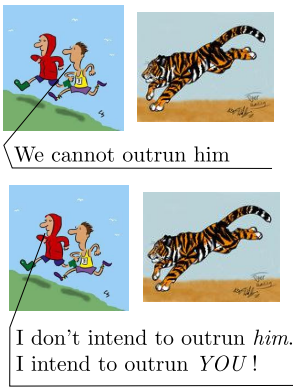
Grenoble Universités
Département de Physique.

Lectures on Mathematical Foundations of Darwinian Evolution.

Bahram Houchmandzadeh



http://houchmandzadeh.net/cours/Evolution/evoltion.html
First Version : December 21 2012
Present Version : October 3, 2017

Contents

1 Introduction.

As it is well known, the foundation of the theory was laid out by Darwin and Wallace in the mid XIXth century, when people had no idea of how inheritance worked. However, the basic principles which were enunciated at this time constitute our basic understanding of how evolution works. These principles are:

1. For any trait z considered (production of β-Gal enzyme by the bacteria *E.Coli*, the top speed reached by the tiger, and anything in between), a population of individuals will show *variations*. We can describe/measure the distribution of this trait in the population which we call $p(z)$.

2. Some trait are under selective pressure : individuals having their trait around some specific value z_m are more apt to pass their progeny to the next generation. The aptitude of some individuals compared to others (or the mean of the population) is called their *fitness*.

3. Offspring resemble their parents, which means that the progeny of parents with trait z_1 will have a value of the trait close to z_1. If we think in terms of sexual reproduction, the progeny of parents with traits z_1 and z_2 will have a trait value distributed according to some function $L(z; z_1, z_2)$, where the function L measure the closeness to their parent.

4. As a consequence of the above principles, fitter individuals will increase their proportion in the population as time flows and new generations replace older ones.

Point 4 is the essence of Darwinian evolution. However, as we will see, these principles stated above are still vague enough to let a lot of space for interpretations. Let us discuss some of these issues before getting any further.

First of all, what is the origin of *variations* ? We can think that variations happen in random and preexist any selection, or they happen because of selection, individuals changing their trait (with more or less success) in order

to be fitter. The second idea is nowadays attributed to Lamarck[1], the first to Darwin. Darwinism won the day and Lamarckism was discarded as one of the big error of science[2]. The first experimental proof however that variations preexist selection was provided by Luria & Delbruck in 1943[3], and even this point was still contested by Cairn in 1988[4].

The above discussions concerned the genes. But are all the traits controlled solely by genes ? The answer is no and there are some behavior (think gene expression level) that vary between isogenic individuals and can be transmitted to the next generation. The first striking proof was provided in the production level of the βGal enzyme in *E.Coli* by Novick and Weiner[5] in 1957 and the discipline now called *epigenetics*[6] is a fast developing field. We will limit however this lecture to the classical view : traits are controlled only by genes and variations happen because of random mutations. These terms will of course be defined precisely when we get to the heart of the matter.

The second point constitute another dogma of the evolutionary theory : selection happens at the individual level. Let me repeat this point: SELECTION HAPPENS AT THE INDIVIDUAL LEVEL[7]. What we mean by that is that in a more or less stable environment, when the total number of individuals is more or less constant, the only way for an individual to increase its proportion (the number of its progeny in further generations) is by producing more viable offspring than its neighbors. This a very strong assumption. First of all, what is an individual ? If you think of multicellular organism, then why not apply the same rule to the cells constituting the organism[8] ? And then, at the level of each cell, you have many genes in a genome, why not applying the rule to individual genes competing against each other[9] ? At

[1] Jean-Baptiste Lamarck, 1744-1829 ; Charles Robert Darwin,1809-1882.

[2] The Godwin point in Evolution literature is reached when one scientist calls the other Lamarkian.

[3] Luria, S. E.; Delbrück, M. (1943). "Mutations of Bacteria from Virus Sensitivity to Virus Resistance". Genetics 28 (6): 491–511.

[4] Cairns, J.; Overbaugh, J.; Miller, S. (1988). "The Origin of Mutants". Nature 335 (6186): 142–145

[5] Novick, A. & Weiner, M. (1957) ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON, Proc. Natl. Acad. Sci. USA 43, 553–566

[6] They should in all honesty call the field Lamarckism, but the term is too negatively loaded.

[7] Two people are trying to escape a tiger. The first one realizes the futility of their trial and inform the second one than there is no chance they could outrun the tiger. The second runner informs the first one that he has no intention of outrunning the tiger, but only outrunning his companion.

[8] This will bring us for example to look at cancer from an evolutionary point of view.

[9] The human genome for example contains 50% of repeating elements called LIN and

higher scale, some individuals associate to form tribes and hordes. Shouldn't we take tribes competing against each other as the level of selection ? To these days, the level of selection is one the hottest topics debated[10] among evolutionist and we will hit part of these debates in this lecture.

Beside the level of selection, what do we mean by more or less stable environment ? What if the behavior of the individual under selection modifies the environment ? We will see that this last part will bring us to cooperative behaviors which are forbidden in the fundamentalist formulation of evolution[11]. We will also see that maybe the first and second point are not independent and the rate of mutation itself can be a trait under selection which can depend on the variability of the environment. This will also be discussed here.

Another vague term in point two was "more apt to pass their progeny to the next generation". The pre-modern scientists interpreted this sentence in a very deterministic way : if you run faster (than your neighbors escaping the tiger), you will have more progeny. The deterministic approach is still widespread, even if it became clear early (around ~1920) that this should be interpreted in a probabilistic way : more apt mean that you will have a higher *probability* to have more progeny, but bad stuff happens : even if you run 0.01m/s faster than your neighbor, you still have a chance to fell prey to the tiger and your neighbor being spared. The probabilist approach will be the core of this lecture.

Point three provoked the first major crisis in the theory of evolution. As we mentioned, at the time of formulation of the theory by Darwin and Wallace, the exact mechanism of heredity was unknown and the dominant model was the blending model : we are the average of our parents[12]. Jenkin[13] noted that this point is in contradiction with the first point : variations will disappear extremely fast in the population. To see this, consider the trait in the progeny Z as the average of the parent $(X + Y)/2$. Then the Variance in the second generation will be

$$\mathrm{Var}(Z) = 1/4(\mathrm{Var}(X) + \mathrm{Var}(Y)) = 1/2\mathrm{Var}(X)$$

SIN, which are believed to arise just by such phenomena. Genes in conflict constitute another field of study which we will not get into in these lectures. See for example "Genes in Conflict: The Biology of Selfish Genetic Elements" by Austin Burt.

[10] scuffled is maybe a more appropriate description

[11] Thomas Henry Huxley (1825-1895), also called Darwin's bulldog, is a fine first example of fundamentalist evolutionist

[12] Even though the major form of life on earth is bacterial, most of early thinkers were concerned by sexually reproducing multicellular organisms.

[13] Henry Charles Fleeming Jenkin, 1833–1885

so at each generation, we will lose half of the variability of the population. In order to maintain the variability observed in natural population, a huge amount of mutations will be needed at each generation, in which case offspring will not resemble their parent. We know today that the weak point of the argument is not evolution but the blending theory. This was discovered by Mendel[14] and saved the day : the heredity information is not continuous, but quantified. We now call the unit of hereditary information a gene. The work of Mendel was rediscovered around 1900, at the same time that physicists healed *their* ultraviolet crisis by another quantification trick. The fusion of Mendelian and Darwinian models took some time, but by around 1920 it was achieved and the modern evolutionary theory took the name of "new synthesis". We will discuss the matter in some detail below.

So in the following, we will discuss all these matters and many more in some details. The field is extremely vast and we will have to make choices and speak about some selected topics. But the framework will be similar for many other topics for which countless books exists. So let us get started.

[14]Gregor Johann Mendel 1822-1884.

2 Biological foundations and lexicology.

The genetic information, for most organisms, is stocked into a linear polymer called DNA. The reader knows that for most organisms, DNA is present in a double stranded form (the famous double helix of Watson and Crick) which we call chromosome. The chromosome is like a book made of letters and words (the ATCG alphabet), divided into chapters, each chapter containing informations about making a particular protein[1]. Each chapter is called a *gene*. A simple bacteria such as *E.Coli* has around 2000 genes, a fruit fly around 10000 genes and a human around 25000 genes. During replications, in some genes, some letters are modified, an 'A' being replaced by a 'C' for example. These are called *mutations*. Therefore, different individuals will carry slightly modified version of the original gene. The various flavors of a given gene are called its allele. So, for example, for three genes A,B,C, one individual of a given species will have *genotype* $A_1B_1C_2$ and an other of the same species will have genotype $A_2B_1C_3$ and so on, where gene A has for example three allele A_1,$A_2$and A_3. Genes are organized linearly on the chromosome, and the physical position of a gene on the chromosome is called its locus.

Some genes produce a protein which is put directly in use. For example, a gene call $\beta-Gal$ produces a protein called beta-galactosidase which is used by *E.Coli* to digest a sugar called Lactose. Some other genes produce proteins which will regulate the production of protein by other genes. These proteins are called transcription factors. In general, a mutation in a transcription factor has much wider effect than a mutation in a 'worker' protein.

Living organisms and computers are both information processing machines[2]. The way they execute their program has striking similarities and

[1]This view has been much extended. Some chapters of the chromosome contain information only about making RNA which is not going to be translated into protein, but plays directly a regulating role. On the other hand, part of the information is contained outside of the book (chromosome) in the form of methylation/acetylation of the DNA or the histone core.

[2]The smart name should be Turing machines.

differences (Figure 2.1). In a computer hard disk, which is a linear storage device, there is a small table at the beginning called the File Allocation Table (FAT). It stores the position of each file and its length. When the CPU is asked to execute the ooffice program, the CPU look up the table, find the position of the oofice file, move the pointer to this position, and from this position, load up the given amount of bytes into memory where it can get executed. In a living organism, the executable programs are called genes on the linear storage device called DNA and executing a program means producing the corresponding RNA[3] and eventually the protein[4]. The is however no FAT and the position of the gene along the chromosome has no importance[5]; instead, each gene has a tag (called a promoter) inducing the polymerase to begin producing mRNA at this position. Moreover, the tag are more or less powerfull, attracting weakly or strongly the polymerase and therefore producing more or less number of copies of the mRNA. Finally, product of some genes can attach around the promoter of another gene and modulate its force.

We see that by this mechanism, a mutation on one gene can have large effect on the organism if the targeted gene was for example an important transcription factor regulating many genes. Going from mouse to humans needs few mutations on the genes controlling the developmental program.

A haploid organism transmits its chromosomes nearly perfectly to its progeny. Some mutations are introduced during the DNA duplicating process because after all, the machinery is governed by the law of statistical physics and errors have probabilities of around $\exp(\Delta E/kT)$, where ΔE is the energy difference between incorporating a wrong A for example instead of a T, which is few hydrogen bounds[6].

In a diploid mechanism, two copies of each chromosome are present in each cells. each one is provided by one parent, and the two copies are slightly different, carrying some different alleles of the genes. Before transmitting the one copy of the chromosome to its progeny, the two copies are reshuffled

[3] By a machinery called RNA polymerase

[4] Some genes produce only RNA which is used in itself as a piece of information to modulates other genes activity.

[5] This is a big approximation. The position of many genes relative to each other is important and the effect is known as neighborhood effects. In bacteria for example, tightly related genes are packed into an "operon" and genes functioning together are spatially related.

[6] In fact, the nake energy difference will give an error rate of 10^{-4} per copied base. The actual error rate is much lower, in the range of 10^{-9}, as the DNA duplication implies many proofreading mechanism. Highly mutable viruses, in particular RNA viruses such as AIDS, have an error rate close to the 10^{-4}.

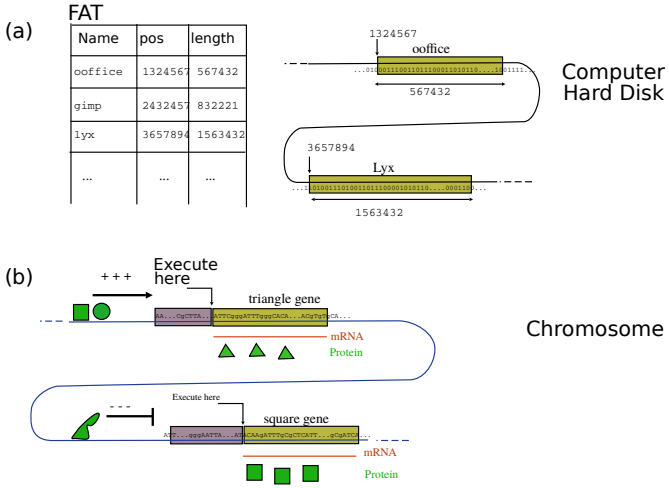


Figure 2.1 – Comparison of a computer (a) and living organism (b) information processing.

in a process called recombination.

Time lines for various discoveries.

1860's: The existence of genes as discrete entity was postulated by Mendel.

1900: Mendel was rediscovered by de Vries, Correns and Tschermak around 1900 AD.

1902: Chromosomes, discovered in the 1880's by Boveri, were postulated to be the carriers of genes.

1920: Morgan postulated that genes were physically separated along the chromose and were mapped from statistical observations of linked traits.

1943: Mitotic chromosomes observed under microscope are a mixture of DNA and proteins. Proteins were supposed to be the carrying of heredity information, until Avery transformed bacteria by DNA injection and showed this latter molecule to be genes carrier.

1953: The race to discover the structure of DNA was won by Crick and Watson, who benefited from the work of Rosalind Franklin.

1960: Jacob et Monod discovered the principle of genetic regulation.

1960: Jacob et Monod discovered the existence of temporary molecule mRNA as the intermediary between DNA and protein. Jacob formulated the "Dogma".

1962: In a series of theoretical thinking and experiments involving Gamov (and NSA cryptographers), Crick and Benner and finally Nirenberg, the genetic code was unravelled.

1960-1980: The cut and past tools of molecular biology (restriction enzymes, polymerase, ligase, ...) were discovered and the first recombinant DNA made.

1983: Mullis developed the PCR technique. Molecular Biology entered its explosive development phase (which continues).

3 A short introduction to stochastic processes.

- concept of probability (counting), the mean and the variance.
- transition rates, master equation, equations of the mean, variance, ...
- Solve explicitly the multiplicative noise of growth.

4 Phenotypes and genotypes.

To the people of XIX$^{\text{th}}$ century, all organisms were sexual. This in turn, made the field of evolution much more complicated and made it necessary to introduce *two* level of information organization : genotype and phenotype. The dogma of the new synthesis then was stated as "phenotype is selected, genotype is inherited". The origin of this problem is as follow. We will tackle again this problem in much more detail in chapter 11.

Let us move back to 1870 and forget about genotype. Individuals have traits quantified by the variable z (height, speed, ...) and the traits is distributed in the population according to the function $q(z)$: $q(z)$ is the proportion of the population possessing a trait value in $[z, z + dz[$.

Let us now suppose the validity of the blending theory of heredity which states that progenies are the average of their parents. If $q_0(z)$ is the distribution of the trait in the parental population, then $q_1 z)$, the distribution of the trait in the progeny is

$$q_1(z) = 2 \int_I q_0(z_1) q_0(2z - z_1) dz_1 \qquad (4.1)$$

which is just another way of saying that the father, mother and child random variables X, Y, Z are related through

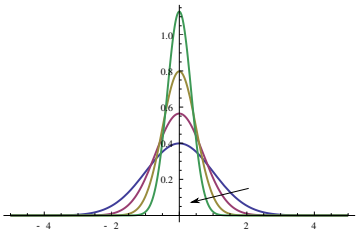


Figure 4.1 – Distribution $q_i(z)$ for 4 successive generations, according to the blending theory.

$$Z = (X + Y)/2 \qquad (4.2)$$

Eq. (4.1) implies that $q_n(z) \to \delta(z - \bar{z})$, where $\bar{z} = \int z q_0(z) dz$ is the mean in the initial generation. This is easily demonstrated by observing that in Fourier space, $\tilde{q}_n(k) = \tilde{q}_0^{2^n}(k/2^n)$. Even more straightforward is to look at the variance

$$Var(Z) = Var(X)/2$$

which again shows that we loose half of the variance at each generation. So a population cannot maintain variability, which is the fuel for the Darwinian

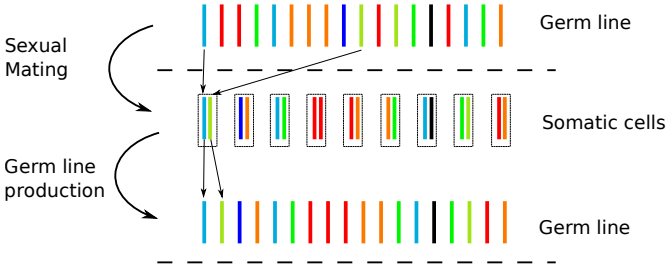


Figure 4.2 – Particulate inheritance. The allele of a given gene are represented by colors. In a somatic cells, two genes are paired and the phenotype of the individual is the result of this pairing.

evolution. This also implies that all mutations are lost after the first mating round.

The above assertion is contrary to all observation and experimental tests conducted. The most elegant experiments was conducted by Mendel and led to the particulate inheritance theory for sexual organisms. In these framework, genes are discrete elements having different alleles. Individuals are carriers of haploid germ cells. During sexual mating, germ cells pair together to form somatic, diploid cells which then proliferate to form a whole individuals. These individuals then produce germ lines by the fission and un-pairing of chromosomes of the somatic cells, and so on. The diploid form of life can be seen as a temporary period between two haploid life cycle. As sexual reproduction provokes only changes in the way alleles are paired, the allele frequency is not changed and no variation is lost from one generation to the other.

The phenotype of an individual however depends on how two different alleles have been paired together. But even the phenotype variability can be shown to be constant in the absence of selection, a fact known as Hardy-Weinberg equilibrium.

For the sake of simplicity, let us suppose that the gene under investigation has only two alleles A and a, and let us call $f(AA)$,$f(Aa)$ and $f(aa)$ the frequency of the different possible genotypes[1]. The allele frequency is the

$$p \quad = \quad f(A) = f(AA) + (1/2)f(Aa)$$

[1]For the moment, we assume that to each genotype a unique phenotype is associated, so genotype and phenotype frequency are similar. We will have more to say on this subject in the quantitative genetic chapter.

$$q = f(a) = f(aa) + (1/2)f(Aa)$$

and of course, $p + q = 1$. Under random mating assumption, the genotype frequency in the next generation is

$$f'(AA) = p^2 \tag{4.3}$$
$$f'(Aa) = 2pq \tag{4.4}$$
$$f'(aa) = q^2 \tag{4.5}$$

and allele frequency in the next generation is obviously

$$\begin{aligned} p' &= f'(AA) + (1/2)f'(Aa) \\ &= p^2 + pq = p \end{aligned}$$

We see that obviously, the allele frequency is conserved, which is trivially due to the model. Less obviously however, we see that after the first generation, phenotype frequencies are also conserved and reach the equilibrium value (4.3-4.5).

The crisis provoked by Jenkins and the contradiction between evolution and blending theory was thus solved by introducing the particulate theory of inheritance and the distinction between phenotype and genotype. The particulate theory wouldn't have been discovered as easily if we were haploid, non-sexual organism[2].

[2]Asimov wrote an SF novel on a world with 3 stars : gravitation is so complicated there that biology and genetic engineering were much more advanced than physics.

5 Are mutations random ?

Around 1940, the main points of evolutionary theory were clarified : scientists knew about genes, chromosomes and Mendelian (quantified) inheritance. They even had a visual representation of the gene. In drosophila for example, in some cells, the chromosomes undergo many round of duplication and all these chromosomes stay sticked to each other and perfectly aligned, making them visible under the microscope. Treating them with some agent, one could stain them differentially and see bands, and therefore make a physical map of where different genes are located. Genes were known to have different alleles[1], the origin of which was attributed to mutations. The chemical structure of the gene (the DNA) was still to come, but the crucial, unknown point was how the mutations occurred. Was the organism inducing specific mutation on its DNA in order to better adapt to various stresses or where these mutations random ? This is the question to which Luria and Delbruck[2] answered brilliantly and set the modern evolutionary theory on firm basis. The crucial point was to design an experiment which could without doubt address the question. The experiment consisted of looking at the number of bacteria surviving a virus infection.

Let us suppose first that organisms can induce mutation in their genes, with some efficiency. Take M batches, each containing N individuals ($N \gg 1$). Expose all the batches to a deadly agent. If each individual has a (very low) probability, say λ to induce the "resistance" mutation in its genes[3], then the probability of observing n survivors in a given batch is a Poisson distribution[4]

$$p(n) = e^{-\lambda N}(\lambda N)^n/n! \tag{5.1}$$

[1] Again thanks to Drosophila and works of Morgan (1866-1945, Nobel 1933) on this organism. Morgan was the first to *induce* mutations in drosophila by the use of chemical agents and showed that they are heritable.

[2] Luria, S. E.; Delbrück, M. (1943). "Mutations of Bacteria from Virus Sensitivity to Virus Resistance". Genetics 28 (6): 491–511.

[3] The mutation is in the gene, because all the progeny of the resistant ancestor are also resistant.

[4] The Poisson distribution is a limiting form of the binomial distribution. The probability to observe n survivors among N individual is

$$p(n) = C(N, n)\lambda^n(1-\lambda)^{N-n}$$

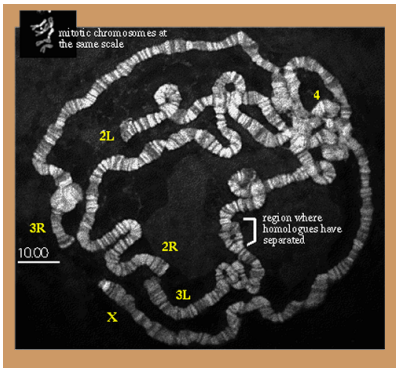


Figure 5.1 – Salivary Gland stained polytene chromosomes and their comparison to mitotic ones. (Image from http://www.msg.ucsf.edu/sedat/polytene_chrom.html.)

In principle, it would be easy to check the validity of this expression. Constitute M test tubes each containing N bacteria, expose each test tube to a deadly virus, plate the medium of each test tube after some times on petri dishes[5] and count the number of colonies, and hence the number of survivors. From the number of survivor in test tubes, we can recover the probability distribution : if m test tubes have n survivors, then $p(n) = m/M$.

In practice, we will need a huge number of test tubes to recover the distribution[6] which was out of the reach of researchers. However, if we want only to check that the distribution is Poissonian, we can use a small number of test tubes (say, around 100 for each set of experimental parameters) and compute only the average survivors per test tube $\langle n \rangle$ and its variance V. The distinctive property of Poisson distribution is that

$$V = \langle n \rangle \tag{5.2}$$

Therefore, if for each set of parameters, we find the variance to mean ratio (VMR) $V/\langle n \rangle \approx 1$, we can be fairly sure that the distribution is a Poisson one.

where $C(N, n)$ is the binomial coefficient. It is not hard to show that the above distribution converges to (5.1) when $\lambda \ll 1$, $N \gg 1$ and $N\lambda \sim 1$.

[5] This is called counting Colony Forming Unit : the technique was invented by R. Koch (and his wife, who knew the basics of cooking food) around 1880.

[6] At these times, researchers did not have an army of Ph.D. students to perform such analysis.

This is what Luria was doing around 1940, and to his surprise, he was finding VMRs of order 1000, 3 orders of magnitude superior to a Poisson prediction. Clearly, something was wrong with directed mutations hypothesis, but would it be compatible with a random mutation one ? The experiment was performed by inoculating each test tube containing nutriment with a small number of bacteria and let them grow to their saturation value N_s, and then expose them to the deadly virus. It is crucial to note that during growth and before reaching saturation limit, bacteria were not exposed to the virus, so any mutation which happened during this stage is purely random. The alternative to directed hypothesis is to suppose that a small proportion of these random mutants confer resistance to the deadly virus, *when* and *if* the exposition happens.

The initial number of bacteria in each test tube was estimated to be around $N_0 = 1000$; This is a large number and we can therefore treat their growth as deterministic :

$$N(t) = N_0 e^{\alpha t}$$

until some saturation is reached[7]. α is the growth rate of bacteria, around 30 min^{-1}. Suppose now that mutants happen by purely random mutations and they pre-exist before exposure to the virus. We can suppose that at each bacterial duplication, one of them transform to a resistant mutant with a (very) small probability a. Let us more over suppose that the mutants have the same growth rate than the wildtype. Now, the number m of mutants is very small and we have to treat it stochastically. The rate transition for the mutants to increase their number by one unit is

$$W^+(m) = \alpha m + aN$$

the first term is the growth rate of the mutants, the second one transformation of WT into mutant. There is no death, so $W^-(m) = 0$. From these two very simple rate, we get the evolution of the mean and variance of mutant populations :

$$\frac{d\langle m\rangle}{dt} = \left\langle W^+(m)\right\rangle = \alpha \langle m\rangle + aN$$

which solution is

$$\langle m\rangle = atN_0 e^{\alpha t} = atN(t)$$

It is interesting to note that the average number of mutants grows faster than

[7]The exact curve would a logistic one, but we don't need a precise treatment of $N(t)$ at this stage. To be more precise, the growth is given by the equation $dN/dt = \alpha N(1 - N/N_s)$ whose solution is $\kappa N_s \exp(\alpha t)/\left(\kappa \exp(\alpha t) + 1\right)$ with $\kappa = N_0/N_s - N_0$.

the WTs. However, it seems impossible to design an experiment to measure this increased growth speed, as the number of mutants is always very small.

The equation for the second moment is given by

$$\begin{aligned}\frac{d\left\langle m^2\right\rangle}{dt} &= 2\left\langle mW^+(m)\right\rangle + \left\langle W^+(m)\right\rangle \\ &= 2\alpha\left\langle m^2\right\rangle + (2aN+\alpha)\left\langle m\right\rangle + aN\end{aligned}$$

whose solution is

$$\left\langle m^2\right\rangle = a^2t^2N_0^2e^{2\alpha t} + 2\frac{a}{\alpha}N_0e^{\alpha t}\left(e^{\alpha t} - 1 - \alpha t/2\right)$$

this seems a little complicated. But note that the first term on the rhs is just $\left\langle m\right\rangle^2$. In the second term, for long time $\alpha t \gg 1$, we can neglect the linear term, so the variance $V = \left\langle m^2\right\rangle - \left\langle m\right\rangle^2$ is simply

$$V = 2\frac{a}{\alpha}\frac{N^2}{N_0}$$

and the VMR reads

$$\frac{V}{\left\langle m\right\rangle} = 2\frac{N}{N_0\alpha t} = \frac{2(N/N_0)}{\log(N/N_0)}$$

We see here that the VMR is amplified by a factor of order N/N_0, which in simple growth can be $\in [10^3 - 10^9]$ and in principle is orders of magnitude higher than a simple Poisson process. The origin of this variability can be understood as the following : a mutation can occur at any time, but the one who occur at the beginning of the growth will yield a huge number of resistants, hence the variability. The argument however should be taken with care, as the probability density of mutation is lower at the beginning. The net effect however is as we computed above. I found few years ago a very general solution of the Luria-Delbrück experiment where many of the simplifying assumptions can be relaxed[8].

When Luria & Delbruck plugged the experimental numbers, they found that the measured variability is even higher that what is expected from the above expression. Part of the additional variability comes from experimental errors : for example the initial number of bacteria is ill measured and can

[8] Houchmandzadeh, B. (2015). *General formulation of Luria-Delbrück distribution of the number of mutants.* Phys. Rev. E, **92**(1), 12719.

vary by a factor of 10. There are also simplifying hypothesis, such as not using the full logistic growth. Note however that we treated the number of bacteria as a deterministic variable. If the experiment had began with very few bacteria (say ~1), we should have also treated the bacterial growth as stochastic and would expect even higher variability.

The story is not yet finished. In 1988, Cairns et al.[9] published a paper going against the Luria and Delbruck experiment. They showed that if mutations are not resistance to lethality, then maybe some mutations are directed. In their experiment, they used $Lac-$ bacteria plated on agar which contained Lactose as the sole source of energy and then counted the distribution of $Lac+$ mutants. The distribution is somewhere between the Poisson and the Delbruck distribution. I don't intend to go into the detail of the Cairns experiment, as there were many counter experiments and a big amount of controversy, and many many models for the interpretation of their result.

[9]Cairns, J.; Overbaugh, J.; Miller, S. (1988). "The Origin of Mutants". Nature 335 (6186): 142–145

6 Fundamentals of population genetics.

[0-dimentional model (well mixed population). Fisher-Wright model. Mean field treatment and Hardy-Weinberg equilibrium. Moran model. Neutral case. full treatment. Kimura equation. Hint at spatially extended.]

Around 1920's, ideas of genes as "quantum" of inheritance information and Darwinian evolution began to merge to form the basis of the new synthesis. Around this time, the idea of a gene having multiple alleles was clear. Each allele of the gene modifying in certain way a trait, it was a short step to go one step further : the fitness of the individual depending on its traits (called its phenotype), an allele of a gene can cause the individual carrying it a higher fitness than the base population. Mendelian genetics had cured the inconsistencies of the blending theory of inheritance, so the question was now to write a coherent mathematical model describing the propagation (or disappearance) of an allele in the population taken as a whole. This field is now called *population genetics*.

6.1 The Fisher-Wright model, deterministic approach.

The first model of population genetic was introduced by Fisher and Wright in the 1930's and bear their name. We'll suppose from now on that the populations reproduce asexually, sex being a small mathematical complication[1]. Let us suppose that the population consist of two types, identical but for one gene, which has two different alleles A and B. We'll refer to individual carrying the allele A as $A-$individual. We also suppose that A individuals have fitness r, compared to fitness 1 for the B individual.

But wait ... I used the term *fitness*, what does it mean ? The fitness implies some idea of "well adapted" to the environment, but of course, this

[1] in these lectures

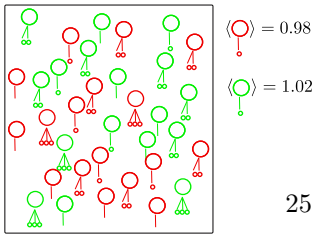


Figure 6.1

cannot be measured. When we think about *measurement* of fitness, there is only one quantity which can be measured : the number of surviving progeny, capable of reproducing.

This is exactly what we mean by fitness and the *only* thing we mean. In order to measure it, select all parents of type A (B), measure the mean number of adult progeny of each individual of each type n_A and n_B. The fitness of A individual (relative to B ones) is $r = n_A/n_B$ (Figure 6.1).

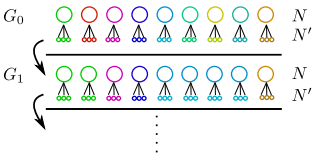


Figure 6.2

The FW model can be formulated as follow : in a population of fixed size N, each individual produces progeny proportional to its fitness. Among the N' progeny, N individuals are selected at random to form the next generation (Fig. 6.2). This scheme keep the number of individuals always at precisely N. Let us call p_0 the proportion (the term frequency is used in most books) of $A-$individual in generation 0. According to the above scheme, the number of their progeny is $\alpha r N p_0$ where the number of the progeny of type B is $\alpha N(1-p_0)$, therefore the frequency of A individuals in the next generation is

$$p_1 = \frac{rp_0}{rp_0 + (1-p_0)}$$

and the change in frequency is

$$\Delta p = p_1 - p_0 = \frac{(r-1)p_0(1-p_0)}{1+(r-1)p_0}$$

which, for small selective advantage $s = r - 1 \ll 1$, can be approximated by

$$\Delta p = sp(1-p) \tag{6.1}$$

This is the deterministic Fisher equation, and shows that at small frequency, a mutant spread exponentially and begin to saturate when it becomes dominant. Measuring time in generation unit, the above equation can be written as

$$\frac{dp}{dt} = sp(1-p) \tag{6.2}$$

with the solution

$$p(t) = \frac{Ce^{st}}{1 + Ce^{st}}$$

So at small frequency, a beneficial mutant will increase exponentially the size of its population. Very often, specially in bacterial studies, the fitness and the rate of exponential growth are used as synonymous.

6.2 The Fisher-Wright model : stochastic approach.

The FW model contains a key ingredient of evolutionary thinking : the importance of mere luck (which we may call stochasticity in order to look serious). One great aberration of early evolutionary thinking was its deterministic approach : if you are fitter (than your neighbor), you are going, for sure, to dominate it through generations[2]. This opinion had its reverse way of thinking which was even more dangerous : if you are dominating today (compared to your neighbor), it is because you are/were superior to him[3]. Another formulation of the same opinion is that if a trait is observed in an organism, it is because it is beneficial and has been selected for.

The value of excess relative fitness s had no importance in this framework. But what if $s = 10^{-3}$ or 10^{-9} ? Clearly the domination fate should take this value into account ? Suppose, to take the image of the two runners on the cover of this book, that one runner runs at $v = 5$m/s and the other at 5.01m/s. Which one is going to be eaten by the tiger ? Well, if you do the experiment by taking 10^6 pairs of runners you would see that the faster one survive in *average* slightly more than the other (by less than a 0.2% margin), but is this relevant ? Again, suppose that among Wild type humans running at 5m/s, a mutant with running speed of 5.01m/s appears. It is very possible that before he is able to spread his genes, because of mere bad luck, he is going to be eaten way and its beneficial mutation will disappear with him. On the other hand, it is possible that both you and your neighbor had a speed of 5m/s, but because of bad luck, he got eaten and you did the gene spreading business.

[2] In terms of number of gene copies, which again, is the only measure of evolution

[3] For old Greeks, success was a sign of blessing by gods and a winning general could be considered somehow a god himself. For example, Octavius, after defeating Marc-Anthony, accepted to be deified in Greece, even though he had only contempt for this tradition back at home. It is amusing to observe that this tradition found its way into evolutionary thinking much later.

Evolution should therefore be seen as a play between two forces : the deterministic one due to the value of fitness and the stochastic one due to all the random processes which can happen. For physicist, trained in statistical physics in his young age, this is an obvious framework : particles sediments only if gravitation dominates over thermal noise. For evolutionary biologist, this approach took a little longer and unfortunately the adaptationist program, as Gould called it, is still widespread.

The beauty of the FW model was to introduce the stochasticity into evolution in a very elegant and generic way. The criteria for the importance of luck, as we shall see below, is very simple:

$$Ns < 1$$

where s is the excess relative fitness and N the population size.

6.2.1 Genetic drift.

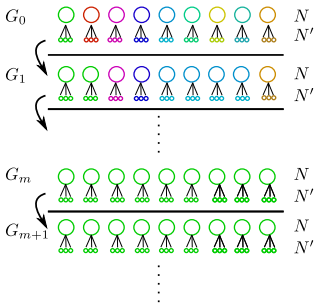


Figure 6.3

The best way to illustrate the importance of stochasticity (or noise, as we call it in physics), is to consider the *neutral* case where all individuals have the same fitness. In a deterministic approach (eq. 6.1), if the proportion of individual A is p at generation zero, then its proportion will be p at generation 100. But if we play the Fisher-Wright game, selecting N individuals at each generation among the N' progeny, we will observe that after some generation, *all* individuals are the descendant of a single individual in generation 0 ! This phenomenon is called the genetic drift (Figure 6.3) and when all individuals are descendants of single one, we say that the ancestor has been *fixed*.

The mathematical problem of the fixation of a neutral allele is best treated using the coefficient of co-ancestry. We suppose that each individual produces exactly g progeny, and among all the progeny, exactly N individuals are selected to make up the next generation.Let us pick at random to individual at generation n and call f_n the probability that they have a common ancestor. Obviously, at generation 0 where we consider all individuals different, $f_0 = 0$. Let us now pick up two individuals at generation n at work it

backward[4]. The event E that these two to have a common ancestor is either E_1 that they have the same parent or E_2 that they have different parents ($E_{2,1}$) which are co-ancestor ($E_{2,2}$). As the logical AND (for independent events) and OR (for exclusive events) are translated into multiplication and addition in probabilities,

$$\begin{aligned} P(E) &= f_n \\ P(E_1) &= 1/N \\ P(E_2) &= (1-1/N) \times f_{n-1} \end{aligned}$$

Which gives us the simple recurrence relation

$$f_n = \frac{1}{N} + (1 - \frac{1}{N}) f_{n-1}$$

We can very easily solve the above recurrence equation. But before doing so, note that we can rewrite this equation as

$$f_n - f_{n-1} = -(1/N) f_n + 1/N \tag{6.3}$$

which is equivalent to the differential equation $df/dt = -(1/N)f + 1/N$ which converges exponentially, over a time scale of N, toward 1. We have two important information here : even when all parents have the exact same fitness, the population becomes homogeneous with a co-ancestry coefficient of 1 ; this homogenization occurs overs time scale of N and is very fast in small communities.

If we wanted to solve exactly the recurrence equation (6.3), it is enough to note that we can rewrite it as $f_n - 1 = (1-1/N)(f_{n-1} - 1)$ and as $f_1 = 1/N$, we would get

$$f_n = 1 - (1 - 1/N)^n$$

6.2.2 FW Transition rates with selection.

Let us now relax the neutrality hypothesis and suppose that a mutant has a higher fitness than wild type individuals : WT and mutants produce g and gr progeny respectively. Let us suppose that we have at some point we have n mutants (and $N - n$ WT). What is probability $P(m|n)$ of having m

[4] A more elaborate way of this kind of computation is called coalescent theory

mutants in the next generation ? Well, this is the plain binomial distribution

$$P(m|n) = C_N^m q^m (1-q)^{N-m}$$

where q is the probability of picking a mutant among the progeny and C_N^m is the binomial coefficient. For the probability q we have

$$q = \frac{nrg}{nrg + (N-n)g} = \frac{nr}{N + n(r-1)}$$

We could therefore, from any given initial state $P_0(m)$, compute by recurrence the probability $P_k(m)$ of having m mutants at generation 0 :

$$P_k(m) = \sum_n P(m|n) P_{k-1}(n) \tag{6.4}$$

This is very easily done by computer simulations. The analytical solution however is much harder to obtain[5]. Equation (6.4) is an example of a Master equation, and indeed, there are hard to solve, specially when they are multi steps.[6] The problem can be tackled in the diffusion approximation of Kimura, which is given in §6.4.

6.3 The Moran Model.

In 1962, Moran introduced a simple model, equivalent to FW one but much easier to tackle mathematically. In this model, birth and death occur continuously. When one individual dies, it is immediately followed by another one duplicating.

6.4 Kimura's diffusion equation.

Master equation of these types however can be approximated by a continuous Fokker-Planck equation for large population size. Calling $x = n/N$ and $Nu(x) = P(n)$, we have

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial \left[a(x)p(x,t)\right]}{\partial x} + \frac{\partial \left[b(x)p(x,t)\right]}{\partial x^2}$$

[5] In fact, nobody has done it

[6] A one step stochastic process is one where $P(m|n) = 0$ if $|m-n| > 1$.

7 The dynamics of mutation spread.

7.1 The Fisher Wave.

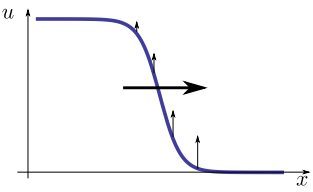
Mutants appear in a small region of space and spread to adjacent position. The simplest equation used to capture this dynamics is to complement the deterministic Fisher equation (6.1) with a diffusion term in order to account for the spread :

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + su(1-u) \tag{7.1}$$

where s is the excess relative fitness and $u \in [0,1]$ is the mutant relative density. This equation was proposed simultaneously by Fisher[1] and Kolmogorov[2] in 1937, and it is known as the FKPP equation.

The FKPP in now widely used in many different field outside population genetics. It has the grace of having traveling wave solution whose speed and width can be exactly computed[3] which we are going to investigate now.

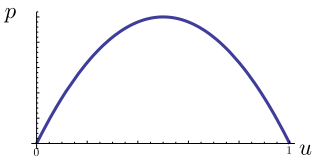
We look for solutions in a 1 dimensional infinite space where $u(-\infty) = 1$ and $u(+\infty) = 0$, *i.e.* with the mutants fixed at the left side space and (still) absent at the right side space. The part of the space connecting this two regions is called the front. A good metric to measure the width of this front is $W = \int_{\mathbb{R}} u(1-u)dx$: only region where $u \neq 0, 1$ contributes to the width and the contribution is proportional to the deviation from these two values.



[1] Fisher, R. A. (1937). *The wave of advance of advantageous genes.* Annals of Eugenics, 7, 355–369.

[2] Kolmogorov, A. N. (1991). *A study of diffusion equation with increase in the amount of substance.* In Selected work of N.A. Kolmogorov (pp. 242–271).

[3] For a review, see Vansaarloos (2003), *Front propagation into unstable states*, Physics Reports , 29–222.



We are looking for traveling wave solution $u(x,t) = u(x-ct)$ which transforms eq.(7.1) into an ordinary (non-linear) second order solution $Du'' + cu' + su(1-u) = 0$. We note that the equation does not depends explicitly on the space variable x, so let us set $p = -du/dx$ and rewrite everything in terms of p and u :

$$\frac{d^2u}{dx^2} = -\frac{dp}{dx} = p\frac{dp}{du}$$

which finally get us to the equation

$$Dpp' - cp + su(1-u) = 0 \tag{7.2}$$

where $p = p(u)$ and $p' = dp/du$. If we can solve this equation, we can get back to $u(x)$ by integrating $dx = -du/p(u)$. But even if we cannot get the exact solution of (7.2), we can extract useful information on the speed and width of the front. Let us suppose that when $p'(0) = \lambda \neq 0, \infty$, *i.e.* at the origin, the slope of $p(u)$ is finite. Then dividing eq.(7.2) by u and letting $u \to 0$, we have

$$D\lambda^2 - c\lambda + s = 0$$

which is a second order algebraic equation and has a solution for

$$c \geq c_{min} = 2\sqrt{Ds}. \tag{7.3}$$

The above condition means indeed that FKPP equation does not have a unique solution : for each $c \geq c_{min}$, there is a front propagating with this speed. This does not seem to be a realistic answer to a well defined problem. If however we go one step further and perform a stability analysis, we will see that all these solution, except the one corresponding to c_{min} are unstable and the front, after a transitory period, will indeed propagate at speed c_{min} and take the shape corresponding to this speed. We omit the stability analysis here, which can be found in the cited review.

The Width of the propagating front can be extracted as easily as its speed :

$$\begin{aligned} W &= \int_{\mathbb{R}} u(1-u)dx \\ &= -\frac{1}{s}\int_{\mathbb{R}} \left(Du'' + cu'\right) dx \end{aligned}$$

$$= \frac{c}{s}$$

and therefore, the stable front has the width

$$W_{min} = 2\sqrt{D/s} \tag{7.4}$$

The two main results of the Fisher Wave equations are then that the speeds scales as $\sqrt{s}$ and the width as $\sqrt{1/s}$. The relevant range of evolutionary biology is however when $s \ll 1$, which is exactly, as we saw earlier, when the deterministic approach breaks down. In particular, the divergence of the width for small s does not seem very realistic.

The small s limit is where the demographic noise becomes important. In order to take this noise into account, Doering et al.[4] added a phenomenological noise term to the FKPP equation :

$$\frac{\partial u}{\partial t} = D\nabla^2 u + su(1-u) + \sqrt{bu(1-u)}\eta(x,t)$$

where $\eta(x,t)$ is a white noise. The addition of the noise term corrects for many of the insufficiency of the FKPP equation at small s. We can however do a more rigorous approach by directly solving an individual based model such as the spatial Moran model. This is the intent of the next section.

7.2 The Spatial Moran process.

[4]Doering, Charles R. and Mueller, Carl and Smereka, Peter, *Interacting particles, the stochastic Fisher-Kolmogorov-Petrovsky-Piscounov equation, and duality*, Physica A: Statistical Mechanics and its Applications (2003), 243–259.

8 Neutral speciation.

Observation of the stunning biodiversity in various ecosystems is one of the factors that led Darwin and Wallace to formulate the theory of evolution. The finches of Galapagos (Figure 8.1) are the standard example cited in any textbook of the field. Even at a single trophic level, *i.e.* considering species which use the same resources, the biodiversity is always large. In spite of many competing theories the question of the causes of biodiversity is still unanswered today.

The most dominant view is the adaptationist one: each species is adapted to its local environment and biodiversity is just a reflection of the heterogeneity of available resources. In this view, neutral biodiversity (species using the same resources and being similar in fitness) can exist only because of geographical barriers between close ecotypes. The possibility of having speciation at the same trophic level at the same geographical location (*sympatry*) has been ruled out by Ernst Mayr in his famous book[1], with far-reaching consequences on evolutionary thinking.

The reason behind this "outlawing" of neutral biodiversity is called the "exclusion principle"[2]. In a series of beautiful experiments conducted at the

[1]Mayr, E. (1942). *Systematics and the Origin of Species: From the Viewpoint of a Zoologist*. Harvard University Press.

[2]Hardin, G. (1960). The Competitive Exclusion Principle. science, **131**:1292–1297.

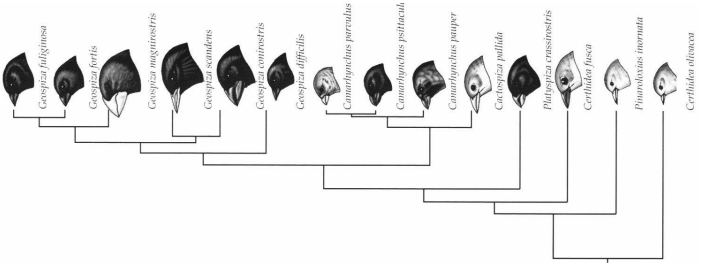


Figure 8.1 – The finches of Galapagos. [Lomolino et al., Biogeography, 2006]

Moscow's zoological institute during the 1930's[3], Gause observed that when two close species of Paramecia (an amoeba) are cultured together in a test tube, one will always drive the other to extinction. As the years passed, this observation became a dogma : "no two species can survive on the same resources at the same place". On the other hand, it was believed that in the absence of geographical barrier, each species will *diffuse* in order to occupy the whole available space[4]. The combination of these two believes led Mayr to rule out sympatry

The generalization of the Gause rule however is very fragile. First of all, the exclusion principle has never been observed in nature (footnote 2). People loved so much the exclusion principle that failing to observe it in nature, they invented resource differentiation everywhere (this birds loves lower leaves, the other higher ones) or asserted that observed species resources only partially coincide. But even if the exclusion principle were true, would it forbid sympatry ? The first question mark is the time to extinction : how long does it take for a neutral species to go extinct ? If this time in the order of geological times, the exclusion principle has no value. Second, does "everything everywhere" belief true ? We will see that this is not the case, and species, in the neutral case, tends to segregate and form boundaries, in the absence . When these fact are weighted, we will see that there is no case against sympatry. The prejudice against sympatry was caused by people thinking about evolution only by words, and not weighting the mathematical consequences of the causes they were proposing.

The need for an alternate/complimentary model arose as Ecologists began to gather large data on biodiversity and observed general patterns everywhere. The adaptionist program is specific to each species and each habitat. Observing general patterns could seem at first at odd with this view and this is the reason many researcher tried to look at generic models of resource distribution and their uptake, before the advent of neutral models shook up this general view. Let us first review some of the generic data gathered by ecologists.

One of the most striking observed "law" is the species-area relationship which states that the number of species S in an area exhibits a power law dependence on the size A of the area considered: $S = kA^z$ with z in the [0.2,0.3] range for most habitats (Figure 8.2). It has to be noted however

[3]G.F. Gause (1935), *Vérifications expérimentale de la théorie Mathématique de la lutte pour la vie*, Herman et Cie Editeur, Paris .

[4]This is called the rule of "everything is everywhere". For a review of this "theory", see O'Malley, M. (2007). *The nineteenth century roots of "everything is everywhere"*. Nature reviews. Microbiology, **5**:647–51.

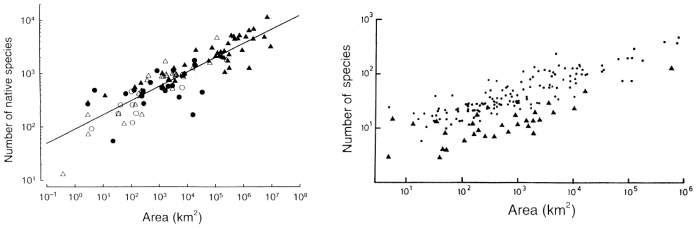


Figure 8.2 – Species-Areas relationship for Vascular plant (left panen) and Land birds (right pannel). [Willamson et al., J. Biogeography, 28, 827-830 (2001)].

that this "law" is not engraved in stone. As it can be observed, the data are very noisy (Figure 8.2 displays only the best selected cases) and even though, they can often be fitted as well by other curves.

An alternative and more precise measure of biodiversity for a fixed area is the abundance curve: collecting species in a given area and measuring the abundance of each species leads to the abundance curve $\phi(n)$, which is the histogram of the number of species having abundance n. Abundance curves taken from very different habitats began to show very similar patterns [5] and it became more and more obvious that they can be unified through a single parameter (Figure 8.3).

The third observation came from measurements of biodiversity in islands close to a continent. It was observed that the number of species in islands decreased as a function of its distance from the continent and increased with the size of the island (Figure 8.4). To explain the third observation, MacArthur and Wilson[6] took a bold approach. They supposed that (i) all species at the same trophic level are *equivalent* ; (ii) species migrate from continent to islands, with the rate of migration a decreasing function of the distance ; (iii) due to random sampling from one generation to the other, species become extinct in islands, with the extinction rate a decreasing function of the size of the island. The number of species present on the island is then a dynamic equilibrium between migration and extinction (Figure 8.4).

MacArthur and Wilson's article, considered as a cornerstone of biogeog-

[5] for a review, see Hubbell, S. P. (2011). *The Unified Neutral Theory of Biodiversity and Biogeography*, Princeton University Press.

[6] MacArthur, R. H., & Wilson, E. O. (1963). An equilibrium theory of island zoogeography. Evolution, **17**: 373–387.

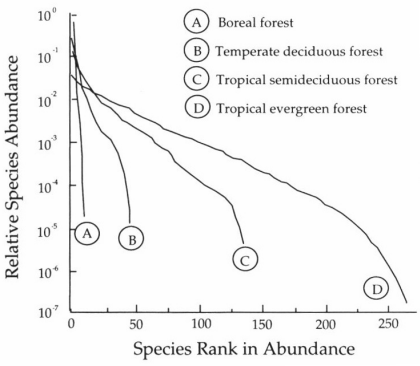


Figure 8.3 – The abundance curve collected in various type of forests. [From Hubbell's book].

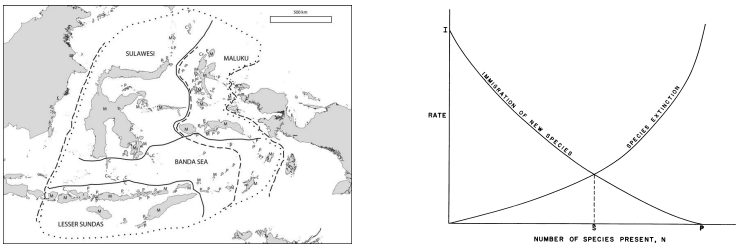


Figure 8.4 – Left panel : A typical example of islands used to gather biodiversity Data [Cartens, JBI 2012]. Right Panel: The MacArthur and Wilson phenomenological explanation .

raphy, was a radical departure from Mayr and the adaptationist program, and proved extremely successful. The next radical step then was taken by Hubbell who applied the same idea to the whole continent : all species at a given trophic level are equivalent, new species appear by mutation and become extinct by genetic drift. The biodiversity curve is then a function of a single number that takes into account the mutation rate and the size of the community. Hubbell's book founded what is called the neutral theory of biodiversity and provoked an incredibly wide and heated debate in the ecological community, which is still ongoing.

8.1 The neutral model.

In retrospect, it seems strange that the idea of neutrality, considered very early by population geneticists such as Malecot (1948) and Kimura (1985), took so much time to permeate the ecological/evolutionary thinking; I believe that this is partly due to the influence of Mayr's book[7]and the prevalence of the competitive exclusion principle. The main idea however is very simple : neutral macroecology is similar to neutral population genetics. The latter deals with alleles of a gene, their frequency and its change because of genetic drift and (neutral) mutations, where the former deals with the equivalent concepts of species, their abundance and its change because of ecological drift and neutral speciation (presumably because of the accumulation of many mutations at the individual levels) and so on. New species emerge with a rate ν . It takes some times τ_a for a new species to become abundant by pure genetic drift. If the arrival time of new species τ_e is much shorter than τ_a, many equivalent species will coexist at the same geographical location and their abundance will be a dynamic interplay between emergence of new species and extinction of existing one. To investigate these assessment, we again use the Moran model to which we add the mutations.

Consider a community consisting of N individuals and S species, with species i having n_i individuals (Figure 8.5a). All individuals, regardless of their species, are equivalent in their reproductive/death rate. When an individual dies, it is immediately replaced by the progeny of another one. Because of mutations, the progeny can differ from its parent with probability ν, thus forming a new species appearing with abundance 1. After its appearance, the species abundance is a stochastic function of time ; if an individual

[7]Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). *Sympatric, parapatric or allopatric: the most important way to classify speciation?* Phil. trans. Royal Soc. London. Series B, **363**:2997–3007.

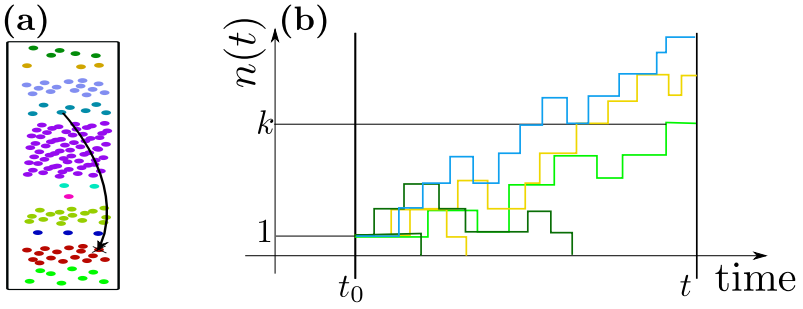


Figure 8.5 – (a) The Moran model of a neutral community composed of various species (distinguished here by their colors), where an individual is replaced upon its death by the progeny of another regardless of its species. (b) Each new species appears with abundance 1 by mutation at some time t_0; Stochastic dynamics of the number of individuals $n(t)$ of few species appeared at time t_0. .

is the sole representative of a species and dies, then this species disappears. As in the previous section, the probability $P(n,t|1,t_0)$ for species i to have n individuals at time t, knowing the species appeared at time t_0, obeys a Master equation where the transition rates are:

$$W^+(n) = \mu(N-n)n(1-\nu)/N \quad (8.1)$$
$$W^-(n) = \mu n\left(N-n+\nu(n-1)\right)/N \quad (8.2)$$

The increase rate $W^+(n)$ is the probability density of death of an individual that does not belong to the considered species $\mu(N-n)$ multiplied by the probability of birth of an individual that belongs to the considered species n/N, times the probability of no mutation $(1-\nu)$. The decrease rate is similar, but takes also into account the probability of an individual dying and being replaced by the progeny of a member of its own species with a mutation.

Let us set the origin of time at $t_0 = 0$. The master equation gives the fate of one particular species. $\langle\phi(n)\rangle$, the average number of species having population size n at time t is the sum of all those who have been generated at an earlier time τ and have reached abundance n at time t :

$$\langle\phi(n)\rangle = \int_0^t f(\tau)P(n,t|1,\tau)d\tau$$

$$
\begin{aligned}
&= \nu \int_0^t P(n, t-\tau|1,0)d\tau \\
&= \nu \int_0^t P(n, \tau|1,0)d\tau
\end{aligned}
$$

where $f(\tau)$ is the probability per unit of time of generating a mutant and is equal to ν (time is measured in units of generations $1/\mu$). Defining the mutation pressure as $\theta = N\nu$, the quantity ϕ can be obtained at the limit of large times and shows that an equilibrium is reached.

[include detailed calculation here].

For large communities, using proportions $\omega = n/N$ and abundances $g(\omega) = N\langle\phi(n)\rangle$, the result takes a simple form

$$g(\omega) = \theta\omega^{-1}(1-\omega)^{\theta-1} \tag{8.3}$$

The above computations ignore spatial distances: an individual can be replaced only by the progeny of its neighbor rather than by everyone in the community. A self consistent model of geographical dispersal is incredibly difficult. We can however go one step further and apply the above model to the case of island biogeography, where a small island of size M is close to a continent of size N ($M \ll N$). The population of the island is affected by migration from the continent, but given the large size of the continent, the reverse is not true. We can also neglect mutation inside the island as the mutation pressure is small. So the transition rates in the island are similar to eqs (8.1,8.2) except that a local individual can be replaced by a migrant from the continent with probability m, where the abundances are given by expression (8.3). Defining the migration pressure as $\xi = Mm$, in the limit of large sizes of both the island and the continent, we can compute the relative abundance $g_I(\omega)$ inside the island as[8]

$$g_I(\omega) = \xi\theta \int_0^1 (1-\omega)^{\xi u-1}\omega^{\xi(1-u)-1}u^\theta du \tag{8.4}$$

This expression may seem cumbersome, but it can be easily plotted and depends on only two parameters : θ which itself can be seen as a function of biodiversity on the continent and ξ which is a simple decreasing function of the distance between the continent and the island. Expression (8.4) is the mathematical expression of the original MacArthur and Wilson model and

[8] Vallade, M., & Houchmandzadeh, B. (2003). *Analytical solution of a neutral model of biodiversity.* Phys Rev E **68**:61902.

can be put to experimental verification.

Improving the above model by taking fully into account the spatial dimension seems mathematically intractable. We have been able to slightly improve the continent-island model by treating both communities on an equal footing but going further seems beyond the reach of the mathematical tools we used. Nevertheless, the neutral theory of biodiversity is a falsifiable theory of biodiversity. It has been put to intense test and has been proved successful at interpreting quantitatively available data in island biogeography[9]. As in the previous section, the merit of this model is to provide a first approximation for biodiversity which will always be present, even though many data will necessitate the addition of more ingredients, such as for example, density dependence of replacement rates[**?**, **?**] to explain deviation from this theory.

[9]Rosindell, J., Hubbell, S. P., & Etienne, R. S. (2011). *The unified neutral theory of biodiversity and biogeography at age ten.* Trends in ecology & evolution, **26**:340–8.

9 Predator induced speciation.

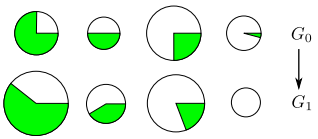
dynamic landscape (l'enfer c'est les autres), phage induced diversification of bacteria (Williams paper, BMC Evobio 2013:13:17).

Neutral speciation gives a minimum background for the biodiversity. There is however a very important fact of the life we have not spoken about yet : we (living organisms) all belong to a food chain; we do have resources (organisms below in the food web) and predators (organisms above in the same graph). As our predators are specialized to us, a good way of escaping them would be to commit speciation.

10 The paradox (?) of altruism and cooperative behavior.

10.1 The Price equation.

The price equation has become a famous one in some evolutionary circle, although why a trivial mathematical concepts can be so much sanctified elude the author of these lines[1].



Consider, at generation G_0, N individuals spread among M group, The number of individual composing group i noted as n_i. Let us associate a phenotypic variable z_i to each group (we don't precise what z is: at this stage it can be anything, from the color of the eye to the proportion of altruists in this group); let us moreover, suppose that the group i has fitness f_i. Now, after reproduction, in generation G_1, the number of individuals in group i become n_i', and the variable z_i changes to $z_i + \delta z_i$. Again, at this stage, we don't need to be to precise about what δz_i is. If groups are ranked by the color of the eye for example and reproduction is perfect, then $\delta z_i = 0$. If z_i is the proportion of altruists, then we can have $\delta z_i < 0$. For the moment, δz_i is a given parameter, which we can link to other variables when we have a precise model in mind.

by the very definition of the fitness f_i we have $n_i' = f_i n_i$. Now let us compute some averages.

$$\left\langle z^0 \right\rangle = \sum_i n_i z_i / N = \sum_i w_i z_i$$

where $\left\langle z^0 \right\rangle$ is the average phenotype in G_0 and $w_i = n_i/N$ is the weight of

[1] Like Galois, one the few mathematicians known to the public at large, Price also has had a tragic end.

group i. By the same token, we get

$$\langle z^1 \rangle = \sum_i n_i'(z_i + \delta z_i)/N' \tag{10.1}$$

where

$$N' = \sum_i n_i' = \sum_i f_i n_i = N \sum_i f_i w_i$$

defining the average fitness in G_0 as

$$\langle f \rangle = \sum_i f_i w_i = N'/N.$$

so we can write eq. (10.1) as

$$\langle z^1 \rangle = \frac{1}{\langle f \rangle} \sum_i w_i f_i (z_i + \delta z_i)$$

The change in the average phenotype is

$$\begin{aligned} \Delta \langle z \rangle &= \langle z^1 \rangle - \langle z^0 \rangle \\ &= \sum_i w_i \left(\frac{f_i}{\langle f \rangle} - 1 \right) z_i + \sum_i w_i \frac{f_i}{\langle f \rangle} \delta z_i \end{aligned}$$

so the change in the average is the sum of two terms : the first term contains only the phenotype, the second only the change in the phenotype. We can repackage this slightly more

$$\langle f \rangle \Delta \langle z \rangle = \sum_i w_i \left(f_i - \langle f \rangle \right) \left(z_i - \langle z \rangle \right) + \sum_i w_i f_i \delta z_i$$

where we have used the fact that $\sum w_i (f_i - \langle f \rangle) \langle z \rangle = 0$. Note that in the above expression, all the $\langle\rangle$ on the right hand side are averages taken in G_0. The above equation is called the price equation and is usually written as

$$\langle f \rangle \Delta \langle z \rangle = \mathrm{Cov}(f, z) + \langle f \delta z \rangle$$

What is interesting in this equation is that even if $\delta z_i < 0$ in all groups, the shift in average $\Delta \langle z \rangle$ can be > 0. This is just a demographic effect, first noted by Simpson in the 50's[2].

[2]Simpson, Edward H. (1951). *The Interpretation of Interaction in Contingency Tables.*

Now suppose that the phenotypic trait z_i is the proportion of altruists in each group. Suppose that groups with a high number of altruist grow faster, *i.e.* $\mathrm{Cov}(f, z) > 0$. We see that if there are some range of parameters where even when altruists have a smaller fitness than non-altruists, *i.e.* their proportion in each group decreases ($\delta z_i < 0$) , the proportion of altruists in the whole population can increase. This demographic twist seems strange to us when we mix absolute numbers and proportions.

Of course, this effect cannot be sustained in the long term. In the extreme case when from example only one group subsists, then altruists are doomed. In order to maintain the altruists, every m generation, some mixing and redistribution into new groups is needed. This mechanism, called haystack, can be observed in some virus. We'll come to that a little later.

Price equation applicability is however very restricted. It is not predictive if the fitness depends not only on the proportion of type A, but also on other parameters (such as the group size). Then the quantities $\mathrm{Cov}(f, p)$ and $\langle f \delta p \rangle$ cannot be computed without knowing the distribution of these other parameters.

Give a numerical example.

introduce the first version of Fisher Fund. Theor.

Let us suppose that $w(z)$ designates the absolute fitness (the mean number of surviving progeny).

Huxley and Kropotkin, Hamilton and Price, kin selection, group selection and the common good.

Journal of the Royal Statistical Society, Series B 13: 238–241.

11 Quantitative genetics : selection of continuous traits.

11.1 General concept of heritability.

Until the end of the XIX, how we inherit a trait was unknown. The most shared belief (including Darwin and Wallace) was the *blending* theory, where we are the average of our parents. This theory is not (mathematically) compatible with evolution and natural selection, as pointed out by Jenkins. Then in 1901, the Mendelian theory of heritability was rediscovered and things were put into order by the founding fathers of population genetics.

The Mendelian theory however was too simple for most of the trait. Mendel was looking at traits which are governed by one or two genes ; the number of possible traits were few, in the range of two or four : is the pea yellow or green ? round or elongated ? Most of the trait are controlled by many genes and therefore, the value of the trait becomes continuous[1]. Body height and weight, cows milk production, rice yields, ... are but a few examples. Suppose that M genes $(g_1, g_2, ...g_M)$ control a trait; each gene i has n_i alleles and we call the genotype of an organism the particular combination it has : $\mathbf{g_k} = (g_1^{k_1}, g_2^{k_2}, ...g_M^{k_M})$. We call z the value of the trait and suppose that it is entirely determined by its genotype:

$$z = F(\mathbf{g}) \tag{11.1}$$

Note that this is entirely theoretical thinking. It is very rare that we know all the genes involved and the function which relates the genotype to phenotype.

A parent transmits exactly, without error, its genome to its progeny. Therefore, if we believe equation (11.1), we will expect to find identical phenotypes in parents and child. As we all know, this is not the case. The phenotype is a function of the parent's genotype and other factors such as

[1] And its distribution in the population very often Gaussian. Suppose that M genes control the trait Z ; each gene X_i has many different alleles and their effect are additive, then $Z = \sum X_i$ and so the distribution of Z is normal if M is big enough, by the virtue of central limit theorem. It the effects are multiplicative, then Z is log-normal.

the environment : even if you have a dutch genotype, not eating enough in your childhood will keep you below 180cm in height. Let us denote $y = F(\mathbf{g})$ and call it, broadly speaking, the genotype. Then

$$z = y + \xi$$

where ξ is a random variable. The randomness of ξ is partly due to the randomness of the environment, partly to the intrinsic noise of cellular processes[2], and any other variability we don't know about. In order to characterize ξ, we should be able to measure the probability density $f(z|y)$, the probability of observing the trait z knowing the genotype y. Such a measure is not totally out of reach : take one individual, make N copies of it, let them grow in the wild, collect the adults and measure the distribution. Do that for all available genotypes. Even if this thought experiment seems too complicated, farmers and agro-engineers use similar but less complete data analysis.

Now, consider a real population. A real population is not iso-genic[3], but different genotypes are represented in the population with different frequencies. Let us note Y and Z the random variables of genotypes and phenotypes due to this distribution. We can write very generally

$$Z = Y + \Xi$$

and if the response to environment is independent of the genetic background (you grow shorter than average if you eat less than average, whatever your genetic background), we'll have

$$Var(Z) = Var(Y) + Var(\Xi)$$

11.2 The selection process.

Now, here's the problem : we (or mother Nature) can only select for what we see, *i.e.* the phenotype. When we select for phenotypes in a given interval $[z, z + h]$, some individuals with genomes we are not interested in will get into the pool: because of the noise ξ, their phenotype will be in the good interval. If the noise is small, this is not a big issue; but if the noise is large,

[2]The noise at the cell level, also called "non-genetic individuality", has seen a huge experimental and theoretical development during the end of 90's and begining of 00's. See Raj & Van Oudenaarden[**?**] for a review.

[3]Well, may be soon we'll have cloned cows, but for the moment we don't

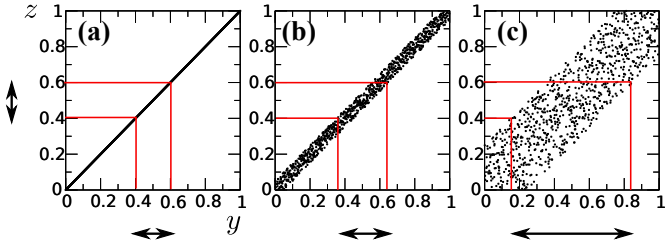


Figure 11.1 – phenotype as a function of genotype at: (a) no noise ; (b) moderate noise ; (c) high noise. Vertical arrow denotes the selection pressure: here we select for phenotype $\in [0.4, 0.6]$; Horizontal arrows show the selected genotype.

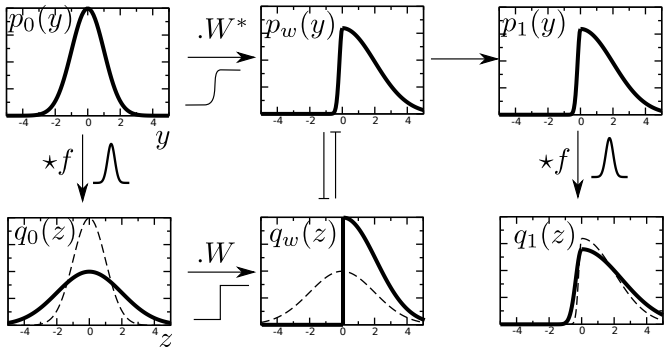


Figure 11.2 – The scheme of the selection process.

everybody will get into the pool and selection is worthless : you are not going to increase the yield of your rice crop. The amount of noise, which hinders our selection process, is related to a quantity called *heritability*. Let us precise a little more what are its consequences (fig.11.2).

Suppose that the genotype of the population follows the distribution[4] $p_0(y)$. The phenotype distribution $q_0(z)$ is given by

$$q_0(z) = \int_{y \in I} p_0(y) f(z|y) dy \tag{11.2}$$

[4]the probability density

We now select a subpopulation according to some criterion[5] $W(z)$. The selection function $W(z)$ is the proportion of individuals in phenotypic class $[z, z+dz[$ which are used to produce the next generation. The phenotype distribution after selection becomes

$$q_w(z) = \frac{1}{\bar{W}} q_0(z) W(z) \tag{11.3}$$

where $\bar{W} = \int_{z \in I'} q_0(z) W(z) dz$ is the normalization constant.

The selected phenotype distribution $q_w(z)$ has a genotype distribution $p_w(y)$. The relative numbers of individual in genotype class $[y, y+dy[$ that are selected is the number of individual in this class developing the phenotype z, multiplied by the selection function of the class $[z, z+dz[$:

$$p_w(y) = \frac{1}{\bar{W}^\dagger} p_0(y) \int_{z \in I'} f(z|y) W(z) dz \tag{11.4}$$

$$= \frac{1}{\bar{W}^\dagger} p_0(y) W^\dagger(y) \tag{11.5}$$

where $\bar{W}^\dagger = \int_{z \in I'} p_0(z) W^\dagger(z) dz$ is the normalization constant. Note that equations (11.4) and (11.2) look similar, but in the latter case, the integration is over the variable z.

We let the selected population to produce the next generation. The genotype is reproduced exactly[6]

$$p_1(y) = p_w(y) \tag{11.6}$$

As before, because of the noise, the phenotypic distribution of the new generation is

$$q_1(z) = \int_{y \in I} p_w(y) f(z|y) dy \tag{11.7}$$

The result we are interested in is the relation between the distributions p_1,q_1 of generation 1, and p_0,q_0 of generation 0, knowing the selection function W and the noise f.

The function $W^\dagger(y)$ is the key to our computations and allows to write the evolution of the genotype as a recurrence relation : $p_{n+1}(y) = \mathcal{L}(p_n(y))$.

Let us note that the normalization constants $\bar{W}$ and $\bar{W}^\dagger$ are equal, which

[5] For example, in figure 11.1, $W(z) = 1$ if $z \in [0.4, 0.6]$, 0 if $z \notin [0.4, 0.6]$.

[6] Recall that we have neglected the finite size of the population and the mutations. Both effects can have major consequences on the fidelity of reproduction.

will slightly simplify are computations :

$$\begin{aligned}\bar{W}^{\dagger}(y) &= \int_y p_0(y) \int_z W(z) f(z|y) dz dy \\ &= \int_z W(z) \int_y p_0(y) f(z|y) dy dz \\ &= \int_z W(z) q_0(z) dz = \bar{W}\end{aligned}$$

11.3 Response to Selection : Breeder's equation.

Operations such as (11.2) and (11.4) are called convolution with kernel f; symbolically, we'll note it by $q = p * f$ and $W^* = W * f$. In the following, we can suppose that the noise and genetic background are independent, so

$$f(z|y) = f(z - y) \tag{11.8}$$

We will also suppose that the noise is symmetric $f(-x) = f(x)$ which implies that $\int_x x f(x) dx = 0$. Finally, because f is a probability density, $\int_x f(x) dx = 1$. It is obvious that such a noise does not induce any difference between the mean phenotype and the mean genotype[7] $\langle z \rangle = \langle y \rangle$, but increases the variance

$$Var(z) = Var(y) + Var(\xi)$$

Broadly speaking, the convolution operation widens and smoothens. Note that the genotype response to selection $p_w(y)$ is smoothened through the noise :

$$\bar{W}.q_w(z) = q(z).W(z) = [(p_0 * f).W](z) \tag{11.9}$$

$$\bar{W}.p_w(y) = p_0(y).W^{\dagger}(y) = [p_0.(f * W)](y) \tag{11.10}$$

The consequence of the above relation is the following : If we apply a selection function to move the average phenotype by S, the average genotype will move to a lesser extent R. Of course, the average genotype is not measurable, but R is also the shift in the phenotype of the next generation. More precisely,

[7] The mean $\langle u \rangle$ of the distribution $p(u)$ is defined by $\langle u \rangle = \int_I up(u) du$. The variance is defined by

$$Var(u) = \int_I u^2 p(u) du - \langle u \rangle^2$$

if we define

$$\begin{aligned} S &= \langle z_w \rangle - \langle z_0 \rangle \\ R &= \langle z_1 \rangle - \langle z_0 \rangle \end{aligned}$$

Then $R < S$. We can write this inequality as

$$R = Sh^2 \tag{11.11}$$

where $h^2 < 1$ is called *heritability.*

The above equation is the central equation of quantitative genetics and of prime importance to farmers. A farmer selects the top 50% of its cows for reproduction ; this selected pool has an increase in milk production of say 5%. Next generation of these cows however will only show a 2% improvement in milk production. The heritability in this case is $h^2 = 0.4$.

Equation (11.11) should however be used with care : h^2 is not a constant, it depends of course on S itself and on the initial distribution p_0 *and* on the selection function W. In some special case, h^2 can be simply related to $Var(\xi)$, but we have to keep in mind that this simplicity *does not* extend to the general case.

11.4 Response to selection: general results.

We saw in the last subsection that the important quantity called heritability. We set the origin at the mean of generation 0, *i.e.* $\langle z_0 \rangle = 0$, then $h^2 = \langle y_w \rangle / \langle z_w \rangle$. Forgetting about the normalization constant[8] $\bar{W}$, we have

$$\langle z_w \rangle = \iint_{z,y} z W(z) p_0(y) f(z-y) dy dz \tag{11.12}$$

$$\langle y_w \rangle = \iint_{z,y} y W(z) p_0(y) f(z-y) dy dz \tag{11.13}$$

To go further, Fourier transforms can be very handy. A function $f(x)$ is related to its Fourier transform $\tilde{f}(q)$ through

$$\tilde{f}(q) = \int_x \exp(-iqx) f(x) dx$$

[8] which will be eliminated in the ratio

$$f(x) \quad = \quad (2\pi)^{-1} \int_q \exp(iqx)\tilde{f}(q)dq$$

Fourier transforms have useful properties : if $f(x)$ is real, then[9] $\tilde{f}(-q) = \tilde{f}^*(q)$; $\int_x x \exp(iqx)f(x)dx = i\tilde{f}'(q)$. Using the Fourier transforms then leads us to

$$\langle z_w \rangle \quad = \quad \frac{i}{2\pi} \int_q \tilde{W}^*(q) \left(\tilde{f}(q)\tilde{p}_0'(q) + \tilde{f}'(q)\tilde{p}_0(q) \right) dq \tag{11.14}$$

$$\langle y_w \rangle \quad = \quad \frac{i}{2\pi} \int_q \tilde{W}^*(q)\tilde{f}(q)\tilde{p}_0'(q)dq \tag{11.15}$$

and the inverse of the heritability is

$$\frac{\langle z_w \rangle}{\langle y_w \rangle} = 1 + \frac{\int_q \tilde{W}^*(q)\tilde{f}'(q)\tilde{p}_0(q)dq}{\int_q \tilde{W}^*(q)\tilde{f}(q)\tilde{p}_0'(q)dq} \tag{11.16}$$

We see here all the complexity encoded into the heritability, which depends on the selection function, the noise and the initial distribution of genotype. There is however one important case where expression (11.16) can be greatly simplified, and this is when both function f and p_0 are gaussian.

Consider the function $n(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ and

$$p_0(x) = \sigma^{-1} n(x/\sigma) \; ; \; f(x) = s^{-1} n(x/s)$$

The Fourier transform reads $\tilde{p}_0(q) = n(\sigma q)$ and $\tilde{p}_0'(q) = -\sigma^2 q n(\sigma q)$. The expression for heritability reads then:

$$h^2 = \frac{\sigma^2}{\sigma^2 + s^2} = \frac{\mathrm{Var}(y)}{\mathrm{Var}(z)} \tag{11.17}$$

The above expression is what is classically found in the literature and is often called the "breeder's equation".

[9] a^* is used here for the complex conjugate of a. $r'(u)$ is used for dr/du.

11.5 Non-gaussian genotype distribution and noise.

Expression (11.17) is not general and depends on the gaussian nature of the noise *and* genotype distribution. It shows however that there are particular cases where the shape of selection function does not enter the heritability. We can look for all such cases, by noting that one necessary condition would be for the noise and the genotype distribution to have the same *shape*, $p_0(x) = a^{-1}f(x/a)$, which implies that $\tilde{p}_0(q) = \tilde{f}(aq)$. More over, let us look for functions such that

$$\tilde{f}'(q)\tilde{f}(aq) = b\tilde{f}(q)\tilde{f}'(aq)$$

where b is a constant. It is easy to check that stretched exponentials

$$\tilde{f}(q) = \exp(-|q|^{\alpha})$$

belong to this class of function[10] if $b = a^{1-\alpha}$. For this case, we have

$$h = \frac{a^{\alpha}}{1 + a^{\alpha}}$$

The gaussian case corresponds to $\alpha = 2$ and $a = \sigma/s$. Let us define

$$n(x) = A\int_0^{\infty} \exp(-q^{\alpha})\cos(qx)dq$$

where A is a normalization constant ; the Fourier Transform of $n(x)$ is a stretched exponential. Let us as in the previous section, define

$$p_0(x) = \sigma^{-1}n(x/\sigma)\ ;\ f(x) = s^{-1}n(x/s)$$

then

$$\begin{aligned} \mathrm{Var}(y) &= \sigma^2 I \\ Var(\xi) &= s^2 I \end{aligned}$$

[10] It can be shown that this is the only class having such a property.

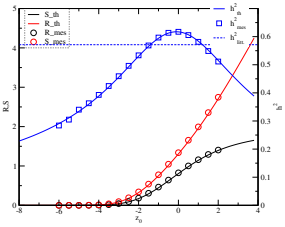


Figure 11.3 – Selection over non-gaussian genotype : $p_0(y) = 1/4\Pi(y/2)$ (uniform distribution over $[-2, 2]$); $W(z) = H(z - z_0)$ (step selection) ; $f(x) = n(x, 0, 1)$. Symbols : numerical simulation over 10^6 individuals. Solid curves : theoretical expressions. Left scale for R,S ; right scale for h^2. We see that the value of h^2 can be markedly different from the classical prediction (dash line).

where $I = \int_x x^2 n(x)dx$. The scaling parameter $a = \sigma/s$ can be written as the ratio of the two variances and

$$h^2 = \frac{[\mathrm{Var}(y)]^{\alpha/2}}{[\mathrm{Var}(\xi)]^{\alpha/2} + [\mathrm{Var}(y)]^{\alpha/2}} \tag{11.18}$$

Expression (11.18) generalizes the heritability expression (11.17) often seen in the literature.

Note however that for general form of the noise and genotype distribution, the selection function enters the expression of the heritability. The problem is, even if the genotype distribution is gaussian in generation 0, it ceases to be so after the first round of selection, because as we saw, $p_1(y) = p_0(y)W^*(y)$ and this product has no particular reason to be gaussian.

11.6 Gaussian selection and noise.

When the genotype and the noise are gaussian, heritability takes a simple form. If the genotype distribution is not gaussian, the classical expression of h^2(eq. 11.17) can be markedly wrong, specially at weak selection limit (fig. 11.3). However, we can derive another exact and simple form for the heritability *if* the selection function and the noise are gaussians, regardless of the exact form of $p_0(y)$. Note that, in Fourier space, expression (...) can be written :

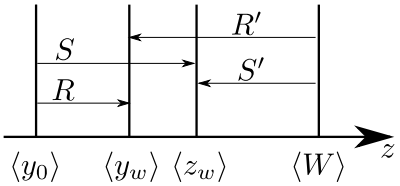


Figure 11.4 – breeder's equation. If the genotype and the noise are gaussian, then $R = h^2 S$. If the selection and the noise are gaussian, then $R' = j^2 S'$.

$$\langle y_w \rangle = \frac{i}{2\pi} \int_q \tilde{p}_0^*(q) \left(\tilde{f}(q)\tilde{W}'(q) + \tilde{f}'(q)\tilde{W}(q) \right) dq \tag{11.19}$$

$$\langle z_w \rangle = \frac{i}{2\pi} \int_q \tilde{p}_0^*(q) \tilde{f}(q)\tilde{W}'(q) dq \tag{11.20}$$

Let us suppose now that both f and W are gaussians of width s and c centered on zero, then their Fourier Transforms are also gaussians and therefore

$$\begin{aligned} \langle y_w \rangle &= \left(c^2 + s^2\right) I \\ \langle z_w \rangle &= c^2 I \end{aligned}$$

And the new breeder's equation reads

$$R' = \frac{c^2 + s^2}{c^2} S' \tag{11.21}$$

We will use j^2 to denote the new heritability relation. Note that $j^2 > 1$ as both $\langle y_w \rangle$ and $\langle z_w \rangle$ are negative, because the origin of phenotype is set at the average of the selection function (Fig. 11.4).

11.7 Adding sex.

As in life, sex complicates the matter under investigation. The main effect of sex here is the modification of inheritance we have considered above. The sexual organisms we consider [11] are now diploid, *i.e.* each gene is present at two copies per cell. One copies is provided by the father, one by the mother. For continuous trait controlled by many genes, the phenotype of

[11] microbes also have a sexual life.

the progeny can be markedly different from those if its parent. Consider for example 4 genes each having two Alleles. Let us suppose that the genotype of the two parents are $ABCD/abcd$ (the slash separetes copies originating from mother and father). We see that a progeny can have a genotype of $ABCD/ABCD$, having inherited only the capital alleles from each parent. Now suppose that each capital letter allele contributes +1 to the phenotype z, and each lower case allele -1. Both parent have $z = 0$, where the progeny displays an extreme phenotype $z = 4$.

To make the matter more complicated, some genes are linked, *i.e.* are found on the same chromosome. As we inherit whole chromosomes, all combination are not possible. For example, if both parent are Ab/aB, *and* the genes are on the same chromosome, then the progeny cannot be AB/AB.

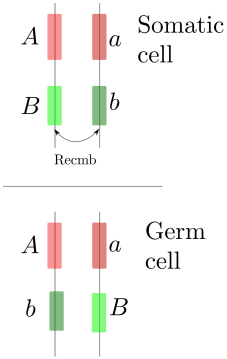


Figure 11.5

And to really muddle the matter, there is recombination. This means that even when two alleles of two genes are on the same chromosome in one parent, they may not be on the same chromosome in the child : there is a cut&paste process which in the germ line, can exchange the place of two alleles on the two chromosome during a process called meiosis. The probability that two alleles of two genes remain on the same chromosome decreases as a function of physical distance between the genes. This was indeed used in the beginning of genetics as a tool to order the genes on the chromosome.

Therefore, the inheritance of a trait looses its general character and now depends on how many genes control the trait and how they are linked. Fortunately, some general assertions remains valid. Indeed, in §11.2, we postulated the inheritance process between selected parent and the progeny (eq. 11.6) as

$$p_1(y) = p_w(y)$$

which is valid only for haploid organism. This is the only expression we have to generalize. Let us introduce the function $L(y|y_a, y_b)$, the probability density for the progeny to have breeding value equal to y given that his parents have y_a and y_b. The exact form of this function will depend on the trait, as we argued above. For a large number of *unlinked* genes for example,

Fisher postulated that the probability should be normal

$$L(y|y_a, y_b) = \exp\left(-(y-\bar{y})^2/2\sigma_g^2\right)$$

where $\bar{y} = (y_a + y_b)/2$ is the mean parental value. This postulate is called the *infinitesimal model*. For our purpose, we don't have to explicit the function L. We only need to state that, in the absence of non-linear effect, the breeding value of the progeny is symmetrically distributed around the mean of its parents :

$$L(y|y_a, y_b) = L(|y-\bar{y}|)$$

Now, having the distribution $p_w(y)$ among the parent, the distribution of genotype among the progeny is

$$p_1(y) = \iint_{\mathbb{R}^2} p_w(y_a)p_w(y_b)L\left(y - (y_a+y_b)/2\right) dy_a dy_b \tag{11.22}$$

Therefore, the mean phenotype of the progeny is

$$\begin{aligned} R = E(Z_1) &= E(Y_1) \\ &= \int_{\mathbb{R}} y p_1(y) dy \\ &= (1/2) \iint_{\mathbb{R}^2} (y_a + y_b) p_w(y_a) p_w(y_b) dy_a dy_b \\ &= \int_{\mathbb{R}} y p_w(y) dy \qquad (11.23) \\ &= \frac{1}{\bar{W}} \iint_{\mathbb{R}^2} y p_0(y) W(z) f(z-y) dy dz \qquad (11.24) \end{aligned}$$

Note that this is exactly what we had in the case of the haploid organisms (eq. 11.13), so the breeder equation and its alternative form remain valid.

11.8 Selection of multiple traits: the G matrix.

It often happens that the same set of genes have influence on multiple traits, for example weight and height. The above computations extend without difficulty to this general case. It is important to have in mind that heritability can be different for different traits; for example, variability on weight is higher than variability on height. The consequence of variable heritability is that in artificial selection, it could be more efficient on a trait a in order to increase

the trait b, instead of selecting directly on the trait b, if heritability on a is higher and if a and b are highly correlated.

Let us note by G_i the set of genes responsible for the trait under consideration Y_k, where

$$Y_k = \sum_i a_i^k G_i$$

For simplicity, we have assumed additivity of the genes. The same genes contributes differently to different traits, which is reflected in coefficients a_i^k. We are going to have many indexes to manipulate, so it will be useful to pack them all into a vector and note, for example, $\mathbf{y} = (y_1, y_2, ...y_M)$, where M is the number of traits under consideration.

We will note $p(\mathbf{y})$ the probability density of genotype[12] $\mathbf{y}$, which relates to phenotype through

$$\mathbf{z} = \mathbf{y} + \xi$$

the distribution probability of the noise ξ being given by the function $f(\mathbf{z}|\mathbf{y})$. $W(\mathbf{z})$ is the selection function and should be understood as follow : the proportion $W(\mathbf{z})$ of individuals with traits 1 in class $[z_1, z_1 + dz_1][$, trait 2 in class $[z_2, z_2 + dz_2[$, ... will be selected for reproduction in the next generation, where $\mathbf{z} = (z_1, ...z_M)$.

Everything we said can now be repeated, replacing scalars everywhere by vectors. Let us the genotype of generation 0 be $p_0(\mathbf{y})$, and $q_0(\mathbf{z})$ its phenotype[13]:

$$q_0(\mathbf{z}) = \int_{\mathbf{y}} p_0(\mathbf{y}).f(\mathbf{z}|\mathbf{y})d^M\mathbf{y}$$

after selection, the distribution of phenotypes become

$$q_w(\mathbf{z}) = \bar{W}^{-1}\, q_0(\mathbf{z})W(\mathbf{z})$$

where $\bar{W}$ is the usual normalization constant. The genotype background of the selected individuals is

$$\begin{aligned} p_w(\mathbf{y}) &= \bar{W}^{-1} \quad p_0(\mathbf{y}).\int_{\mathbf{z}} f(\mathbf{z}|\mathbf{y})W(\mathbf{z})d^M\mathbf{z} \\ &= \quad p_0(\mathbf{y}).W^\dagger(\mathbf{y}) \end{aligned}$$

[12] of course, the term genotype is abusif. The exact term would be the phenotype in the absence of noise.

[13] of course, $\int_{\mathbf{y}} d^M\mathbf{y}$ is just a multiple integral and means $\int_{y_1}\int_{y_2}...\int_{y_M} dy_1 dy_2 ... dy_M$.

the genotype of generation 1 is duplicated

$$p_1(\mathbf{y}) = p_w(\mathbf{y})$$

and results in the phenotype

$$q_1(\mathbf{z}) = \int_{\mathbf{y}} p_1(\mathbf{y}).f(\mathbf{z}|\mathbf{y})d\mathbf{y}$$

By assuming independence of noise from genotype, we can write $f(\mathbf{z}|\mathbf{y}) = f(\mathbf{z}-\mathbf{y})$. The Fourier transform of course generalizes to vectors

$$\tilde{f}(\mathbf{q}) = \int_{\mathbf{x}} f(\mathbf{x}) \exp(-i\mathbf{q}.\mathbf{x})d^M\mathbf{x}$$

where $\mathbf{q}.\mathbf{x} = \sum x_i q_i$ is the scalar product of these two vectors. We can pack the partial derivatives of $\tilde{f}(\mathbf{q})$ into a vector called gradient

$$\begin{aligned} \nabla\tilde{f}(\mathbf{q}) &= \left(\partial\tilde{f}/\partial q_1, ..., \partial\tilde{f}/\partial q_M\right) \\ &= -i\int_{\mathbf{x}} \mathbf{x}f(\mathbf{x}) \exp(-i\mathbf{q}.\mathbf{x})d^M\mathbf{x} \end{aligned}$$

following exactly our footsteps of the previous chapter, we get

$$\langle \mathbf{z}_w \rangle = \langle \mathbf{y}_w \rangle + \frac{i}{2\pi}\int_{\mathbf{q}} \tilde{W}^*(\mathbf{q})\nabla\tilde{f}(\mathbf{q})\tilde{p}(\mathbf{q})d^M\mathbf{q} \tag{11.25}$$

$$\langle \mathbf{y}_w \rangle = \frac{i}{2\pi}\int_{\mathbf{q}} \tilde{W}^*(\mathbf{q})\tilde{f}(\mathbf{q})\nabla\tilde{p}(\mathbf{q})d^M\mathbf{q} \tag{11.26}$$

To go further, we should introduce a little tensorial notations. Of course, we could do the same computations by considering directly the components of our vectors such as y_i or z_k, but it is much more efficient to pack these coefficients into vectors and tensors. We use here the term tensor as tensors of rank 2, *i.e.* matrices.

Given two vectors $\mathbf{u} = (u_1, ...u_M)$ and $\mathbf{v} = (v_1, ...v_M)$, their scalar product is a scalar

$$\mathbf{u}.\mathbf{v} = \sum_i u_i v_i$$

and their tensorial product is the tensor

$$(\mathbf{u} \otimes \mathbf{v})_{ij} = u_i v_j$$

If we imagine vectors as column vectors, the $\mathbf{u}.\mathbf{v} = u^\dagger v$ and $\mathbf{u} \otimes \mathbf{v} = uv^\dagger$. Now, if the probability density of $\mathbf{y}$ is $p(\mathbf{y})$, then its average is given by

$$\langle \mathbf{y} \rangle = \int_{\mathbf{y}} \mathbf{y} p(\mathbf{y}) d^M \mathbf{y}$$

and it is a vector. It's variance is given by the tensor

$$\mathrm{Var}(\mathbf{y}) = \int_{\mathbf{y}} \mathbf{y} \otimes \mathbf{y} p(\mathbf{y}) d^M \mathbf{y} - \langle \mathbf{y} \rangle \otimes \langle \mathbf{y} \rangle \tag{11.27}$$

Usually, the diagonal elements of $\mathrm{Var}(\mathbf{y})$ are called variances and its off-diagonal elements covariances. For example, in simple scalar notation, we will write

$$\mathrm{Cov}(y_i, y_j) = \int_{\mathbf{y}} y_i y_j p(\mathbf{y}) d^M \mathbf{y} - \left(\int_{\mathbf{y}} y_i p(\mathbf{y}) d^M \mathbf{y} \right) \left(\int_{\mathbf{y}} y_i p(\mathbf{y}) d^M \mathbf{y} \right)$$

Notation (11.27) is just a more efficient way of writing the same thing. Now, let us suppose that the distribution $p(\mathbf{y})$ is, broadly speaking, gaussian :

$$p(\mathbf{y}) = A. \exp\left((-1/2)(\mathbf{y} - \mu).\hat{G}^{-1}(\mathbf{y} - \mu) \right)$$

where A is the normalization constant and $\hat{G}^{-1}$ a second rank tensor. Then, a small amount of computation shows that

$$\begin{aligned} \langle \mathbf{y} \rangle &= \mu \\ \mathrm{Var}(\mathbf{y}) &= \hat{G} \end{aligned}$$

Let us suppose that $p_0(\mathbf{y})$ is a normal distribution with mean $\mu = 0$[14]. Its Fourier transform reads

$$\begin{aligned} \tilde{p_0}(\mathbf{q}) &= A \exp\left((-1/2)\mathbf{q}.\hat{G}\mathbf{q} \right) \\ \nabla \tilde{p_0}(\mathbf{q}) &= A \left(\hat{G}\mathbf{q} \right) \exp\left((-1/2)\mathbf{q}.\hat{G}\mathbf{q} \right) \end{aligned}$$

Let us moreover suppose that the noise also is a centered gaussian

$$\tilde{f}(\mathbf{q}) = A \exp\left((-1/2)\mathbf{q}.\hat{N}\mathbf{q} \right)$$

[14] As we have set the mean of generation 0 genotype at as the origin.

$$\nabla \tilde{f}(\mathbf{q}) \quad = \quad A\left(\hat{N}\mathbf{q}\right) \exp\left((-1/2)\mathbf{q}.\hat{N}\mathbf{q}\right)$$

If we note $\hat{P} = \hat{G} + \hat{N}$, we see that the averages (11.25,11.26) simply read :

$$\langle \mathbf{y}_w \rangle = \hat{G}\hat{P}^{-1} \langle \mathbf{z}_w \rangle \tag{11.28}$$

The above expression generalizes the heritability to multiple traits, where the scalar h^2 is replaced by the tensor $\hat{G}\hat{P}^{-1}$. The tensor $\hat{G}$ is often called the genotype variance matrix or simply the G matrix. $\hat{P}$ is referred to as the matrix of phenotype variance.

Let us again stress that the generalized breeder's equation (11.28) applies only when genotype and phenotype distributions are normal.

11.9 Fisher's 'fundamental theorem'

It has been said about the Fisher's fundamental theorem that is neither fundamental nor a theorem. Many still wonder what really Fisher meant when he wrote

> "The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time."

In terms of what we developed in the previous chapters, this statement is a triviality. Let us set heritability $h^2 = 1$, in which case we don't have to make any distinction between genotype and phenotype. Beginning with the distribution $p_0(y)$, the selected population distribution is

$$p_w(y) = \bar{W}^{-1} p_0(y) W(y)$$

and the next generation inherits this distribution $p_1(y) = p_0(y)$. The *mean fitness* of the population is the average of the selection function :

$$\langle W \rangle_i = \int_y W(y) p_i(y) dy$$

and we observe that $\langle W \rangle_0$, the mean fitness of Generation 0 is just the normalization coefficient $\bar{W}$ we use. The change in the mean fitness is

$$\begin{aligned} \Delta W &= \langle W_1 \rangle - \langle W_0 \rangle \\ &= \frac{1}{\bar{W}} \int_y W^2(y) p_0(y) dy - \int_y W(y) p_0(y) dy \end{aligned} \tag{11.29}$$

The *'genetic variance in fitness '* is

$$\sigma_W^2 = \int_y W^2(y) p_0(y) dy - \left(\int_y W(y) p_0(y) dy \right)^2$$

so we can write expression (11.29) as

$$\Delta W = \frac{\sigma_W^2}{\bar{W}} \tag{11.30}$$

If the fitness of generation 0 is taken as the reference ($\langle W \rangle_0 = 1$) Fisher's theorem is exact (for this round of selection). More generally, the *relative rate of increase* in fitness is just the square of the coefficient of variation :

$$\frac{\Delta W}{\bar{W}} = \left(\frac{\sigma_W}{\bar{W}} \right)^2 \tag{11.31}$$

Let us stress that the above equation as such is useless to describe the long term evolutionary dynamics because the variance σ_W^2 also varies (decreases) from one round of selection to the next. Following the same arguments, we will find that the variation of the second moment depends on the third moment and so on. We can build a whole hierarchy of moments, where variation of the $k-$th moment depends on the $k+1$-th moment. It is not obvious however that this approach has any advantage over writing directly the distribution of the $k-$the generation as

$$p_k(y) = \frac{1}{Z} p_0(y) W^k(y)$$

We can follow the argument of previous chapter and develop a multidimensional version of the above statement. The more fruitful development however is to study the inclusion of heritability into the above computation, which is the subject of the next chapter.

Central framework, additive genes, environment noise (stochastic gene expression). The breeder's equation. Fisher's fundamental Theorem.

12 Sexual selection and the Fisher Divergence.