

Evolution and Probability*

Luca Peliti
SMRI (Italy)
luca@peliti.org

January 6, 2019

Abstract

Life forms exhibit such a degree of exquisite organization that it seems impossible that they could have developed out of a process of trial and error, as intimated by the theory of Darwinian evolution. In this general public paper I discuss how differential reproduction rates work in producing an exceedingly high degree of improbability, and the conceptual tools of the theory of evolution help us to predict, to some degree, the course of evolution—as it is routinely done, e.g., in the process leading to the yearly influenza vaccines.

Natural selection is a mechanism for generating an exceedingly high degree of improbability.

R. A. FISHER

A priori vs. a posteriori probabilities.

Figure 1 shows a popular image of the process of human evolution. We see, starting from a quite ape-like ancestor, a sequence of life forms that progressively become more similar to a present idea of a human, of course male, light-skinned, fair-haired, probably anglo-saxon. Even when one considers other life forms, one gets a similar picture, as in the case of the evolution of the horse (fig. 2). In these representations the evolutionary process appears as the progressive discovery of an archetype (*Homo sapiens* or *Equus caballus*) which existed previously, at least in the world of ideas.

If we accept this image, the theory of evolution faces a great problem: How can this archetype be reached in a reasonable time by the Darwinian mechanism of selection and mutation? To make one example: How can we think that such a complex structure as the human eye (on the left in Figure 3) could be formed by a trial-and-error process? In fact, one can see that the formation of eyes in vertebrates is not the only phenomenon of its kind. During the evolution of Life on Earth, other life forms have developed comparable structures. One of these life forms is the common octopus. Looking at the details of the anatomy of the octopus eye (on the right in Figure 3), we see that, although its structure is quite close to that of

*©Luca Peliti 2019

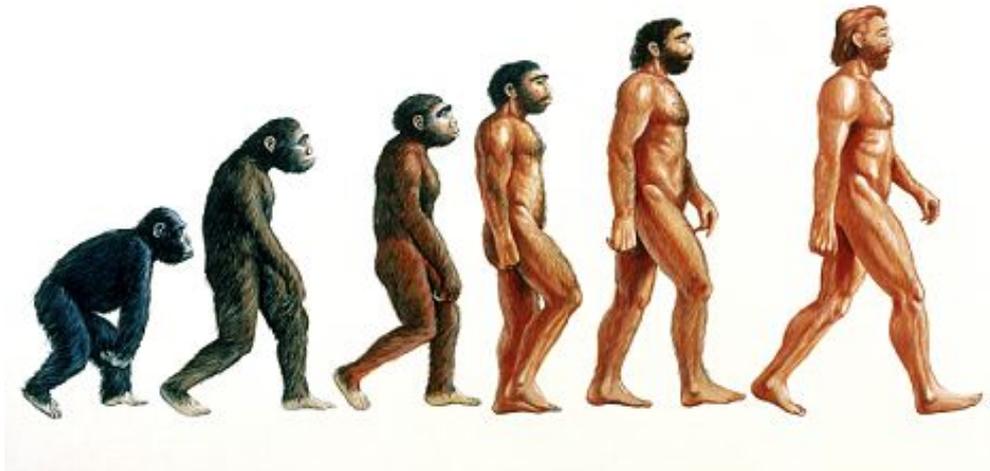


Figure 1: A popular representation of the steps of human evolution. From the *Daily Mail*.

the vertebrate eye, it is the outcome of a radically different process. For example, the retina (1) in the vertebrate eye is made of cells connected to the brain via nerve fibers (2) which gather to form the optical nerve (3): the nerve fibers are placed *above* the retina, so that the light must cross them in order to reach the light-sensitive cells. Indeed, the point where the optical nerve crosses the retina to enter the eye corresponds to the blind spot (4). The octopus eye seems to have benefitted of a more rational design: the nerve fibers (2) lie below the retina (1), and therefore there is no blind spot, and the connections with the remainder of the nervous system take place in an optical ganglion (not represented) placed behind the eye globe.

We can try and make these ideas more precise by considering the evolution of a typical protein. Let us follow an argument by Hoyle and Wickramasinghe [1, cap. 2], who try to prove that the usual evolution theory based on the mutation-selection mechanism cannot explain the formation of complex biomolecules from a

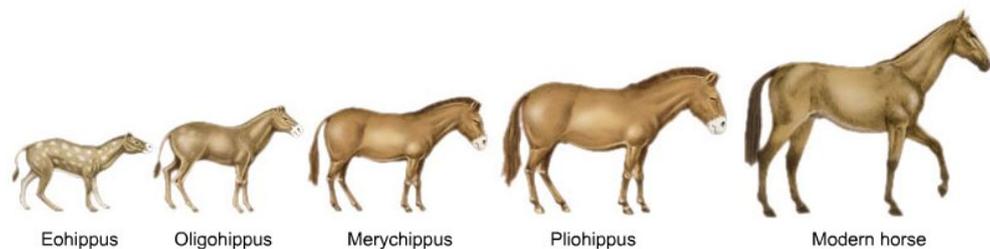


Figure 2: A representation of the evolution of the horse. From *Pinterest*.

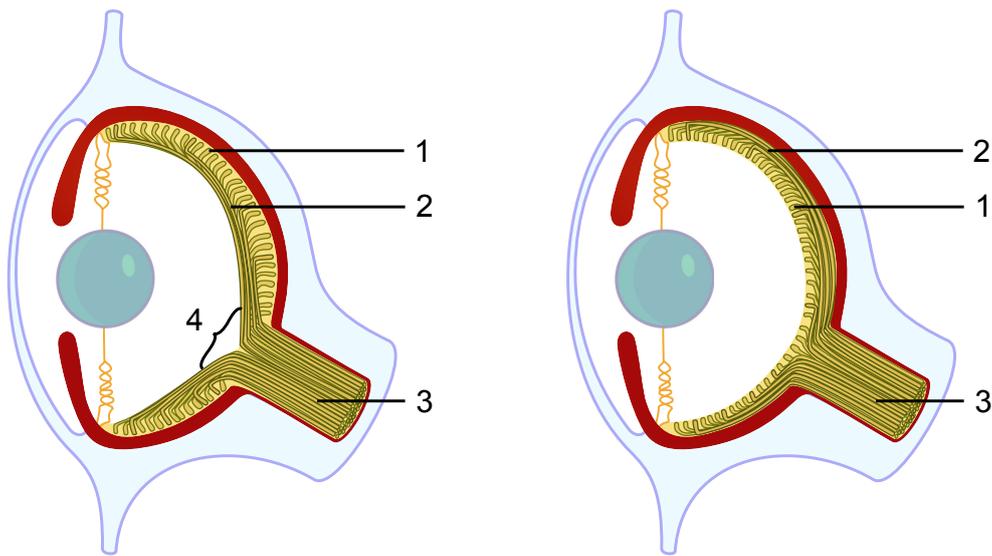


Figure 3: Left panel: Scheme of the structure of the vertebrate eye. Right panel: Scheme of the structure of the octopus eye. 1: Retina, 2: Nerve fibers, 3: Optical nerve, 4: Blind spot. From *Wikipedia*.

random association of atoms. Following their line of argument, let us evaluate the probability that placing amino acids at random in a linear sequence we obtain a working enzyme. Living cells have a repertoire of 20 amino acids and proteins are made of linear chains of amino acids (called polypeptide chains). A typical enzyme (a kind of protein whose function is to fasten, sometimes by several orders of magnitude, the unfolding of chemical reactions) is about a hundred amino acids long. If we suppose that only one particular amino acid sequence can make a functional enzyme, the probability to obtain this sequence by randomly placing amino acids will be of the order of one over the total number of amino acid sequences of length 100, i.e., $20^{100} \approx 10^{30}$. Of course this is a very pessimistic estimate, since it supposes that placing even only one wrong amino acid makes the sequence non-functional. But a more realistic calculation does not reach substantially different conclusions [1, p. 24]. If this is the case, it will be hard to escape Hoyle and Wickramasinghe's conclusion that life on Earth got started by structures of extra-terrestrial origin (whatever it may be) which had already solved the basic problems of biochemistry. And therefore that the passage from non-living to the living started somewhere else in the vast extension of the Universe, or even outside it.

We can recast this argument in the following way. Let us suppose, following Dawkins [2, Chap. 11], that 1000 evolution steps are necessary to obtain a working eye, and that at each step there are only two possibilities: the right or the wrong one. If the choice is made blindly, the probability of making the right choice at any step is equal to $\frac{1}{2}$. Then the probability of always making the right choice at every step is equal to $1/2^{1000} \approx 1/10^{300}$. It is clear that the fact that our lineage made the correct choice at each step looks nothing less than miraculous!

However, neither the eye nor the protein need to be perfect—in the sense of possessing their present form—in order to provide advantage to the organisms that

produce them. It is unrealistic, therefore, to evaluate the probability of producing them in the way that we just described. The eye is the result of a “Research and Development” process made by many life forms during long stretches of time, starting from the formation of small photosensitive patches, that came to be later organized in more complex and effective structures. And the more effective structures were retained because they provided advantages to the life forms that carried them—advantages that materialized as a more numerous offspring.

In the argument we just discussed we assumed that one does not know if one is on the right path until the end is reached. Let us suppose instead that each step made in the good direction provides a small advantage in terms of survival or fecundity to the being that makes it. More precisely, let us imagine to send in this “Garden of forking paths” a group of, say, 100 individuals, who perform their choice (right or wrong) at each step, and then reproduce. Let us assume that those who make the right choice at the right moment have slightly more offspring than those who make the wrong one: for instance, that for each offspring of an individual who made the wrong choice, there are 1.02 offspring (on average) of one who made the right choice. Thus, after the first step, we shall have 50 individuals on the right side and 50 on the wrong side and, after reproduction, 102 on the right side and 100 on the wrong side. This is surely a very small difference: the fraction of individuals on the right path is 51% rather than 50%. However, if we wait a few generations, the fraction of individuals on the right side will increase: after 10 generations the ratio of the number of offspring of an individual who had made the right choice to that of one who made the wrong choice will be about 1.22, and after 100 generations it will be about 7.25. Thus after 100 generations about 88% of the population will be on the right path. Imagine that at this point those who are on the right path face a new bifurcation. Half of them will take the right side, and half the wrong one. Thus we shall have 44% of individuals on the right path, 44% who made the wrong choice at the last fork, and 12% of those who had made the wrong choice at the beginning. After waiting 100 generations more, each individual who made the right choice twice in a row will have (on average) 7.25 more offspring than those who made the wrong choice at the last step, and each of them will have (on average) 7.25 more offspring than those who got their first step wrong. Summing up, there will be 88% of the population on the right path, a bit less than 12% having got the last step wrong, and less than 1% of those who got their first step wrong. Continuing this way, and assuming to be able to wait for 100 generations at each step, in 100 000 generations (a long time, but reasonable on the geological scale) our population will be made for 88% of individuals whose ancestors made all the correct choices, about 12% of those who got only the last step wrong, and the offspring of all the others will make a completely negligible fraction of the population. Looking at this path in retrospect it appears as extremely improbable, since its a priori probability is extremely small. However, the population took advantage at each step of the hints of the selection, so that its a posteriori probability is large enough. Moreover, it may well be that our guess on the a priori probability is too pessimistic, for two reasons: one, that it is not necessarily true that all the right choices must be made in a rigid sequential order, and two, that it is often possible to make functionally equivalent structures by following radically different paths, as shown by the existence of the octopus eye. Indeed, mollusks exhibit a number of different eye structures. Based on these considerations, a realistic estimate of the number of generations needed to develop the eye [3] yields a value of about 300 000 generations, a large number, but which requires a few hundred thousand years in a history of life that is estimated

to have lasted billions of years.

Evolution: The ladder and the tree.

In fact the path of evolution is not defined a priori but is an outcome of the evolutionary process itself. As stressed by Jacob [4],

Natural selection has no analogy with any aspect of human behavior. However, if one wanted to play with a comparison, one would say that natural selection does not work as an engineer works. It works like a tinkerer—a tinkerer that does not know exactly what he is going to produce but uses whatever he finds around him whether it be pieces of string, fragments of wood or old cardboards: in short it works like a tinkerer who uses everything at his disposal to produce some kind of workable object.

We can identify traces of this tinkering process by paying attention to those small revealing imperfections where some structures in living organisms appear excessively contrived with respect to what an engineer could have planned from scratch. A well-known example of this structure in mammals (and indeed, in all tetrapods) is the left recurring laryngeal nerve (cf. fig. 4), which performs a long detour under the aortic arch to end up connecting to structures under the larynx a few inches away from its branching point from the vagus nerve. The detour amounts to about 15' in giraffes! Comparative anatomy allows us to explain this odd fact. As we see in figure 5, one of the nerves connecting the fish brain to the gills has become the laryngeal nerve. In the fish this nerve lies behind a neighboring arterial arch (number 6 in the figure). During evolution this arch, that has become the ductus arteriosus, has kept a position close to the heart, and therefore the nerve, which passes behind it and could not cross the arch has been forced to grow longer and longer. This has even more striking implications, as argued in [5]:

The recurrent course of the nerve from the brain, around the great vessels, to the larynx, is shared by all extant tetrapods. Therefore we may infer that the recurrent laryngeal nerve was present in extinct tetrapods, had the same developmental origin, and followed the same course. The longest-necked animals of all time were the extinct sauropod dinosaurs, some of which had necks 14 meters long. In these animals, the neurons that comprised the recurrent laryngeal nerve were at least 28 meters long.

Evolving life forms must modify their structures along the way, without preventing the working of the organism. This is as paradoxical as trying to change an aircraft's propellers during the flight. Therefore, in the process, some less relevant aspects can be neglected with respect to more urgent ones: thus in the giraffe, the cost of lengthening the nerve a few millimeters at each time can be neglected with respect to the advantage of having a longer neck, what allows the animal to reach the leaves of higher trees.

As any DIY fan knows well, the pleasure of tinkering does not lie in building up things according to instructions, as when assembling IKEA® furniture, but in inventing new usages for the objects in one's possession. In some cases, these objects solve, possibly in an original way, already present problems: thus one can use old telephone books to build a stool. In other cases, an unexpected combination of tools allows one to build up an instrument that did not exist before: possibly the best known example is the way in which Gutenberg made use of the wine press to press paper sheets against the types. Working as a tinkerer, evolution shows some

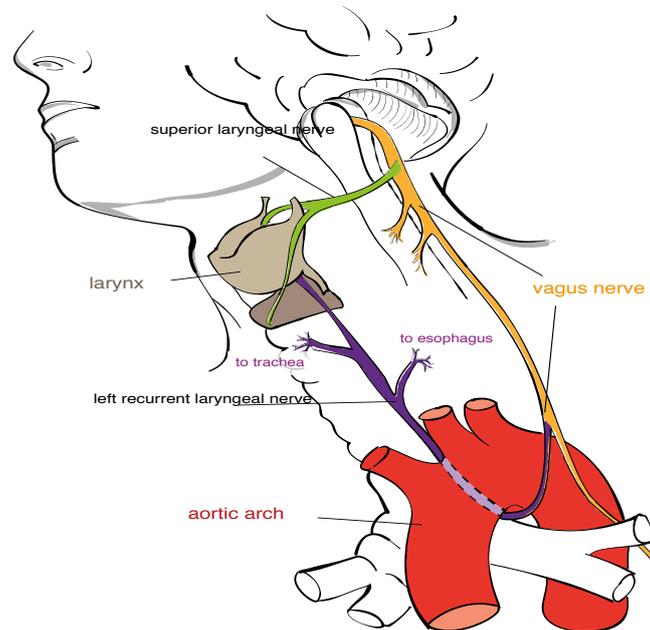


Figure 4: The left recurrent laryngeal nerve is a branch of the vagus nerve, which connects with almost all intrinsic muscles of the larynx. Its name derives from the recurrent path it performs with a descent to the thorax, followed by a rise upwards, in the neck, to the larynx. From *Wikipedia*.

similar phenomena. Here is how Jacob [4] summarizes the theory of lung evolution in terrestrial vertebrates, according to Mayr [7]:

Its development started in certain freshwater fishes living in stagnant pools with insufficient oxygen. They adopted the habit of swallowing air and absorbing oxygen through the walls of their esophagus. Under these conditions, enlargement of the surface area of the esophagus provided a selective advantage. Diverticula of the esophagus appeared and, under continuous selective pressure, enlarged into lungs. Further selection of the lung was merely an elaboration of this theme—enlarging the oxygen uptake and vascularization. To make a lung with a piece of esophagus sounds very much like tinkering.

This shows how the tinkering evolution can lead by trial and error to a major innovation, creating an organ that did not exist before, like the lung.

Coming back to the example of the garden of forking paths, we were perhaps too pessimistic in assuming that a single wrong path choice at a fork would necessarily lead to ruin. Since there are so many combinatorial possibilities in living life forms, it may well be the case that the path leads—after a great deal of branching—to a different working structure, possibly with a different function. And we can realize that this is possible by looking at the prodigious variety of life forms that exist at present. Gould [8] remarks that

The model of the ladder is much more than merely wrong. It never could

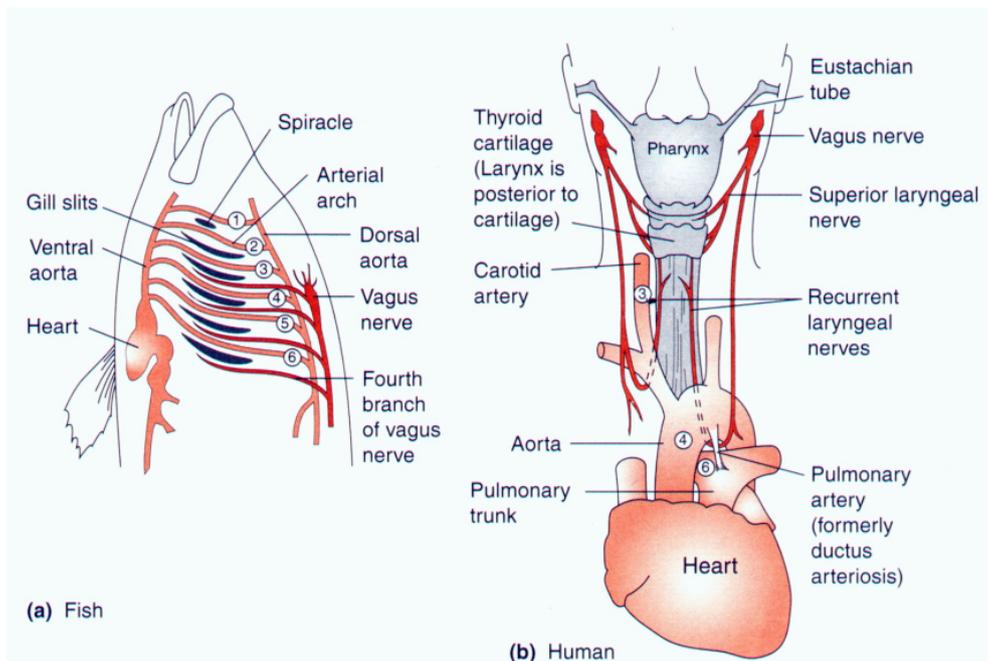


Figure 5: Scheme of the evolution of the recurrent laryngeal nerve, from fishes to mammals. The ancestor of this nerve passed close to the arterial arch marked with 6 in the image. This arch has become the ductus arteriosus in mammals. The branches between arch 3 and 4 and between arch 4 and the last one are no more present in mammals. From [6, p. 45]

provide the promised illustration of evolution progressive and triumphant – for it could only be applied to unsuccessful lineages.

This is the case of the equine family, which exhibits at present no more than six species (of which only one of horses), and of the *Hominidae* family, the great apes, which contains eight species: three of orangutans, two of chimps, two of gorillas, and *Homo sapiens*. Many other families cannot fit into a linear scheme. Among the mammals, the *Mustelidae* family (which contains the badger and the ferret) contains at least 57 species, the *Cricetidae* one 600, and the *Muridae* family (which contains mice and rats) more than 700. And we can also think at the vast variety of species of insects. A possibly apocryphal story reports that the evolutionary biologist Haldane answered in this way a theologian who had asked him what he had figured out about the Creator by investigating the creation:

The Creator, if He exists, has an “inordinate fondness for beetles”.

Haldane was referring to the fact that there are more than 300 000 species of beetles against no more than 9000 species of birds and slightly more than 10 000 species of mammals. Thus, if it may make some sense to look at the evolution of the horse starting from the arrival point (but it is a bit arbitrary to start from the horse, rather than from the zebra or the humble donkey) it will make much less sense to describe the evolution of beetles starting from a randomly chosen present species.

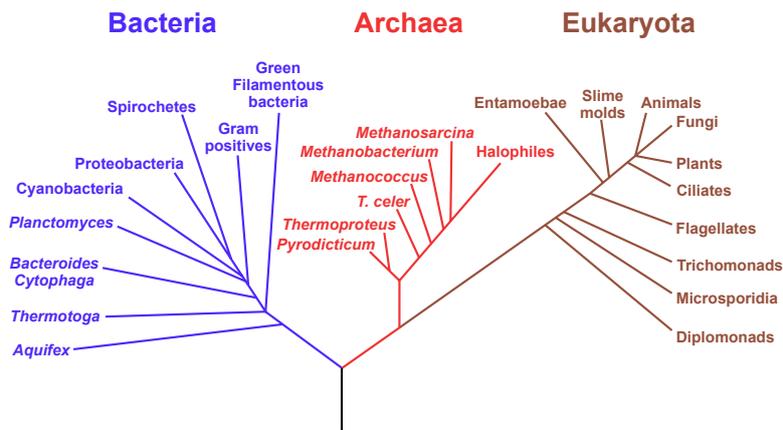


Figure 7: A modern schematic view of the phylogenetic tree of living forms. From Wikipedia.

the *Fungi* (the proper mushrooms) just some terminal twigs of the same branch. The three branches now correspond to *Eukaryota*, namely organisms made of cells which possess a well-identified nucleus, and which encompass all multicellular organisms and many unicellular ones; *Bacteria*, unicellular organisms, which also encompass most of the agents causing infectious diseases (viruses are not considered living beings in this classification); and *Archaea*, a large group of unicellular organisms that differ from all other life forms by some fundamental aspects of the cell membranes and of the mechanisms of protein synthesis. Many archaea are extremophiles, that is they can survive at high temperatures and pressures, where ordinary bacteria do not survive. It is thought therefore that they were better suited to thrive in the environment of the young Earth, and that therefore they are older than the bacteria, what justifies calling them *Archaea*. Although the existence of this branch was not suspected as recently as about 50 years ago, it is now believed to be one of the major actors in evolution. According to current opinion, as shown in the figure, the eukaryots branched from the archaea rather than from the bacteria. It has been suggested that the eukaryots originated from cases in which some archaeas (very different from their present descendants) formed symbiotic structures with bacteria that they had engulfed. This stresses how many surprises have been found, also recently, in the study of the fundamental phylogenetic organization of life.

We can summarize this discussion by stressing two main points: one, that an evolutionary history that appears extremely improbable a priori looks much more probable a posteriori, looking back from the arrival point, since a constant selective pressure has acted during the whole path, favoring advantageous variants with respect to disadvantageous ones. Stated otherwise, the a posteriori probability of a given evolutionary path is a conditional probability, *conditioned by the fact that the concerned life form is presently alive*. It is thus very different from the a priori probability. The second point is that one should not forget that the organisms are not constrained to follow a preexistent path, but can follow a great number of viable paths, as shown by the great variety of life forms.

The random nature of natural selection.

A number of probabilistic techniques have been developed since the thirties of last century to understand the details of the evolutionary process. They form the discipline known as Population Genetics. The best known exponents of this discipline are R. A. Fisher (British), S. Wright (American) and M. Kimura (Japanese). One can consider, for instance, a population formed by a fixed number of individuals which evolves according to the Darwinian mechanism of reproduction, selection and mutation. The simplest model that describes this situation is called the Wright-Fisher model, and is schematically represented in figure 8. Columns represent the

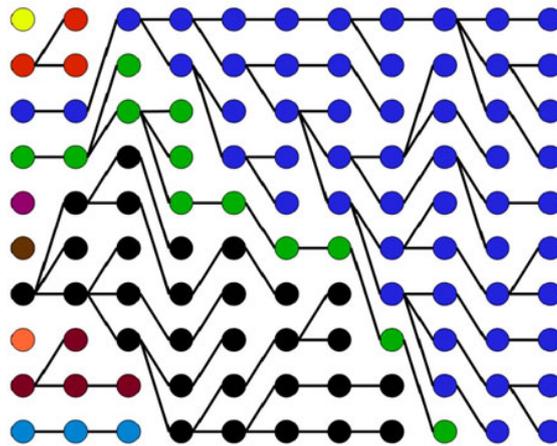


Figure 8: Scheme of the Wright-Fisher model of the evolution of an asexually reproducing population. Columns correspond to successive generations of a population of 10 individuals. Different phenotypes are denoted by different colors. One assumes that each individual carries the same phenotype as its parent. Each individual is connected to its parent (in the previous generation) by a black line.

composition of the population in successive generations, from left to right. The colors represent the features (phenotypes) of the individuals that form the population: different features correspond to different colors. The fitness of a phenotype is measured by the average number of fecund individuals that an individual possessing that phenotype can produce in the given conditions. One assumes an asexual reproduction mechanism and that the phenotype of the offspring is equal to that of its parent, except for mutations (in the represented case there appear no mutations). The new generation is obtained by choosing at random a parent in the previous generation (with probability proportional to its fitness). The parent-child relationship is represented by a black line in figure 8. Due to inevitable fluctuations, some individuals (like the yellow or the magenta ones in the first column) will have no children, while others (like the red or the black one) will have more than one. (In the represented case, all phenotypes have the same fitness.) Iterating the process, we see that some phenotypes tend to invade the population: in the figure, this happens before with the black phenotype, and eventually with the blue one, which covers the whole population starting from the tenth generation. Kimura has stressed that

this phenomenon (called fixation of a trait) can take place in populations made of a small number of individuals even if the phenotypes are equivalent from the selection viewpoint. There is even a small but non-vanishing probability that a sub-optimal phenotype undergoes fixation. Since many populations in nature are made of small numbers of individuals, Kimura suggested that a large part of the diversity of life forms is due to the random fixation of different but selectively equivalent variants in different populations. Figure 9 represents the evolution of the frequency of the

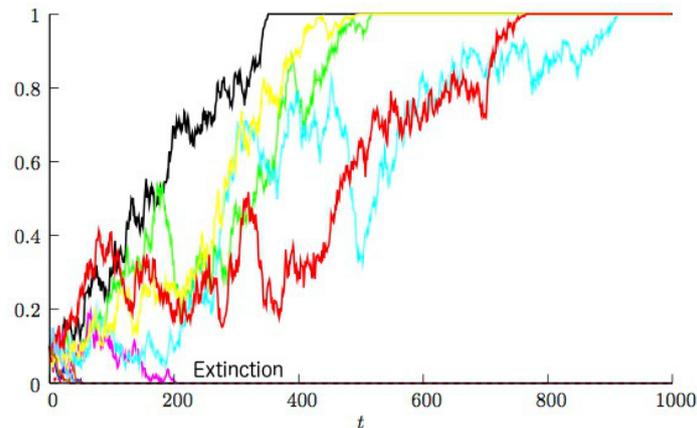


Figure 9: Evolution of the frequency of the optimal phenotype in a Wright-Fisher model. Population of 500 individuals. The process has been repeated 10 times, starting from the same initial condition. The optimal phenotype has a 1% larger fitness than the others. The initial frequency of the optimal phenotype is 10%. The optimal phenotype fixates in 5 out of 10 cases, and goes to extinction in the other cases.

optimal phenotype in a population of 500 individuals. The fitness of the optimal phenotype is larger by 1% with respect to the others, meaning that an individual that carries it has 1% more offspring (on average) than one that doesn't. The initial fraction of individuals carrying the optimal phenotype is 10%. In this situation we observe that the optimal phenotype goes to fixation within 1000 generations in 5 cases out of 10, and gets extinguished in the other cases. Thus natural selection has a positive effect for this case only 50% of the times, due to the small population size. Actually one of the most relevant results obtained by Kimura is that if in a population there is a single mutant having a fitness advantage, say, of 1% with respect to the remainder of the population, the probability that the trait eventually goes to fixation is equal to its relative advantage, that is, in our case, is of 1%. (Or even, taking into account that in sexually reproducing populations each individual has two copies of the genome, and the mutation appears on only one of them, this probability will be only half of 1%.)

Is evolution predictable?

Given that sometimes even natural selection is not effective, what hope can we have to predict evolutionary processes? Or should we be contented to use it to explain a posteriori the properties of living beings, without daring to frame predictions?

Actually predictions based on the theory of evolution have been stated and verified many times. In a very interesting case, first Darwin and later (independently) the other discoverer of natural selection, A. J. Wallace, made a precise prediction. This prediction was verified with complete satisfaction only many years later and the conclusion was reached a few years ago, in 1992, with the discovery of the “smoking gun”. This is a story worth telling [10]. As we know, pollination plays an essential role in the life cycle of flowering plants, and they often produce nectar to lure insects into visiting their flowers. It is clearly advantageous for the plant to develop mechanisms that help the visit of a particular species of insects rather than another, because this increases the probability that their precious pollen is carried to flowers of the same species. Under the pressure of evolution, many flowering species have produced elaborate ways to discriminate among the different species of insects that collect their nectar. Darwin dedicated one of his treatises to “The Various Contrivances By Which British And Foreign Orchids Are Fertilised By Insects”. Knowing of Darwin’s interest in orchids, an orchid grower sent him in 1862 a box containing in particular the beautiful star-shaped flower of *Angraecum sesquipedale*. The name is an allusion to the fact that the nectar is located at the bottom of a tube almost a foot long. In a letter to a friend, Darwin wrote “Good Heavens what insect can suck it” and in another letter a few days later he suggested that “in Madagascar there must be moths with probosces capable of extension to a length of between ten and eleven inches [25.4–27.9 cm]”. Darwin published this prediction in his book on orchids, published the same year. In 1907, more than 20 years after Darwin’s death, a subspecies of the moth *Xantopan morganii*, already remarked in Congo by its long proboscis, was identified in Madagascar. This moth has a wingspan of about 16cm, but its proboscis is truly colossal, more than 20cm long, and it forms a large coil before the head when not in use. Given Darwin’s prediction, this subspecies was called *Xantopan morganii praedicta*. We had however to wait until 1992, about 130 years after Darwin’s prediction, to be able to observe moths of this subspecies in the act of sucking nectar from *Angraecum*’s flower and of carrying its pollen from plant to plant (cf. fig. 10). *Angraecum sesquipedale* is now known as “Darwin’s orchid” among the aficionados.

In the recent years there has been a growing effort to go from reconstruction and analysis of the evolutionary processes to the prediction of their future behavior. The interest is not only academic. We can produce an effective influenza vaccine only if we have reliable predictions on the evolutionary dynamics of its virus. Indeed, the evolution of influenza virus exhibits two apparently contradictory features: on the one hand its mutation rate is so high that the dominating strain changes substantially every few years, and on the other hand in each epidemic season most infections are due to closely related strains. This second feature allows the production of an effective vaccine, provided the dominant strain has been identified with an advance long enough to allow the production of enough doses. A group of the World Health Organization (WHO) is dedicated to the identification of the prevailing strain for each coming epidemic season. It exploits world-wide collected data on the prevalence of the different viral strains and on the effectiveness on the different strains of the immunity distributed in the population due to the previous epidemics. [11] Every six months, the WHO publishes its recommendations for the vaccine composition, about six months before the beginning of the epidemic season for each hemisphere.

How can one make predictions for an evolutionary process, when it is determined by so many random factors? When only few data were available, mostly



Figure 10: Caught in the act! *Xantopan morganii praedicta* sucks *Angraecum sesquipedale*'s nectar. From [10].

referring to macroevolutionary processes (species evolution), it was natural to be skeptical about the possibility of repeating and therefore of predicting such processes. But things are now changing, since we can now have high throughput data on sequences and on phenotypes, we can perform experiments in parallel on many samples, and we are developing more effective methods to analyze the behavior of complex dynamical systems. This opens the way to the possibility of formulating and validating predictions about evolutionary processes on an accessible time scale: and this could help us to solve important problems like the selection of the influenza vaccine, but also to fight against the development of antibiotic resistance (especially in hospital environment) or in the prediction of cancer development in a single patient: since cancer evolution reproduces at a small scale and in a shorter time span a process of Darwinian selection.

While it will probably be always impossible to predict the evolution of the *genotype* of life forms, it could be possible to attain a less ambitious goal, namely to predict—within limits—the evolution of their *phenotypes*. The genotype of an individual is the whole of its inheritable characteristics: in practice, it is represented for a good part by its genome, that is by its DNA sequence. As we know the DNA is described by a very long sequence written in a four-letter alphabet: A, T, C and G. In the simplest organisms, this sequence is already a few million letter long, and the number of possible variants is enormous. Even if we consider just those variants with only a few differences from a given sequence, the number of possible variants is already of the order of the genome length. Many of these variants are irrelevant, because they do not modify the phenotype of the organism, that is the whole of its

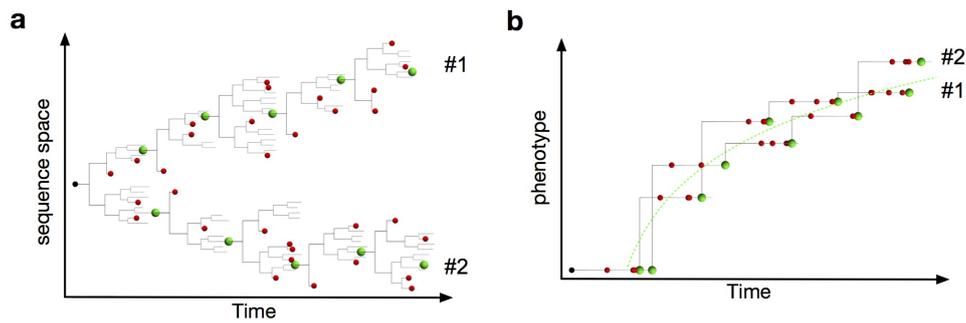


Figure 11: Schematic representation of a selection-mutation process in genotypic and in phenotypic space. Comparison of the evolutionary process in two parallel populations in two different representations. Left: Evolution in sequence (genotype) space is a stochastic process with many equiprobable pathways. Mutations produce new variants, represented as nodes in a tree. These variants are only weakly checked by selection. Those positive selected (advantageous) are represented by green dots, and those negatively selected (disadvantageous) by red ones. The two populations follow diverging paths. Right: Fitness and other phenotypic traits evolve in a regular way. The pathways shown contain the same mutations (green and red) as on the left panel, but now different sequences produce similar phenotypic effects. Positive and negative selection effects lead the evolution towards a single phenotypic path, represented by the green line. The populations evolve in parallel close to this path. From [12].

observable inheritable characteristics. But the observable features are the ones that fall under the edge of natural selection. This point is schematically represented in figure 11, which has been published in a recent paper on the predictability of evolution. An experiment made by a group in Harvard on the common beer yeast *Saccharomyces cerevisiae* [13] has shown that this is not just a theoretical construction. In this experiment 640 yeast strains have been allowed to evolve in parallel for 1000 generations under identical conditions. Their fitness was measured from time to time by measuring their growth rate and comparing it to that of the ancestral strain. Thus the experimenters could evaluate the fitness variations among the different strains, and how the fitness of a strain depended on the fitness of its ancestor. Since natural selection acts on the phenotype it is quite reasonable that evolutionary dynamics is more predictable at the phenotypic rather than at the genotypic level. Exploiting this concept it has been possible to develop a predictive model [14] of the evolution of the influenza virus A/H3N2 which should allow to improve on the WHO method, reaching a predictability horizon of the order of one year.

One of the founders of the evolutionary synthesis, Dobzhansky, famously stressed that “Nothing in biology makes sense except in the light of evolution”. [15] We can add to this that evolution cannot be understood except in the light of probability theory. And that the tools of probability theory can already shed light on evolutionary processes that touch us directly and hopefully will help us to control them in the near future.

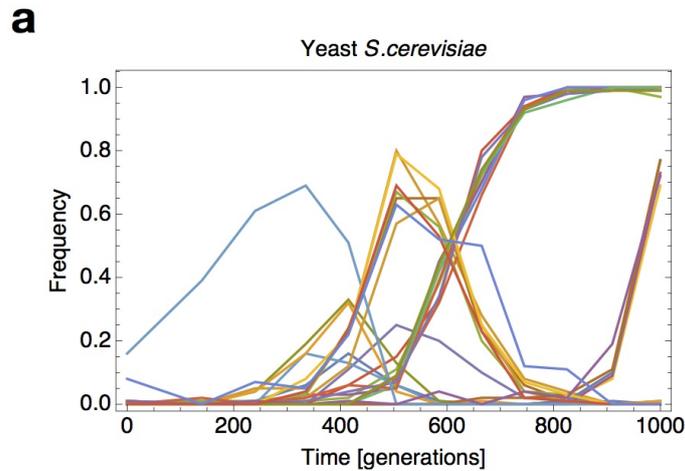


Figure 12: Frequency of several variants in *Saccharomyces cerevisiae* populations evolving in parallel. Although the variants are genotypically different, one can identify some similarities in their dynamics. Typically some variants first grow in frequency, but are later overcome by newly arising ones. This shows that parallel evolution exhibits some degree of predictability. From [12], on data by [13].

References

- [1] Hoyle, F. and Wickramasinghe, N. C., *Evolution from Space: A Theory of Cosmic Creationism* (New York: Simon and Schuster, 1981).
- [2] Dawkins, R., *The Blind Watchmaker* (London: Penguin Books, 1986).
- [3] Nilsson, D.-E. and Pelger, S., A pessimistic estimate of the time required for an eye to evolve, *Proc. R. Soc. Lond. B* **256** 53-58 (1994)
- [4] Jacob, F., Evolution and tinkering, *Science* **196** 1161-1166 (1977)
- [5] M. J. Wedel, A monument of inefficiency: The presumed course of the recurrent laryngeal nerve in sauropod dinosaurs, *Acta Palaeontologica Polonica* **57** 251-256 (2012)
- [6] M. W. Strickberger, *Evolution* (3rd ed.) (Sudbury MA: Jones and Bartlett, 2005)
- [7] Mayr, E., From molecules to organic diversity, *Federation Proceedings* **23** 1231-1235 (1964), reprinted in: Mayr, E. *Evolution and the Diversity of Life* (Cambridge (Mass.): Belknap, 1976) cited by [4]. For a technical description, see Packard, G. C., The evolution of air-breathing in Paleozoic gnathostome fishes, *Evolution* **28** 320-325 (1974). More recently, this model has been the target of some criticism. See, e.g.: Farmer, C., Did lungs and the intracardiac shunt evolve to oxygenate the heart in vertebrates? *Paleobiology* **23** 358-372 (1997)
- [8] Gould, S. J., Life's little joke, in: Gould, S. J., *Bully for Brontosaurus: Reflections in Natural History* (New York: Norton, 1991).

- [9] Kimura, M., On the probability of fixation of mutant genes in a population, *Genetics* **47** 713-719 (1962)
- [10] Arditti, J., Elliott, J. Kitching, I. J., and Wasserthal, L. T., 'Good Heavens what insect can suck it' – Charles Darwin, *Angraecum sesquipedale* and *Xantopan morgani praedicta*, *Botanical Journal of the Linnean Society* **169** 403-432 (2012)
- [11] One can find the description of the protocol used by the WHO at the URL http://apps.who.int/gb/pip/pdf_files/Fluvaccvirusselection.pdf?ua=1
- [12] Lässig, M., Mustonen, V., and Walczak, A. M., Predicting evolution, *Nature Ecology & Evolution* **1** 0077 (2017)
- [13] Kryazhimsky, S., Rice, D. P., Jerison, E. R., and Desai, M. M., Global Epistasis Makes Adaptation Predictable Despite Sequence-Level Stochasticity, *Science* **344** 1519-1522 (2014)
- [14] Łuksza, M., and Lässig, M., A predictive fitness model for influenza, *Nature* **507** 57-61 (2014)
- [15] Dobzhansky, Th., Nothing in Biology Makes Sense Except in the Light of Evolution, *American Biology Teacher* **35** 125-129 (1973)