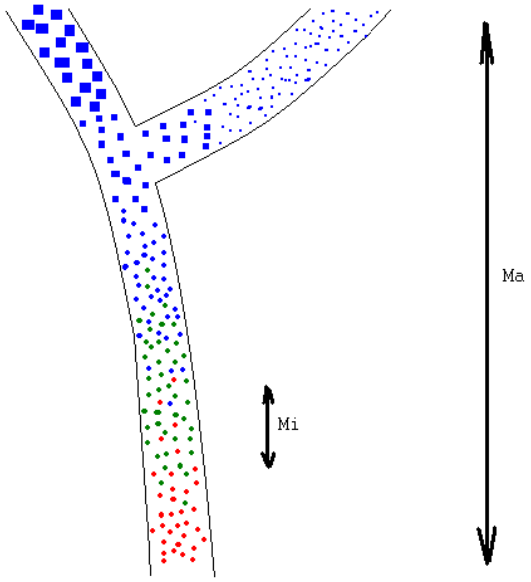


Fig. I51. Time scales of Micro- and Macroevolution.

Naturally, Micro- and Macroevolution are not two independent processes. After a period of time when two or many alleles at some locus segregate within the population, one of them may become fixed, and the rest become lost, a process called allele replacement. Allele replacements can be viewed as an interface between Micro- and Macroevolution: fixation of a new allele is the result of Microevolution and is an elementary Macroevolutionary event.

17. Evolutionary biology today: generalizations, theories, and relevance

The current state of evolutionary biology is a patchwork of knowledge and ignorance. We know a lot about past evolution of life, although not about its origin.

Despite a continuous stream of new fascinating discoveries, many key facts about the history of life on Earth have already been firmly established. A number of important generalizations emerge from these facts.

A lot is also known about Microevolution. Intrapopulation variation and the driving forces of Microevolution have been studied at different levels and its dynamics are to a significant extent understood theoretically. Even the processes which lead to formation of new species do not look mysterious any more. However, parameters of natural selection are still known rather poorly, and Microevolution of phenotypes remains obscure.

In contrast, we still lack a comprehensive theory of Macroevolution, and our understanding of the subject mostly depends on generalizations about past evolution. For many ot them, a simple, plausible, proximal explanation can be provided. For example, a generalization stating that a mass extinction is always triggered by rapid and drastic changes of the global environment is not at all surprising: organisms are adapted only to some scope of the environmental conditions, and may die if these conditions change too much and too fast.

In contrast, many other generalizations could possibly be explained only by a theory of the corresponding facet of evolution. It is essential to understand the relationships between generalizations, which simply summarize the data, and theories, which has the power to explain them through some general principles. For example, three Kepler's laws which state that:

1. The orbit of every planet is an ellipse with the Sun at a focus.
2. A line joining a planet and the Sun sweeps out equal areas during equal intervals of time.
3. The square of the orbital period of a planet is directly proportional to the cube of the semi-major axis of its orbit.

are generalizations describing the patterns in orbital motions of plantes. The theory which explain these patterns is provided by Newton's laws of motion and universal gravitation, which were discovered almost a century after the Kepler's generalizations. In contrast, the second Mendel's law, segregation of phenotypes in F_2, is a generalization, and the explaining theory, random union of haploid gametes carrying discrete alleles, was proposed by the same person.

Thus, generalizations, however important, are only stepping stones which should eventually lead from facts to their deep understanding. For some generalizations, theories of the corresponding facets of Macroevolution that have the power to explain them have already been developed. However, a number of key generalizations, in particular those dealing with evolution of complex adaptations, remain enigmatic.

The biggest obstacle to the theory of Macroevolution is poor knowledge of fitness landscapes. Understanding Microevolution is of very little help here, because variability within any population sheds light only on a tiny fraction of the fitness landscape. Studies of Macroevolution are relatively more successful at the population level, where it is often possible to avoid the task of attributing phenotype and fitness to each genotype. Only now, armed with information on many complete genomes, we got a chance to understand Macroevolution of molecules, cells, and organisms. This is the current frontier of evolutionary biology.

The goals of studying Micro- and Macroevolution are so different that it often leads to confusion. The genetical mechanism of Microevolution, Darwinian natural selection operating on Mendelian variability within populations, was uncovered in the XX century and proudly called "Synthetic Theory of Evolution". Then, this "Synthetic Theory" was accused of not being able to predict the course of Macroevolution. This is true, but misses the point. The course of Macroevolution is determined by large-scale properties of fitness landscapes, and the "Synthetic Theory" can only tell us - correctly - that a lineage evolves in the direction which increases its fitness. In a similar way, Newtonian law of gravity is not enough to predict the trajectory of a stone which rolls downhill: although gravity forces it to move, its trajectory is determined by the landscape, including, perhaps, numerous small bumps, which can make the trajectory essentially unpredictable. Thus, understanding Microevolution is insufficient for understanding Macroevolution.

Many years ago, Feodosij Grigor'evich Dobrzhanskij (known to English speakers as Theodosius Dobzhansky), one of the founders of modern evolutionary biology, boldly proclaimed that "nothing in biology makes sense, except in the light of evolution". It is easy to dispute this claim. Many monumental discoveries, *e. g*. that DNA in cells consists of two complementary antiparallel strands, were made, and make a lot of sense, independent of any evolution. Biology still has to do a lot of discovering of what the

modern organisms are made of. For such work, the fact that these organisms are products of long evolution and are related to each other is often, at most, only marginally relevant.

However, the importance of evolutionary biology outside its traditional domains of paleontology, comparative morphology, and population genetics is growing. Comparative analysis of genomes, which elucidates their functioning, is impossible without understanding evolution. Evolutionary analysis becomes more and more important in the studies of spatial structures of macromolecules. Discovery of striking evolutionary conservatism of some genes governing the key steps in development left a profound impact on developmental biology. Evolution is crucial for the analysis of many diverse phenomena such as senescence, interactive behavior, and sexual reproduction. Many advances in understanding diseases were made by studying proteins, which correspond to human disease-related proteins, in mice, fruit flies and even fungi. Evolution of human pathogens is of paramount importance for public health. Artificial evolution of molecules, based on the Darwinian mechanism, becomes a promising tool for drug design. This tendency will persist, because there remain fewer and fewer undiscovered facts about life that are simple enough to make sense even outside the evolutionary paradigm. Thus, Dobzhansky was right - but way ahead of his time.

Also, evolutionary biology made a larger impact on thinking outside natural sciences than any other field of them, including even relativity and cosmology. This impact is often exaggerated: contrary to what some people claim, it is impossible to reject of prove the existence of God by studying evolution of life. Still, the genesis of life in general and of our own species in particular must be of profound interest to any curious person and might, indeed, affect even our core beliefs.

18. Logic and structure of this book

The key obstacle to conveying the beauty of St. Basil's Cathedral (Fig. I52) by words is that you must start somewhere and then keep going, one word at a time. Which dome to describe first - and what to leave to the end? Writing a textbook is even more hopeless - scientific truth resides in much more than mere three dimensions, and surpasses in beauty everything made by humans. Reality can never fit comfortably into a Procrustean bedstead of a necessarily unidimensional text. Any sequence of chapters is

bound to be imperfect, and a careful reader will always see hidden faults, tensions, and compromises. Still, some layouts of a textbook are worse than others, and a reasonable layout must reflect both the intrinsic makeup of the subject and the current state of its knowledge. Below, I outline the logic behind this book, and its resulting structure.

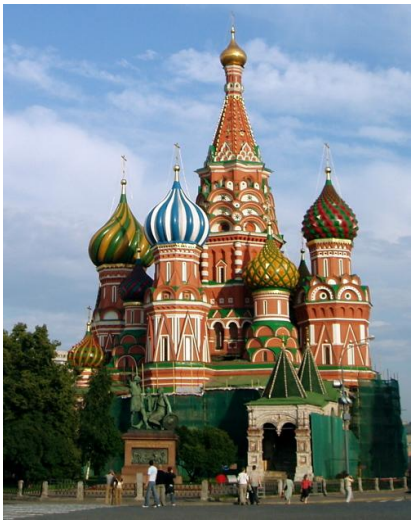


Fig. I52. An object that defies verbal description.

It makes sense to begin coverage of evolutionary biology from history of life on Earth - or from its past evolution, which is the same thing. Indeed, without data on past evolution we would know very little about Macroevolution, because its currently available theory is rudimentary. Thus, history of life is the subject of Part 1. Naturally, Macroevolution is emphasized there, and presentation is organized around facts and methods of uncovering these facts. Questions "what?" and "when?" are emphasized, and questions "why?" and "how?" are mostly ignored.

From the very beginning, a reader will be asked to believe in two bold claims, Weak and Strong, about the past of modern life. Biologists are so used to these claims that evidence for past evolution are often considered only in the context of debates with creationists. This cannot be justified: understanding of why common ancestry of all modern species is regarded as a definite fact provides a solid foundation for studying all aspects of evolutionary biology. Indirect evidence for past evolution, based on properties of modern life, are enough to settle the matter. After this, it makes sense to talk about

phylogenies of sets of modern species, and to review methods of inferring phylogenies. These topics are covered in Chapter 1.1.

Direct evidence of past evolution, provided by fossils, complement indirect evidence and also supply a lot of crucial data. Such direct evidence cannot be studied outside the context of Earth sciences. Indeed, fossils reside within rocks and, moreover, rocks are the key source of data on past environments. Thus, Chapter 1.2 provides a very brief review of Earth sciences, of which paleontology is an important branch.

Together, indirect and direct data brought enormous amount of information on past life. Chapter 1.3 deals with this information from the chronological perspective, and covers evolution of life from first unambiguous fossils formed ~3.5 Bya until the present.

Recent history of the human lineage obviously deserves special attention and is covered in Chapter 1.4.

To complement a chronological perspective on the history of life, we will consider timeless generalizations. Each generalization describes a particular pattern, encountered repeatedly in the course of past evolution. Some generalizations are well-understood, but many others remain obscure. A number of key evolutionary generalizations are treated in Chapter 1.5.

Although truly profound Macroevolution is extremely slow, substantial changes of lineages may occur fast enough to be observed directly. Chapter 1.6 considers a number of instances of rapid evolution, and, thus, provides a smooth transition to Part 2.

Part 2 deals with Microevolution and thus focuses on minor changes in evolving populations, at the scale of their genetic variation. In contrast to Part 1, Part 2 is organized around the Synthetic Theory of Microevolution, which consists of Darwinian selection and Mendelian genetics. A theory-based exposition of any subject is always preferable, as long as a good enough theory is available, and studies of Microevolution is the only branch of evolutionary biology where such luxury is already affordable. Of course, the relevant data also have to be considered. The question "why a genotype has the fitness it has?" is mostly ignored.

First, we need to introduce population, the key object of Microevolution, and a variety of experimental, statistical, and dynamical methods of studying populations. This task is performed by Chapter 2.1.

The key feature of the Darwinian mechanism of evolution is the central importance of heritable variation between competing individuals. Chapter 2.2 considers within-population variation at the level of genotypes and phenotypes, together with small-scale properties of genotype > phenotype maps.

The five factors of Microevolution - mutation, selection, mode of reproduction, population structure, and random drift - are introduced and described in Chapter 2.3.

The genetical theory of Microevolution, the analysis of dynamics of populations under simultaneous action of all its five factors, is reviewed in Chapter 2.4.

Of course, the goal of theory of Microevolution is to understand what happens in natural populations. Data on natural selection, and implications of properties of Microevolution for Macroevolution are all elucidated by this theory and are treated in Chapter 2.5. Surprisingly, studies of Microevolution are only of limited utility for understanding Macroevolution. Indeed, selection that occurs within natural populations mostly has nothing to do with adaptive evolution and the range of within-population variation is too narrow to illuminate large-scale properties of fitness landscapes.

Coverage of Microevolution ends with consideration of the process of speciation, understood as evolution of incompatibility. This subject can be attributed to both Microevolution and Macroevolution, and here I treat Microevolution inclusively. Species and speciation is the subject of Chapter 2.6.

Part 3 considers what little is understood about Macroevolution. Here, presentation is organized around the concept of levels of organization. At each level, the relevant theories are used, to the extent possible, and questions "why?" and "how?" are emphasized. Topics covered in Part 3, especially in Chapter 3.2, are the least understood and the most important in all evolutionary biology.

Chapter 3.1 treats Macroevolution of genomes. At the sequence level, Macroevolution is just the record of past allele replacements, so that this subject is tightly connected to Part 2, and comprehensive theory, both statistical and dynamical, is available for many aspects of genome evolution.

Chapter 3.2 presents the currently available rudiments of understanding of Macroevolution of complex, functioning phenotypes at the molecular, cellular, and organismal levels. This understanding is based on theory that deals with inferring complex fitness landscapes and with dynamics of evolving complex objects. This theory

can then be applied, so far with very limited success, to data on actual Macroevolution of complex phenotypes. Thus Chapter 3.2 is tightly related to chapter Chapter 1.5.

Chapter 3.3 considers Macroevolution of external properties of individuals, which describes them as members of populations. Thus, the same subject can also be viewed as evolution at the level of populations. Not surprisingly, this subject, which ignores details of processes inside organisms, is much easier and better-understood than evolution at the functioning levels.

Chapter 3.4 considers Macroevolution of ecosystems.

Finally, Part 4 deals with applications of evolutionary biology to human biology and welfare, and to issues outside natural sciences.

Introductory texts on mature subjects deal only with firmly established facts and concepts. Introductions to Calculus or Inorganic Chemistry written 100 years from now will probably look familiar. In contrast, an author of even an introductory text on Evolutionary Biology has to take sides in ongoing scientific debates (I already did), and to be prepared to make some errors. To avoid overloading the book, I found it necessary to ignore some opinions which still have serious adherents, although I tried to present different points of view when, in my judgment, all of them are still viable. Generally, the degree of certainty declines towards the end of this book.