An introduction to adaptive evolution in population genetics based on fitness flux theory

by

Pablo Mateu Hoyos

Library-based Project
ICFP Master's in Physics
École Normale Superiéure - PSL

Year 2022-2023

Supervisor 1: Andrea Mazzolini
Supervisor 2: Aleksandra Walczak

Contents

Abstract

Evolution of species is a hugely developed experimental fact whose first formulations date back to the XIXth century. However, evolution takes place in time scales that make its direct observation almost impossible. Only with an appropriate theoretical background the current state of a population can be interpreted and subsequent predictions on its future dynamics can be made. The aim of this work is to provide an introduction into the theoretical formulation of evolution based on population genetics. Concretely, we will focus on evolutionary adaptation. Our ultimate goal will be to derive a quantity able to reproduce the insights of adaptation in evolution driven by genetic drift, mutations and time-dependent natural selection. For that, we will define and prove the fitness flux theorem that establishes fitness flux as a universal measure of adaptation in such conditions.

1 Introduction

The modern history of evolution dates back to the early XIXth century, when Jean-Baptiste Lamarck proposed what is considered the first fully formed theory of evolution. It was later on, in 1859, when Darwin published its *On the origin of the species*, in which he stressed that evolution was adaptive, that is, always towards an increase in the capability of survival. Darwin succeeded in spreading the concept of evolution among the scientific community, but it wasn't until the developments in biology that took place between the 1920s and the 1940s that natural selection, the evolutionary mechanism he introduced, was accepted. Simultaneously, based on his experimental studies at the beginning of the XXth century, Mendel formulated the first version of Mendelian genetics. The first studies integrating Darwinian evolution with Mendel's laws of genetics using statistical approaches were performed during the decades of 1920 and 1930 and have three proper names: R. A. Fisher, J. B. S. Haldane, and S. Wright. In their attempt, Fisher, Haldane and Wright gave birth to a new discipline of biology: population genetics.

Population genetics studies the genetic basis of evolution. Hence, it deals essentially with two aspects: Mendel's laws of segregation and the characterisation of the evolutionary forces acting on populations. Concretely, population genetics aims to describe not only what are these evolutionary forces, but also the mechanisms through which they reflect on populations and lead species to changes in their genetic structure over generations. Population genetics differs from many other fields of biology in that its insights are essentially theoretical, rather than observational or experimental. It is necessarily so: as time scales of evolution are typically greater than lifetimes, it is in general difficult to explore directly the effects of evolution. However, population genetics is a highly contrasted field that has been tested in several experiments, especially over the last years, for biological systems with very fast evolutionary time scales. One important example is the *Escherichia coli* long-term evolution experiment carried out by Richard E. Lenski since 1988.

Ever since its introduction, population genetics has been extended and integrated with several disciplines from biology, mathematics and physics. The objects of study in population genetics are mainly frequencies of alleles or genotypes (broadly, the number of individuals within a population with the same genetic structure) and by «evolution» is meant their variation in time. Therefore, evolution, as understood from population genetics, is a compilation of several phenomena that affect the population as a whole and correspond to a wide range of complex mathematical models in which multiple forces other than just natural selection participate actively. In 1930, Fisher formulated his first version of the Fundamental Theorem of Natural

Selection, in which he provided a mathematical basis for adaptation under natural selection. In general, not all evolutionary forces lead to adaptive evolution, but, as long as natural selection is governing the process, evolution should be adaptive. Here, we apply the prescription of evolution provided by population genetics to obtain a more general characterisation of adaptation. As a consequence, the structure of the text is divided in three parts.

In first place, we introduce the simplest non-trivial theory of evolution based on population genetics [1, 2]. In the context of this work, it is not the complexity of real populations but the key features of the process that we are interested in. Here, evolution is described as a combination of three evolutionary forces: mutations, random genetic drift and natural selection, each of them represented by a fundamental model of evolution, for which the effect these forces have on the population is quantified by computing the variation of allele frequencies over time. Then, we formulate Fisher's theorem and discuss the insights of its non-validity [3, 4]. Finally, we integrate the three forces in one single Equation, the Kimura-Ohta Equation.

In the second part, we are concerned with the statistical framework for such a model of evolution. The properties of the Kimura-Ohta Equation motivate the necessity of some concepts on stochastic thermodynamics for microscopic systems out of equilibrium. We discuss the stochastic behaviour of populations under mutations, genetic drift and natural selection, and we derive a set of prescriptions and theorems to deal with such systems and calculate macroscopic averages corresponding to the dynamics of their evolution [5].

As of now, we have fully developed the evolutionary and statistical theories representative of the population subjected to genetic drift, mutations and selection. The last step is to integrate the two of them to find a mathematical expression of adaptation. Following Fisher's intuition, we define a new magnitude, the (cumulative) fitness flux Φ, for which we deduce an analogous version of Fisher's theorem, the fitness flux theorem, in which the average fitness flux is shown to be an always increasing magnitude, thus an almost universal measure of adaptation [6]. To conclude, comments on the interpretation and validity of the theorem are done, as well as on its generalisation to more realistic populations, more complex genetic structures or other evolutionary effects, and its reduction to Fisher's theorem in the strong selection limit.

Four Appendices are included to provide the reader a more detailed calculation of some important quantities and results. There is no further theory here, so they shouldn't affect the overall understanding of the text.

2 Basic concepts of genetics

The definitions, examples and concepts of this section have been mainly extracted from references [1] and [7].

Population genetics is built around the notion of *gene*. There are mainly two ways of defining what a gene is. From the physical (in the sense of information) point of view, a gene is the elementary unit of hereditary information, that is, the entity transmitted from parents to offspring during the reproductive process that influences the hereditary traits. Thus, all the genetic information of one individual is ultimately contained in their genes, whose interaction with the individual's environment determines the different morphological, physiological and behavioural traits. Genes may exist in a certain number of alternative forms called *alleles*. For example, the gene that determines eyes color can take different alleles, each corresponding to one possible color. The variety of allele composition within a population is called *genetic variation* or genetic diversity. When variation is such that at least more than one allele is present in a proportion superior to 1%, genetic variation can be referred to as *polymorphism*, and its

study is of fundamental importance in population genetics.

From the biochemical point of view, a gene is a coding or functional region of the *DNA*. DNA is the genetic material of every being. It consists of a double-helix-shaped molecule formed by two long linear strands of four types of nucleotides: adenine (A), thymine (T), cytosine (C) and guanine (G). Each of the nucleotides of one strand is connected to one and only one nucleotide of the opposite strand, such that only A-T and C-G pairings are allowed. Therefore, the chemical composition of one strand is entirely determined by the chemical composition of the other one. The global sequence of nucleotides is called *code*, and a code that serves for one function is said to *code for* that function.

DNA contains the genetic information necessary for the cells to sythesise proteins, which are fundamental for most of the structure and physiological processes carried out by organisms. In this process, called *genetic expression*, cells *transcript* the information contained in the DNA sequence, *process* it and *translate* it into specific actions. However, not all the DNA sequence codes for the same function. Moreover, some fraction of the DNA sequence is non-coding and some other fraction might be impossible to transcript. But for cells to carry out concrete tasks it is necessary to receive concrete and readable instructions, meaning that they will only use sections of DNA coded for that task. These sections of DNA that are possible to transcript and code for a known function is what we call genes, and it is in this sense that we claim genes are the most basic units of genetic information. In this picture, alternative alleles are nothing but permutations of the nucleotide sequence that leave the gene function unchanged.

DNA filaments are grouped together in bigger microscopic structures called *chromosomes*. The position of a gene along a chromosome is called the *locus* of the gene. Every cell contains multiple types of chromosomes (23 in the human case), each of which is replicated in *diploid* organisms, due to inheritance from two parents. In such organisms, chromosomes appear in pairs and, therefore, each locus presents two copies of the same gene or, equivalently, two alleles. Whenever the two alleles are the same, the organism is said to be *homozygous* at the locus under consideration. Otherwise, it is said to be *heterozygous*.

The ensamble of alleles at different loci that affect one trait is called *genotype*. The genotype contains the genetic constitution of the organism. For simplicity, and throughout the work, we will restrict ourselves to one locus models and we will neglect multiloci effects. Thus, for a certain gene, we have as many genotypes as possible combinations of two we can form with the k possible alleles for that gene. The expression of an observable trait of a certain genotype is called its *phenotype*. The phenotype is essentially the result of the genetic expression of the genotype due to its interaction with the environment. As a consequence, identical genotypes can lead to different phenotypes under strong influence of the environment.

To clarify this difference, consider the following mental experiment: let's imagine two people with the same genotype for hair color. We call genotype the combination of alleles that determine the hair color and phenotype the hair color itself, which is its physical manifestation. Exposed to the same environment, both of them will have the same hair color. However, if one of them is constantly exposed to solar radiation, their hair will get blonder, leading to a change in phenotype.

Given a trait and its associated genotypes, whether each genotype has a single, unique expression of the trait depends on the way the two alleles at the locus interact with each other. This relationship is referred to as *dominance*, and determines the phenotypic effects the presence of each allele has over the others. We say one allele is completely dominant when the phenotypic effect of the heterozygote is indistinguishable from the phenotypic effect of the associated homozygote. Whenever we have a different situation we say there is either incomplete dominance or codominance.

To recapitulate, consider a gene with two alleles, A_1 and A_2[1]. We shall also assume the organism is diploid and a single locus model, such that the genotype depends only on the composition of the locus under consideration. The possible genotypes for this gene are $\{A_1A_1, A_1A_2, A_2A_2\}$. Organisms with genotypes A_1A_1 and A_2A_2 are homozygous for the gene at the locus studied, while A_1A_2 organisms are heterozygous. We say A_1 is dominant over A_2 if the A_1A_2 phenotype is closer to A_1A_1 than it is to A_2A_2. Dominance is complete when the phenotypes associated to A_1A_1 and A_1A_2 cannot be distinguished and incomplete when when the result is an intermediate configuration. Similarly, if the product of both alleles is manifest in the heterozygote, they are said to be codominant.

3 Population genetics

As we have stressed before, population genetics deals with the quantitative description of the evolution of the genetic structure of populations. In the previous section we have defined the elements needed to understand what is meant by genetic structure. Now we want to answer how genetic structure is quantified and how does it evolve over generations.

Genetic structure is typically represented by a set of allele or genotype *frequencies*. Ideally, if we are able to give the full set of allele frequencies at the locus under consideration, then we will have characterised the whole genetic structure at that locus. However, allele frequencies are not constant in time. There exist several evolutionary forces that act on this quantities and lead them to variations that ultimately change the genetic structure of populations.

Typically, the main evolutionary forces are considered to be four: mutations, random genetic drift, natural selection and migration. Since we are only interested in single locus space-independent evolution, we will ignore migration in the following. In this section, thus, we will present what are this evolutionary forces and how we can model their mathematical structure, computing their effect on allele frequencies and discussing their properties, behaviour and their global effect on populations over time. Eventually, this reasoning will lead us to the formulation of one of the fundamental results in XXth century population genetics: Fisher's fundamental theorem of evolution.

For that, we will essentially follow the explanations and theoretical calculations of references [1] and [2].

3.1 Allele and genotype frequencies

Before trying to explain the way evolutionary forces act on a population, it is important to have a clear picture of what is understood by allele and genotype frequencies.

Consider a population of N diploid individuals and consider the same locus in every individual[2]. Consider also, for simplicity, that the gene under study has only two possible alleles, namely (A_1, A_2) and, therefore, three possible genotypes: $\{A_1A_1, A_1A_2, A_2A_2\}$, whose relative or *genotype frequencies* are defined as:

$$\begin{aligned} x_{11} &= n_{11}/N \\ x_{12} &= n_{12}/N \\ x_{22} &= n_{22}/N \end{aligned} \tag{3.1}$$

[1]Alleles are typically denoted as $(A_1, ..., A_k)$, but for two-allele genes is also common to see the notation (A, a) in many references, which makes the dominance explicit. We will restrict in the future to the notation used in the text.

[2]Recall that locus refers to a position along the chromosome. Thus, we can find the same locus in the two homologous chromosomes of an individual.

where n_{ij} is the total number of genotypes A_iA_j, and the frequencies satisfy the normalisation condition $x_{11} + x_{12} + x_{22} = 1$.

Since every genotype consists of two alleles (one from the father and one from the mother) it is natural to think that both genotype and allele frequencies must be related. Moreover, one often refers to a sample of $2N$ different[3] alleles at the same locus and it can be interesting, even necessary, to characterise their frequencies. From (3.1), we can define the quantities

$$p = \frac{n_1}{2N} = x_{11} + \frac{1}{2}x_{12} \tag{3.2}$$

and

$$q = \frac{n_2}{2N} = x_{22} + \frac{1}{2}x_{12},$$

which represent the *allele frequencies* for A_1 and A_2, respectively. As well as the genotype frequencies, allele frequencies satisfy $p + q = 1$. In this case, n_i is the number of alleles A_i within the population.

Allele frequencies can be interpreted in two ways. One is simply, by construction, as the fraction of the allele in the entire population. The other one is slightly more subtle and is as the probability of obtaining an allele of the considered type by picking an allele at random from the population.

In the following, unless the opposite is specified, all derivations and models will be developed for genes with two alleles. The results obtained will always have a natural adaptation to multiple k alleles and their physical interpretation will remain essentially unchanged in doing so. Thus, for the scope of this work, it will be more interesting to work with models of two alleles and generalise the results afterwards, if needed, without loss of generality.

3.2 The Hardy-Weinberg model

Mendel's laws hold independently of the mechanisms involved in the evolution process. Subsequently, even if there are no evolutionary forces such as drift, mutation or selection, mating still takes place and allele and genotype frequencies can behave non-trivially from one generation to the next one. In fact, as we shall see, genotype frequencies are partially determined by the pattern of mating. Therefore, prior to analyse one by one the most important evolutionary forces, one should concentrate on the zero case: evolution of genotype frequencies under mating.

The most simple (and important) model is the Hardy-Weinberg model. The Hardy-Weinberg model states that genotype frequencies can be expressed in terms of allele frequencies (and viceversa) through a simple mathematical relation:

$$x_{11} = p^2 \quad , \quad x_{12} = 2pq \quad , \quad x_{22} = q^2, \tag{3.3}$$

Note that Equation (3.3) is a particular (invertible) case of (3.2). Indeed, the Hardy-Weinberg law is valid only under a series of important assumptions:

- Random mating and infinite population size. Random mating states that reproductive mates are chosen with complete ignorance of their genotypes. In other words, the genetic structure of the offspring is not biased by a particular choice of the parents. For example, if homozygous individuals were to mate only with other homozygous individuals mating

[3] In this context, different means that the alleles come from the same locus in separate chromosomes. We say the alleles are *different by origin*. For example, the two alleles of the same individual are necessarily different by origin. If they are also different in their DNA sequence, they are said to be *different by state*

would not be random. Furthermore, population size needs to be very big, in theory infinite.

- Non-overlapping generations. A non-overlapping generations model states that the birth, maturation and death cycles of one generation, say t, and the next one, $t+1$, do not cross each other: t individuals must die before $t+1$ individuals mature.

- Absence of evolutionary forces. Allele frequencies are not affected by selection, mutation, random genetic drift, migration or any other type of evolutionary force.. The only driving process involved is reproduction. As we shall see, however, drift is intrinsic to finite populations, hence the necessity for population size to be infinite.

Additionally, individuals need to be diploid[4], reproduction sexual, allele frequencies sex-independent and the considered gene bi-allelic.

The Hardy-Weinberg model has some important implications. The first one is related to the random mating. A closer look reveals that genotype frequencies (3.3) coincide with the probabilities of picking the two alleles at random from the population: alleles are treated as a *pool* of $2N$ equivalent elements and, due to random mating, every choice of k alleles is a collection of k independent events.

Secondly, random mating between $\{A_1A_1, A_1A_2, A_2A_2\}$ individuals has six possible offspring outcomes. By using Mendel's laws to compute all these outcomes and their related frequencies, it can be shown that

$$x'_{11} = p^2 \quad , \quad x'_{12} = 2pq \quad , \quad x'_{22} = q^2, \tag{3.4}$$

and, therefore, using (3.2)

$$p' = p \quad , \quad q' = q, \tag{3.5}$$

which means not only that Hardy-Weinberg frequencies represent an equilibrium state, but also that after a perturbation from equilibrium Hardy-Weinberg proportions are reached again after one generation of random mating.

Finally, due to the way it is formulated, the Hardy-Weinberg model separates life history in two steps: the genetic recombination (offspring) and the maturation process. When more complex models are formulated, this allows to keep the results obtained for the first step and consider the variations introduced in allele frequency only acting on the second step. In this sense, the model is similar to many physical models, such as non friction free fall or the ideal harmonic oscillator, in which complications are always introduced over the basis of the ideal model, and constitutes, in spite of its apparent oversimplifications, a reference model. Moreover, in many species, evolution time scale is too big for evolutionary forces to be directly observed. Hence, in most of the cases, experimental observations will agree with an apparent Hardy-Weinberg equilibrium.

To conclude, it is interesting to underline that when male and female frequencies are different, equilibrium is still reached after two generations, rather than one, and that Hardy-Weinberg law can be easily generalised to k-allelic genes.

3.3 Mutations

The first evolutionary force we are going to consider is *mutations*. Mutations can be defined as the spontaneous change of one allele to a new, different allele. From the biochemical point

[4]This is not strictly true. HW applies also to the diploid phase of an individual whose state of reproductive maturity is haploid. For simplicity, we will restrict to the diploid assumption.

of view, mutations are typically a result of a variation in the nucleotide sequence that results in the gene changing from one allele to another. Although the internal mechanism may be considered partially random, standard population genetics treats mutations in a deterministic way. The action of mutations over allele frequencies is called *mutation pressure*.

While genetic drift, as we will see later on, guides evolution towards *fixation*, mutations lead to diversification in allele composition. In this sense, mutations are essential to increase polymorphism and represent the ultimate source of genetic variation within a population. Mutations are always present, always possible, but their evolutionary force is so small that even for large populations it takes many generations to observe a significant effect.

In order to characterise mutation pressure we can start from a Hardy-Weinberg model in which mutation between alleles is permitted. Let's consider, then, a population of N diploid individuals, with N large enough to neglect changes in allele frequency due to chance[5], and such that the gene at the locus studied can adopt two alleles, A_1 and A_2, which have frequencies p and q. One introduces now the *mutation rate* μ as the number of A_1 alleles that mutate to A_2 per generation, and ν as the analogue for the inverse process. Mutation rates can be interpreted as probabilities of mutation and their typical values range from 10^{-6} to 10^{-4}.

Consider that we start at generation t, when the A_1 allele frequency is p_t. An allele A_1 in generation $t+1$ can be obtained in two complementary ways from generation t: we pick an A_1 allele and it does not mutate, or we pick an A_2 allele and it mutates. Since mutation and random choice of alleles can be considered independent processes, the allele frequency of A_1 in generation $t+1$ is given by

$$p_{t+1} = (1-\mu)p_t + \nu q_t. \tag{3.6}$$

An often more illustrative representation is the variation in allele frequency after one generation, $\Delta_m p \equiv p_{t+1} - p_t$, where the subscript m stands for we are only considering mutations,

$$\Delta_m p = -\mu p_t + \nu q_t. \tag{3.7}$$

In this case, after some algebra, we can obtain the time Equation for the A_1 allele frequency in terms of the initial condition,

$$p_t = \frac{\nu}{\mu+\nu} + \left(p_0 - \frac{\nu}{\mu+\nu}\right)(1-\mu-\nu)^t. \tag{3.8}$$

One illustrative solution is plotted in Figure 1 for the case $\mu \sim \nu$. An important implication of Equation (3.8) is that balance between the two flows is reached with a typical time $\tau = -1/\ln(1-\mu-\nu)$. Hence, for reversible mutation there is always an equilibrium value

$$\hat{p} = \frac{\nu}{\nu+\mu}. \tag{3.9}$$

Since both τ and $\hat{p}$ depend only on the mutation rates, equilibrium under mutation is always reached after the same approximate time with independence of the initial allele frequency.

3.4 Random genetic drift

So far we have considered models in which the size of population was infinite. However, real populations, even if usually very large, are finite. *Random genetic drift* is the force that

[5]Finite population effects, the most important one among which is random genetic drift, can be understood as border effects. Thus, we can neglect them as long as we are far from their ratio of action.

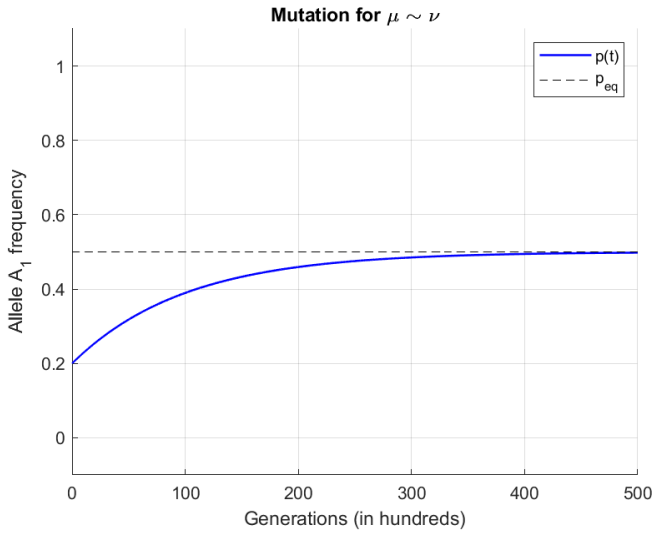


Figure 1: Allele frequency evolution for initial value $p_0 = 0.5$ and mutation rates $\mu = 2\nu = 3 \times 10^{-5}$. An equilibrium value is reached. As it can be seen, the typical equilibrium time is of the order of a few tenths of thousands of generations.

arises when the intrinsic randomness of evolutionary process is mixed with the non-infinity of population size.

Since typical populations are very large, one might think that finite population effects could simply be neglected. Obviously, this is the case when just evolution over a few generations is considered and, even if finite, population size is both large and stable. But this could be said also for mutations. In general, when acting on stable populations, all evolutionary forces need a typical time of thousands or tenths of thousands of generations to express themselves. Thus, even if in short-time scales it can be usually ignored, random genetic drift is a cumulative force whose long-time effect is crucial to understand how evolution works and how dominance between all the forces involved is distributed. Conversely, random genetic drift becomes extremely important for species with small (some hundreds or thousands of individuals) population size. For that reason, random genetic drift is the solely force able to explain and give immediate prediction for some extreme situations such as *bottlenecks* or massive extinctions.

We have stressed before that, contrary to genetic drift, mutations tend to increase polymorphism in populations, introducing in the while the concept of fixation. We aim to explain now what is meant by fixation and why drift removes variation from populations.

3.4.1 The Wright-Fisher model

Consider, once again, a population of N diploid individuals. We also assume random mating and constant and finite N, as well as non-overlapping generations. Therefore, the population at generation t can be considered as a pool of $2N$ alleles, all of which die and are replaced when moving towards the next generation. Again, the gene at the locus under study can adopt two alleles, A_1 and A_2, with frequencies p and q.

Each allele in generation $t+1$ is the result of a binomial choice in generation t: we run over the whole population and ask, allele per allele, whether it is copied into the next generation or not. Thus, if the frequency of A_1 in generation t is $p = n_1/2N$, the probability of having a frequency $p' = n_1'/2N$ for A_1 in $t+1$ is

$$\mathcal{P}(p'|p) = \binom{2N}{2Np'} p^{2Np'} q^{2Nq'} \equiv T_{p'p}. \tag{3.10}$$

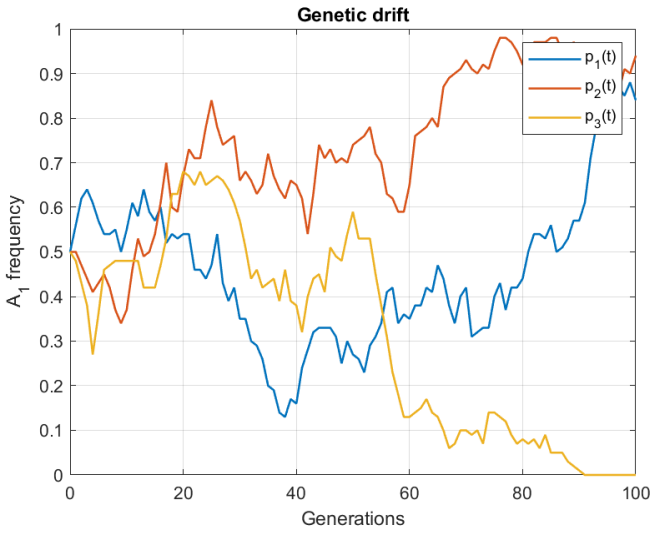


Figure 2: Three different computer simulations for the A_1 frequency under Wright-Fisher's model, with $p_0 = 0.5$ and a $N = 100$ population. Fixation is reached for allele A_2 in the yellow one, while it is almost reached for A_1 in the blue and red curves. Randomness is clearly reflected.

Equation (3.10) implies that, in general, even if we start with a defined p_0, our prediction after one generation is a probability distribution for all possible frequencies, rather than one concrete value. When building the second generation, this means that each value for the frequency can be obtained, with different probability, from all the possible values from the first generation. Hence, it is convenient to define the total probability of obtaining a frequency p' of the allele A_1 at generation $t+1$ given, not a frequency, but a probability distribution of frequencies at t,

$$\mathcal{P}(p') = \sum_p T_{p'p}\, \mathcal{P}(p). \tag{3.11}$$

Using Equation (3.11) it is possible to plot some examples of evolution under random genetic drift. The result is shown in Figure 2.

Fixation is the most significant feature of genetic drift. It can be defined as the process for which the frequency of one allele, after many generations, goes to one, causing the extinction of the other allele(s) in the population. In bi-allelic populations, extinction of one allele is due to fixation of the other and viceversa, which is not true anymore in populations with multiple alleles. In any case, extinction and fixation correspond to the only steady-states of Equation (3.11). This means that, even if randomness prevents anticipating the state of the population at generation t, different alleles will progressively extinct until one allele eventually fixes the population and homozygosity is reached.

The probability of fixation of A_1, $\pi(p_0)$, being p_0 the initial allele frequency, can be easily reasoned. When fixation is reached, all the alleles are identical by state. But not only: due to the way generations are replaced in the Wright-Fisher (W-F) model, the alleles must be also identical by descent, that is, they must share one common A_1 ancestor. The probability one specific allele fixates can then be mapped to the probability of picking that one specific allele, which is $1/2N$. Now, if we have n_0 alleles of type A_1, each of them leads to the same state of fixation. Thus, the total probability of fixation of A_1 is simply

$$\pi(p_0) = \frac{n_0}{2N} = p_0. \tag{3.12}$$

The existence of a non-zero probability of fixation implies that ultimately genetic drift always removes polymorphism from populations. Whether it takes longer or shorter depends on the population size and (on average) on the closeness of the initial configuration of alleles to the steady-state reached. Hence, while mutations introduce constantly new alleles that keep variation within species, genetic drift moves populations towards homozygosity.

To understand this better, consider a large population such that the gene at the locus can adopt four alleles. Mutation acts in every direction, with specific rates of mutation, while drift, when adding two more alleles with respect to our previous formulation, can act in two ways: fixation of one of them or, more likely, extinction of one of them. Whenever one allele is removed from the population, genetic variation is by definition reduced. However, mutation rates from the surviving alleles towards the extinct one are generally nonzero, meaning that still present alleles can mutate and «create» copies of the lost allele. In this sense, mutations restore genetic variation.

3.4.2 The neutral theory

Being inverse forces, it is reasonable to ask whether evolution can be modelled by the combination of mutations and random genetic drift. The branch of population genetics that deals with the interaction between genetic drift and mutations is called *neutral theory*. Neutral theory states that evolution is completely due to positive/negative balance between mutation and genetic drift. Concretely, the evolutionary process is somehow divided in two independent steps: evolution of genetic structures is due to neutral mutations, that is, mutations which have no consequences on survival and reproduction of individuals, but only on evolution[6], and the way these mutations spread and distribute over time is regulated by genetic drift.

The most illustrative model in neutral theory is the *infinite-allele model*. It consists of a population of N diploid individuals, in which both genetic drift and mutations are acting. In order to introduce mutations in the simplest way, it is convenient to assume that the gene under study has an infinite number of possible alleles, the mutation rates between them being u.

We define $\mathcal{G}$ as the probability that two alleles picked at random are identical by state but different by origin. By definition, $\mathcal{G}$ is an inverse measure of the genetic variation in the population: when all the alleles are equal, $\mathcal{G} = 1$, and when all the alleles are different $\mathcal{G} = 0$. Starting from generation t, the probability of obtaining two alleles different by origin but identical by state at generation $t + 1$ is given by

$$\mathcal{G}_{t+1} = \left(\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}_t\right)(1-u)^2 . \tag{3.13}$$

The first term is the sum of two complementary mechanisms. In the first one, the same allele from generation t is copied twice in generation $t + 1$, which occurs with probability $1/2N$[7]. In the second one, two already equal by state but different by origin alleles from t are copied into $t + 1$. For this, we need to require two independent conditions: pick one allele and then pick a different one (probability $1 - 1/2N$), but such that these two were already equal by state (probability $\mathcal{G}_t$, by definition). The second term is the probability of non mutation, $(1 - u)^2$, necessary to keep unaltered the result of the choice over the population.

[6]For example, one could consider lethal mutations, which independently of their abundance are always decisive in survival.

[7]Note that the choice of the first allele is completely arbitrary and occurs with probability 1 (at least one allele is always picked). This wouldn't be the same if we were asking for the probability of picking specifically allele i (which is $1/2N$).

Since N is large and u is small, we can expand this expression both up ot order u^2 and $u/2N$ to obtain

$$\Delta\mathcal{G} \equiv \mathcal{G}_{t+1} - \mathcal{G}_t = \frac{1}{2N}\left(1 - \mathcal{G}_t\right) - 2u\mathcal{G}_t, \tag{3.14}$$

whose interpretation is clear: the variation in homozygosity is due to two terms, the first one, the drift term, which increases it, and the second one, the mutation term, that reduces it. Equilibrium is reached when $\Delta\mathcal{G} = 0$, leading to a steady-state solution

$$\hat{\mathcal{G}} = \frac{1}{1 + 4Nu}, \tag{3.15}$$

which predicts two asymptotic regimes for the distribution of polymorphism:

- $u \gg \frac{1}{2N}$: mutation is dominant, equilibrium is heterozygous, $\hat{\mathcal{G}} \to 0$.
- $u \ll \frac{1}{2N}$: drift is dominant, equilibrium is homozygous, $\hat{\mathcal{G}} \to 1$.

One of the most defining consequences of neutral theory is the prediction that the *rate of substitutions* is independent of the population size. We define the rate of substitution as the average time needed for a new mutant allele to fixate the population. By means of Equation (3.12), one solely allele introduced by a mutation will have, initially, $\pi = 1/2N$. On the other hand, by the probabilistic definition of the mutation rate, the mean number of mutations per generation will be $X = 2Nu$. Consequently, the rate of substitution, k, can be obtained as

$$k = \Pi\, X = \frac{2Nu}{2N} = u. \tag{3.16}$$

The immediate product of k being size-independent is that neutral evolution takes place at the same speed both for small and large populations. Result (3.16) is of great significance in the context of population genetics, since it acts as a consistency check for neutral theory in studied species. It might seem in disagreement with observation, but one has to note that time is expressed in terms of generations. If we consider physical time, the larger the population, the faster generations succeed, and the faster evolution takes place.

3.5 Natural selection

Neutral theory gives a theoretical satisfactory prescription on how populations evolve. Not only is it able to describe in a coherent way the interaction between two contrary forces, but also, the same way HW model did, it provides some predictions about polymorphism and rate of substitution that have been successfully verified in many experiments. Nevertheless, neutral theory alone fails to explain evolution of proteins and genes in which some other non-random phenomena are involved, namely selection. Thus, a third evolutionary force needs to be considered: *natural selection*.

3.5.1 The principles of modern natural selection

Natural selection was first introduced by Darwin in *The origin of species* (1859), and is built around the concepts of «individual» and «survival of the fittest». Although the core still remains the same, the theory has been improved over the last century, not only by making it more rigorous, but also by incorporating these concepts into models that study the change in allele frequency due to natural selection. Now, natural selection is a highly quantitative evolutionary theory, able to predict how natural selection operates on different components of fitness and how it is structured at different levels of population structure, not just the individuals.

The modern theory of natural selection can be summarised in three axioms [1]:

- In all species, more offspring are produced than can possibly survive and reproduce.
- Organisms differ in their ability to survive and reproduce, in part owning to differences in genotype.
- In every generation, genotypes that promote survival in the current environment are present in excess at the reproductive age and, thus, contribute disproportionately to the offspring of the next generation.

In other words, natural selection is based on the concept of adaptive evolution: evolution acts in such a way that species tend to *adaptation*, that is, genetic structure of populations is always improved. Different genotypes have different *fitness*, which means that not all genotypes are equally good for the survival of the species. Fitness is both a result of genetic composition and interaction with the *environment*. Thus, the fittest genotypes, i.e those that increase the survival and reproductive capability of the individuals, therefore populations, tend to increase their frequency from one generation to another.

Natural selection introduces two main modifications with respect to genetic drift and mutations: fitness and environment. The first one acts as a deterministic bias in genotype frequencies that contrasts with the so far encountered models. The second one introduces dependence on the phenotype at genotype level. In this context, selection is strongly affected by dominance, as we will explicitly show when we develop the model.

3.5.2 The fundamental model for natural selection

The model of natural selection we are going to construct is based on the Hardy-Weinberg model, but allowing the fitnesses of genotypes to differ. This underlines again the importance of Hardy-Weinberg model as an ideal basis for more complex evolutionary models. More complex models of selection can be discussed [1], but are out of the scope of this work.

Hence, consider a population of N diploid hermaphroditic (i.e, male and female allele frequencies coincide) individuals which mate randomly and in which generations do not overlap each other. Again, we concentrate our efforts in one locus, whose gene can adopt two different alleles, A_1 and A_2 with frequencies p and q. Then, the HW model predicts genotype frequencies $x_{11} = p^2,\ x_{12} = 2pq$ and $x_{22} = q^2$ for the newborn.

We define now the fitness ω_{ij} of the genotype A_iA_j. The easiest way to understand the role of fitness is in terms of *viability*, which is defined to be the probability that a genotype survives the maturation process (from fertilization to reproductive age). The actual fitness includes not only viability, but also fertility, mating success, developmental time[8] and so forth, but it can be understood as a survival probability of the associated genotype. We shall assume that environment does not change over time, so that neither does fitness.

Fitness is defined in such a way that natural selection acts only on the maturation part and is independent of the process that determines the genetic structure of the newborn. We can, accordingly, distinguish three steps: due to random mating, before selection, the newborn at generation will have HW genotype frequencies. Then selection occurs, and genotype frequencies are modified by fitnesses. After selection, by means of the probability of independent processes, the genotype frequencies at generation t are given by

$$x_{11} = p_t^2 \frac{\omega_{11}}{\langle\omega\rangle} \quad , \quad x_{12} = 2p_tq_t \frac{\omega_{12}}{\langle\omega\rangle} \quad , \quad x_{22} = q_t^2 \frac{\omega_{22}}{\langle\omega\rangle}, \tag{3.17}$$

[8]In general, fitness, which is the quantity that appears in natural selection model, is formed by multiple terms that co-influence each other.

where $\langle\omega\rangle$ is called the *mean fitness* and normalises the new frequencies to one:

$$p_t^2 \frac{\omega_{11}}{\langle\omega\rangle} + 2p_t q_t \frac{\omega_{12}}{\langle\omega\rangle} + q_t^2 \frac{\omega_{22}}{\langle\omega\rangle} = 1. \tag{3.18}$$

Note that HW model is equivalent to say $\omega_{ij} = \langle\omega\rangle$. Each generation can be built with this same procedure, which implicitly parts from the HW model, but since we have introduced different fitnesses, there is no longer a reason to assume allele frequencies at generation $t+1$ to be the same as in generation t. That is, in principle, differences in fitness have to be a source of variation from equilibrium. Using Equation (3.1) we obtain the allele frequencies at generation $t+1$

$$p_{t+1} = \frac{p_t^2\,\omega_{11} + p_t q_t\,\omega_{12}}{\langle\omega\rangle} \quad , \quad q_{t+1} = \frac{q_t^2\,\omega_{22} + p_t q_t\,\omega_{12}}{\langle\omega\rangle}. \tag{3.19}$$

Defining $\Delta_s p \equiv p_{t+1} - p_t$ as the variation in allele frequency due to selecion over one generation, Equation (3.19) reads

$$\Delta_s p = \frac{p_t q_t}{\langle\omega\rangle}\Big(p_t(\omega_{11} - \omega_{12}) + q_t(\omega_{12} - \omega_{22})\Big). \tag{3.20}$$

3.5.3 Relative fitness and the three kinds of selection

The fitnesses we have introduced so far are called *absolute fitnesses*, since they are related to one genotype and independent from each other. However, properties of natural selection are usually better characterised in terms of *relative fitnesses* with respect to one reference genotype, usually the first one.

We are going to adopt the following definition of relative fitnesses with respect to ω_{11}:

$$\begin{aligned} A_1A_1 &: \frac{\omega_{11}}{\omega_{11}} = 1, \\ A_1A_2 &: \frac{\omega_{12}}{\omega_{11}} = 1 - hs, \\ A_2A_2 &: \frac{\omega_{22}}{\omega_{11}} = 1 - s, \end{aligned} \tag{3.21}$$

where s is the *selection coefficient* and h is the *heterozygous effect*. Relative fitness exhibits dominance in the genotype distribution way more explicitly than absolute fitness. We can interpret s and h in this sense as follows:

- $\underline{s}$: the smaller s is, the closer A_2A_2 is to A_1A_1 in fitness and viceversa. The selection coefficient contains information about how much does the environment benefit A_1A_1 over A_2A_2. From now on, we shall assume $s > 0$, that is, genotype A_1A_1 is more fit than genotype A_2A_2[9]

- $\underline{h}$: for $h = 1$, the fitness of A_1A_2 is equal to the fitness of A_2A_2, while for $h = 0$ it is equal to A_1A_1. The heterozygous effect measures towards which homozygote does the phenotypic effect of the heterozygote tend, that is, the dominance between alleles A_1 and A_2. In particular, we have five ranges: $h < 0$ (overdominance), $h = 0$ (A_1 dominant), $0 < h < 1$ (incomplete dominance), $h = 1$ (A_2 dominant) and $h > 1$ (underdominance). The implications of each range of values will be discussed later on.

[9]Note that we are not losing generality. If A_2A_2 were fitter, we could simply rename the genotypes or change the reference one for the relative fitness. As a convention, the reference is taken for s to be positive.

Furthermore, in terms of s and h, Equation (3.20) turns into a more informative expression,

$$\Delta_s p = \frac{p_t q_t s}{\langle \omega \rangle}\Big(p_t h + q_t(1-h)\Big), \tag{3.22}$$

where now $\langle \omega \rangle = 1 - 2p_t q_t hs - q_t^2 s$ corresponds to the mean relative fitness. Equation (3.22) predicts three steady-state solutions (i.e $\Delta_s p = 0$ points) for natural selection:

$$\hat{p}_1 = 0 \quad , \quad \hat{p}_2 = 1 \quad , \quad \hat{p}_3 = \frac{h-1}{2h-1}. \tag{3.23}$$

Based on their linear stability analysis, these solutions draw a total of three different scenarios (recall that $s > 0$) or three different kinds of natural selection:

- $\underline{0 \leq h \leq 1}$: corresponds to $\omega_{11} \geq \omega_{12} \geq \omega_{22}$. Solution $\hat{p}_1$ is unstable, $\hat{p}_2$ is stable, and $\hat{p}_3$ does not exist ($\hat{p}_3 \notin [0,1]$). When dominance is incomplete and $s > 0$, A_1A_1 genotypes are always favoured by he environment and population tends to fixation of A_1. If we were to choose $s < 0$, fixation would occur for A_2. Often called *directional selection*, it was the original selection predicted by Darwin.

- $\underline{h < 0}$: corresponds to $\omega_{12} \geq \omega_{11} \geq \omega_{22}$. Solutions $\hat{p}_1$ and $\hat{p}_2$ are unstable, $\hat{p}_3$ is stable. A_1A_1 is still favoured over A_2A_2, but overdominance implies that the most stable configuration is the heterozygous one. Fixation is never reached, and the greater overdominance is, the more heterozygous population is. Selection under overdominance is referred to as *balancing selection.*

- $\underline{h > 1}$: corresponds to $\omega_{11} \geq \omega_{22} \geq \omega_{12}$. Now, $\hat{p}_1$ and $\hat{p}_2$ are stable, $\hat{p}_3$ is unstable. Underdominance represents a situation where heterozygous configuration is highly unstable. Contrary to incomplete dominance, since now A_2A_2 is fitter than A_1A_2, fixation is reached for both homozygotes. In other words, the initial condition gets away from the heterozygous configuration towards the closest homozygous configuration. Selection in this case is named *disruptive selection*, but is very unlikely to happen in nature.

The numerical solutions for Equation (3.22), along with the three situations recently discussed is illustrated in Figure3.

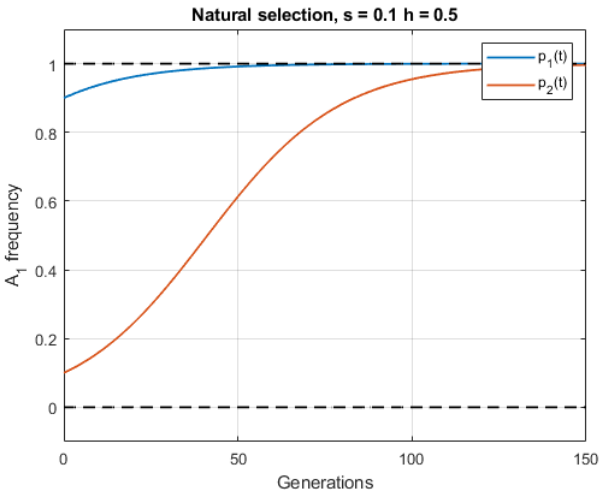
3.5.4 Fisher's Fundamental Theorem of Natural Selection

For the following discussion, the reader should refer to [3] and [4].

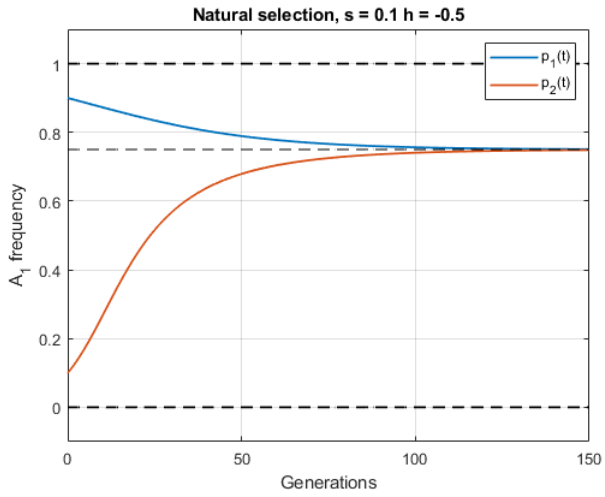
It is known by Darwin's pioneer observations over species that natural selection is *directional.* Natural selection tends to preserve those genes that enhance survival of the population, that is, the fittest ones, in the process known as adaptation. Thus, populations evolve towards fitter configurations and it is reasonable to think that adaptation must be expressed through an average increase in fitness. From Equation (3.22), we have predicted three types of selection that reflect an adaptive behaviour, but the Equation itself does not manifest explicitly any overall sign of adaptation. Thus, we are left with the question: how does fitness increase?

We shall start out discussion by using the expression of the mean relative fitness to rewrite Equation (3.22) in a more convenient way:

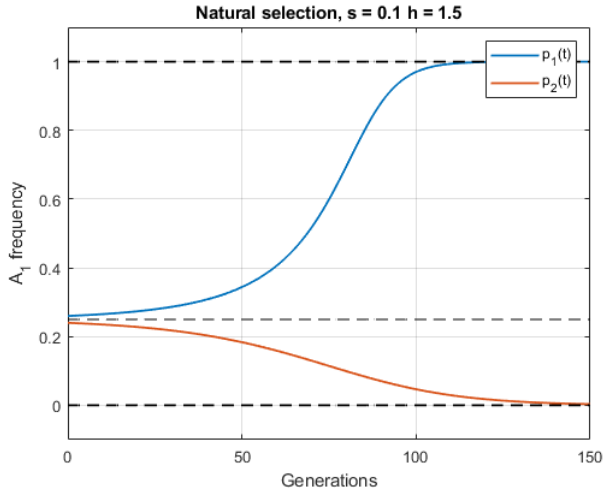
$$\Delta_s p = \frac{pq}{2\langle \omega \rangle}\frac{d\langle \omega \rangle}{dp}. \tag{3.24}$$



(a) Incomplete dominance.



(b) Overdominance.



(c) Underdominance.

Figure 3: Numerical solutions for evolution under natural selection. Each plot corresponds to one of the three characteristic situations found by linear stability analysis. The coloured lines represent the time evolution of A_1 frequency for two different initial conditions. The black lines represent the steady-state solutions given by Equation (3.23).

Now, applying the chain rule $\Delta\omega = (d\langle\omega\rangle/dp)\ \Delta_s p$, Equation (3.24) reads

$$\Delta\langle\omega\rangle = \frac{pq}{2\langle\omega\rangle}\left(\frac{d\langle\omega\rangle}{dp}\right)^2. \tag{3.25}$$

As it can be seen, no matter how the allele frequency varies, mean fitness is always an increasing function of time. It can be shown that the right-side term of Equation (3.25) coincides with the variance in fitness:

$$\Delta\langle\omega\rangle = 2\frac{\mathrm{Var}(\omega)}{\langle\omega\rangle}, \tag{3.26}$$

thus generalising the result to an unspecified number of alleles and genotype fitnesses. The derivation can be seen in Appendix A.

In this more general expression, result (3.26) synthesises *Fisher's fundamental theorem of natural selection*. In Fisher's own words [3][4]: «the rate of increase in fitness of any organism at any time is equal to its genetic variance at that time». In other words, this is exactly the

mathematical formulation of adaptation: as variances are always positive, under natural selection species not only become better adapted to the environment, but the speed of adaptation increases proportionally to how adapted they are.

3.6 Derivation of the Kimura-Ohta Equation

Fisher's fundamental theorem seems to be a definitive achievement in the context of evolutionary dynamics. When selection is not involved, neutral theory generates accurate predictions on how genetic structure of populations changes with time; and when selection is involved, neutral theory is corrected and Fisher's theorem provides an statement of adaptation. Thus, taking all into consideration, it looks like we have derived a successfully complete picture of the evolutionary process. Haven't we?

(Un)fortunately, the answer is no. Fisher's fundamental theorem is undeniably true for simple selection at a single locus, but losses validity when less restrictive assumptions are made over the population. To be more specific, average fitness can decrease when population size is variable, forces other than selection are considered, fitnesses are frequency or time dependent or genetic interaction between multiple loci (epistasis) is not neglected [2][4]. One would say that Fisher's fundamental theorem is not fundamental nor even a theorem. Thus, we are left with two options: either evolution is not really adaptive or fitness is not a good measure of adaptation. Surprisingly, both options are partially true. While forces such as mutations and drift are non-adaptive mechanisms of evolution, evolution should still be adaptive in every process governed by natural selection. As we shall see, the problem is not on Fisher's theorem itself, but on the non-validity of the underlying evolutionary and statistical theories there used when a more complex scenario is described. Hence, an extended version of Fisher's result for a quantity slightly different than fitness should naturally emerge from a more general approach. In that sense, Fisher's theorem is a first, restricted attempt to a mathematical description of adaptation, and it leads the way on how to proceed to derive a result valid under genetic drift, mutations and time-dependent selection.

The first step in generalising our results is obtaining a unique Equation that describes mutations, genetic drift and selection altogether. This is the *Kimura-Ohta Equation* (K-O), which constitutes a diffusion Equation for the probability distribution of allele frequencies in the continuous limit.

3.6.1 The Moran model for random genetic drift

Moran model is a simple model for random genetic drift slightly different from Wright-Fisher model. The motivation for this change will be clear later on, when we see how this slight difference results in a simpler generalisation to continuous. Moran model repeats all the assumptions of the W-F model but two: the population is haploid and it is an overlapping generations model. While W-F model states that all individuals die and are replaced every generation, Moran model considers that only one individual is replaced at each step of time, so that individuals born in different generations coincide in time. Therefore, the concept of generation is loosely defined, and the maximum we can say is that, on average, a population takes N steps to be completely replaced.

Technically, these condition means the difference between Moran and W-F models is introduced not introduced in the probabilistic description, but on the accessible transitions between time steps. In other words, Equation (3.10) from W-F model holds for Moran model,

$$\mathcal{P}(p, t+1) = \sum_{p'} T_{p,p'} \mathcal{P}(p', t), \tag{3.27}$$

but transition probabilities, $T_{p'p}$, have different values:

$$T_{p,p'} = \begin{cases} (1-p')\,p' & \text{if } p = p' + \frac{1}{N} \\ (p')^2 + (1-p')^2 & \text{if } p = p' \\ p'\,(1-p') & \text{if } p = p' - \frac{1}{N} \\ 0 & \text{otherwise,} \end{cases} \tag{3.28}$$

since they respond to a different description of drift at microscopic level (gene level). In this context, the interpretation of $T_{p,p'}$ as transition probabilities is more evident. At each time step only one allele is replaced, so only transitions from $n' \sim p'$ to $n \sim p = (n'-1, n', n'+1) \sim (p' - 1/N, p', p' + 1/N)$ are allowed. To illustrate how probabilities are obtained, consider only the transition, n' to $n'+1$. This transition corresponds to the death of one A_2 and birth of one A_1. Assuming independence in both phases, that happens with probability $Prob = Prob(\text{picking } A_2) \times Prob(\text{picking } A_1) = (1-p')\,p'$.

In order to introduce the continuous limit, also called *diffusion approximation*, it is convenient to change to a generations τ scale as time scale. Since each generation takes N steps, $\tau = t/N$. We shall also change notation from $p \to x \in \mathbb{R}$, where x is a continuous variable. Taking $N \to \infty$, we can define $\Delta x = \frac{1}{N} = dx$ and $\Delta\tau = \frac{1}{N} = d\tau$. With these prescriptions the diffusion Equation for Equation (3.27) reads

$$\frac{\partial \mathcal{P}}{\partial \tau} = \frac{1}{N}\frac{\partial^2}{\partial x^2}\Big(x(1-x)\mathcal{P}\Big). \tag{3.29}$$

We refer the reader to Appendix B for a more detailed calculation.

3.6.2 The continuous limit of mutations and natural selection

The generalisation of mutations and selection to a continuum limit is immediate from Equations (3.7) and (3.20). Note that both theories are already expressed in terms of generations, so we do not need to change the time scale as done for the Moran model, but only expand in series $\Delta p \approx dp/d\tau$ and replace p by x. Mutations read

$$\frac{dx}{d\tau} = -x\mu + (1-x)\nu, \tag{3.30}$$

whereas selection

$$\frac{dx}{d\tau} = x(1-x)\left(f_1(x,\tau) - f_2(x,\tau)\right). \tag{3.31}$$

The quantities $f_i(x,\tau)$ are defined as the fitness of the allele A_i[10], which coincide with the genotype ones for haploid organisms. It can be easily shown that Equation (3.31) is generally valid both for diploid and haploid organisms, but the values of the allele fitness will be necessarily different, and so the explicit expression of the change in allele frequency. In other words, one needs to be careful here: since we are now dealing with haploid organisms, Equation (3.22) will not be valid anymore.

3.6.3 The Kimura-Ohta Equation

The final step is «adding up» Equations (3.29), (3.30) and (3.31). However, while mutation and selection are modelled in terms of allele frequency evolution Equations, genetic drift is in

[10]Note that according to the definition given in Appendix A the fitness of the allele is strictly talking ω_i, rather than f_i, which might be better interpreted as a normalised fitness. In the following we will just refer to it as allele fitness.

terms of their probability distribution. In order to combine the three Equations, one needs to express Equation (3.29) as an *stochastic differential Equation*. Operating the same way for the deterministic Equations (3.30) and (3.31), the full process is described by

$$dX_\tau = \big(m(X_\tau) + g(X_\tau)\, s(X_\tau, \tau)\big)d\tau + \sqrt{\frac{2}{N} g(X_\tau)}\, dW_\tau, \tag{3.32}$$

where have been introduced the quantities

- *response coefficient*: $g(x) \equiv x(1-x)$,
- *mutation coefficient*: $m(x) \equiv -x\mu + (1-x)\nu$,
- *selection coefficient* of A_1 over A_2: $s \equiv f_1(x,\tau) - f_2(x,\tau)$.

The way this is done is out of the scope of this work. For the moment, it is only important to notice that mutations and selection are described as deterministic processes, while genetic drift contains a noise term, dW_τ. This is in agreement with our expectations, since genetic drift has been studied as a purely random process and, thus, must have an stochastic nature.

Inverting the stochastic Equation (3.32), we obtain the normal form of the K-O Equation:

$$\frac{\partial \mathcal{P}(x,\tau)}{\partial \tau} = -\frac{\partial}{\partial x}\Big[\big(m(x) + g(x)\, s(x,\tau)\big)\mathcal{P}(x,\tau)\Big] + \frac{1}{N}\frac{\partial^2}{\partial x^2}\Big(g(x)\mathcal{P}(x,\tau)\Big). \tag{3.33}$$

The K-O Equation provides a general and accurate model to describe the collective effect of the three studied evolutionary forces over a population. We won't show it here, but its generalisation to k alleles is also immediate. As such, the K-O Equation should be regarded as one of the most important milestones in population genetics.

4 Stochastic thermodynamics

In the models developed so far, we have been dealing with very large allele populations. Since the typical size of a chromosome is of the order of μm, allele populations are clearly microscopic systems. For microscopic systems with a lot of components, giving an exact prescription of the dynamical evolution of every component is highly inefficient, and a statistical approach is often required. Even if simpler, statistical descriptions also introduce fluctuations that can affect significantly the system, displacing it from equilibrium. One of the main and more difficult to deal with sources of fluctuations is randomness. Systems with one or more parts of them subjected to random effects are called *stochastic systems*, and the branch of physics that deals with its interactions with the environment is *stochastic thermodynamics*.

Genetic drift is an evolutionary force entirely due to chance: the finite size of population leads to changes in allele frequency when individuals are replaced into new generations. As a consequence, K-O Equation describes a microscopic system which is stochastic by nature and generally out of equilibrium, for being the combination of three evolutionary forces equilibrium is rarely the observed behaviour.

Stochastic thermodynamics seems to be the statistical theory that better defines the type of evolution here described. Thus, if one wants to succeed in a proper characterisation of adaptation, methods for the study of stochasticty need to be introduced. A detailed description of stochastic thermodynamics is out of the scope of this work, so we will rather adjust to introduce the necessary elements for our goals. Even if right now the application of introducing stochastic thermodynamics might not seem evident, it will be clarified by the end of this work.

4.1 Basic concepts of stochastic systems

Let us start giving some more precise definitions [5].

«Stochastic» is vaguely used as a synonym of «random». More precisely, an *stochastic variable* X is a random variable defined by a set of possible values and a probability distribution over this set. For example, the stochastic variable in genetic drift is allele frequency.

An *stochastic process* is a random function evolving in time. That is, is a function of two variables, one of which is an stochastic variable and the other one of which is time.

Stochastic processes appear naturally in physical systems with quantities that vary in irregular and complicated ways over time. When such a fluctuating behaviour is exhibited, the time choice to take average values is ill-defined and it is often convenient to replace the single irregular function by an ensemble of well-defined functions (chosen depending on the process). We say, then, that time evolution of that quantity has been turned into a stochastic process.

Stochastic thermodynamics is the branch that studies stochastic processes in physical systems. Stochastic processes involve strong fluctuations over the average and equilibrium configurations, which manifest themselves as energy exchanges with the environment (also named *bath*) and define quantities analogue to those of thermodynamics. In some specific cases, these quantities coincide with the conjugate thermodynamic variable. Thus, even if its name is confusing, stochastic thermodynamics is essentially the study of these fluctuations.

Stochastic thermodynamics does not apply to microscopic systems, but more specifically to *small systems*. A system is considered small when both the system is of microscopical size and subject to significant fluctuations from the average behaviour. Consider the example of genetic drift alone. Observed over just a few generations, nothing indicates tendency towards fixation: allele frequency varies randomly around some average value. Observed over $t \gg N$ generations, equilibrium is eventually or almost reached and fluctuations are null or small enough to be ignored, so that the system is no longer small.

Let's finally introduce the concept of non-equilibrium. If we assume no external action over the system, the internal energy has to be conserved and fluctuations need to compensate in every direction. Hence, no fluxes of energy between the system and the environment can be, on average, observed and eventually the system relaxes to *equilibrium*. Conversely, when net exchange of energy can be observed the system is said to be *out of equilibrium*. In that case, the environment is crucial to understand the energy dissipation and a external *control parameter*, λ, that monitors the state of the environment has to be introduced. Its time evolution, $\lambda(t)$, is called *protocol*.

4.2 The fluctuation theorems

The previous definitions confirm the original affirmation that K-O Equation describes in general a small system which is out of equilibrium, motivating the introduction of stochastic methods. Evolution of such systems is in general very difficult to predict, but under certain conditions some precise results about energy exchange can be obtained. These results are known as *fluctuation theorems*.

4.2.1 Stochastic description of small systems out of equilibrium

Consider a stochastic system in contact with a thermal bath and evolving in time. We assume that time varies as a discrete variable, i.e, $t_k = k\Delta t$, with $k = 0, 1, ..., n$ (nothing far from reality when considering generations). The state of the system at time t_k is identified by the

generic random variable C_k (for instance, the allele frequency). We define a *trajectory* of the system, Γ, from t_0 to t_n, as a concrete succession of values

$$\Gamma = (C_0, ..., C_n) \tag{4.1}$$

that bring the system from C_0 at t_0 to C_n at t_n, so that the evolution of the system between a initial-final state is determined by its possible trajectories. Seemingly, the state of the bath at time t_k is identified by the control parameter λ_k, and its evolution between t_0 and t_n determined by the protocol $(\lambda_0, ..., \lambda_n)$.

Since multiple trajectories with the same extreme points are in general possible, the system can take different values of C[11] at each step of time. We need to define, therefore, the probability that the system, at t_k, has value C: $\mathcal{P}_{\lambda_k}(C)$. For the same reason, we introduce the probability of transition from C' to C at time t_k, $W_{\lambda_k}(C' \to C)$. Both probabilities are, naturally, normalised[12]. In general, $\mathcal{P}_{\lambda_k}(C)$ will combine $W_{\lambda_k}(C' \to C)$ for different starting values C' and different times t_k.

The most important case of stochastic process in physics and biology is a *Markov chain*. A Markov chain or Markov process is such that the state of the system at time t_{k+1} is completely determined by the state of the system at t_k and unaffected by previous times. In other words, the system has no memory of its evolution. In a formal definition, the probability of finding C at time t_{k+1} is only given by the probabilities of the possible C', all at same t_k,

$$\mathcal{P}_{\lambda_{k+1}}(C) = \sum_{C'} W_{\lambda_k}(C' \to C)\, \mathcal{P}_{\lambda_k}(C'). \tag{4.2}$$

Comparing with Equation (3.27) we notice that genetic drift is modelled as a Markov chain[13] and, by extension, the K-O model. Thus, in the following, we shall restrict ourselves to Markov chains, rather than entering a more general (but not useful) treatment.

4.2.2 Entropy production and energy dissipation

When dealing with stochastic processes it is also crucial to assume that **always** exists at least one equilibrium probability distribution, $P^{eq}_{\lambda_k}(C)$. This equilibrium distribution naturally satisfies $P^{eq}_{\lambda_{k+1}}(C) = P^{eq}_{\lambda_k}(C)$.

A much stronger assumption in addition to equilibrium is that there exists also a so-called *local detailed balance*,

$$\frac{W_{\lambda_k}(C' \to C)}{W_{\lambda_k}(C \to C')} = \frac{\mathcal{P}^{eq}_{\lambda_k}(C)}{\mathcal{P}^{eq}_{\lambda_k}(C')}, \tag{4.3}$$

which is essentially equivalent to say that all transitions are reversible, as expected for thermodynamic equilibrium. Note that the existence of an equilirbium distribution is a necessary but not sufficient condition for local detailed balance.

Given a distribution and its equilibrium associated distribution, it is possible to define the *total dissipation* of a generic trajectory Γ as

$$S(\Gamma) = \sum_{k=0}^{n-1} \log\left(\frac{P^{eq}_{\lambda_k}(C_{k+1})}{P^{eq}_{\lambda_k}(C_k)}\right) - \log\left(\frac{P_{\lambda_n}(C_n)}{P_{\lambda_0}(C_0)}\right), \tag{4.4}$$

[11] Note that here C is the generic variable, not the values along the path Γ.

[12] $\sum_C \mathcal{P}_{\lambda_k}(C) = 1$ and $\sum_{C'} W_{\lambda_k}(C' \to C) = 1$

[13] Just think on how evolution proceeds: every generation is formed depending only on the previous generation. The system's past is not important when moving towards t+1.

where the first term corresponds to the *entropy production*

$$S_p(\Gamma) = \sum_{k=0}^{n-1} \log\left(\frac{P^{eq}_{\lambda_k}(C_{k+1})}{P^{eq}_{\lambda_k}(C_k)}\right), \tag{4.5}$$

and the second term is called *boundary term*. Note that if equilibrium is satisfied, $S = 0$.

In particular, the entropy production measures the irreversibility of a trajectory. To see that, we can define the reverse path of Γ under the protocol λ, that is $\Gamma_R = (C_n, ..., C_0)$, which evolves under the reverse protocol $\lambda^R = (\lambda_n, ..., \lambda_0)$. The forward path has an overall probability of being accomplished $P_F(\Gamma)$, while the reverse path happens with probability $P_R(\Gamma_R)$. By virtue of Equation (4.3) (see Appendix C for the detailed calculation), when the system is at equilibrium, the following equality is satisfied:

$$\frac{P^{eq}_F(\Gamma)}{P^{eq}_R(\Gamma_R)} = e^{S_p(\Gamma)}, \tag{4.6}$$

which means the greater the entropy production is, the more unlikely the reverse path is. In particular, Equation (4.6) tells that the entropy production is zero for a reversible system in equilibrium, as expected from the thermodynamic interpretation, but in general equilibrium does not imply reversibility. Although this seems to contradict the detailed balance condition, one should notice that S_p is not the total entropy, but that of the system's production alone, which is not necessarily zero even if equilibrium is realised and suggests that a more general expression, containing the total entropy, is needed.

Furthermore, this result is naturally not true anymore for systems out of equilibrium. In such systems, definition (4.4) for the total dissipation provides $S(\Gamma_R) = -S(\Gamma)$, and Equation (4.6) can be generalised to

$$\frac{P_F(S)}{P_R(-S)} = e^S. \tag{4.7}$$

4.2.3 The Jarzynski equality

From Equation (4.7) and using the fact that probabilities $P_F(S)$and $P_R(-S)$ must be normalised with respect to integration in dS, it is immediate to obtain the *Jarzynski equality*:

$$\langle e^{-S} \rangle = 1, \tag{4.8}$$

where, as usual, $\langle f \rangle$ stands for average over the $P_F(S)$ distribution. Finally, due to Jensen's inequality[14], we obtain

$$\langle S \rangle \geq 0. \tag{4.9}$$

The interpretation is immediate. First, S is a quantity always increasing in average, but not necessarily in specific processes. Second, when the system is at equilibrium, $S = 0$ by definition, which implies that the forward and reversed trajectories are equiprobable, or, in other words, the process is reversible. Now equilibrium and reversibility are equivalent, respecting the detailed balance condition and solving the apparent problem of Equation (4.6). Hence, Equation (4.9) represents a generalised second law of thermodynamics for small systems out of equilibrium, where S plays the role of the total entropy or entropy of the universe.

This deduction can be checked for well-defined statistical systems with known probability distributions that allow to obtain concrete forms of S. The reader can refer to [8] for the example of the canonical ensemble, where the S obtained coincides exactly with the thermodynamic definition of entropy.

[14] Jensen's inequality: $\langle F[x] \rangle \leq F[\langle x \rangle]$, where averages are taken with respect to a certain distribution and F is a convex function.

5 Fitness flux theory

This entire chapter reproduces the results of [6].

At the beginning of Sec. 4 we said that it was not obvious whether it was necessary or even interesting to introduce stochastic thermodynamics, or what was the connection with the derivations from Sec 3. Note that Equation (4.9) provides a quantity that is, on average, always increasing. Moreover, it is a quantity that depends on the probability distribution, so it is representative of the system we are studying. When discussing Fisher's fundamental theorem, we have stressed the fact that the problem wasn't on the theorem itself, but on the features of the underlying theory used. Averages over statistical systems are determined both by the statistical theory describing the system and its probabilistic dynamics. Whereas selection is modelled as deterministic in population genetics, genetic drift is an stochastic process. Hence, if the same average quantity is wanted for evolution under genetic drift, mutations and selection, the statistical prescription originally used by Fisher needs to be replaced by an stochastic approach based on the K-O Equation.

This extension of Fisher's fundamental theorem to a case in which mutations, genetic drift and time-dependent selection are considered together is called fitness flux theorem. Following Fisher's intuition, the goal is to obtain a fitness-like quantity which is, on average, always increasing and, therefore, accounts for adaptation in this more general scenario of the evolutionary dynamics.

5.1 Fitness flux and fitness sea- and landscapes

Prior the proceed, it is important to stablish some definitions and conventions.

The genotype evolution can be described by a series of observations of allele frequencies $\Gamma = (x_0, ..., x_n)$ at successive times $(t_0, ..., t_n)$, called history of the population. We have intentionally used the notation Γ to identify these observations with the trajectory our system follows and time evolution with our protocol. The *fitness flux* can then be defined as $\phi(x_i, t_i) = s(x_i, t_i)\Delta x_i/\Delta t_i$, where $s(x, t)$ are the time-dependent natural selection coefficients of the allele. Fitness flux represents a measure of adaptation at a given point of the population history. Similarly, the *cumulative fitness flux*

$$\Phi(\Gamma) = \sum_{i=0}^{n} \phi(x_i, t_i)\Delta t_i = \sum_{i=0}^{n} s(x_i, t_i)\Delta x_i, \tag{5.1}$$

is a measure of the total adaptation over the global population history. To see this, one can imagine a substitution process under constant selection, $s(x, t) = s$, in which the allele considered evolves from $x_0 = 0$ to $x_n = 1$. In that case, $\phi = s$, so the cumulative fitness flux just coincides with the selection coefficient.

Physics tells us that studying the properties of a system is sometimes easier in terms of some potential rather than in terms of its associated fields. Following that philosophy, it is useful to introduce what are called *fitness seascapes*, $F(x, t)$. A fitness seascape is essentially a potential whose associated field is the natural selection coefficient, $s(x, t) = \nabla F(x, t)$. Since the selection coefficient defines differences in fitness between alleles or genotypes, fitness seascapes are a continuous representation in time and frequency of the fitness among a population. We can then distinguish three cases:

- $s(x, t) = s = \text{cte} \implies \Phi(\Gamma) = s(x_n - x_0)$.
- $s(x, t) = s(x) \implies \Phi(\Gamma) = F(x_n) - F(x_0) = \Delta F$.

- $s(x,t) = s(x,t) \Longrightarrow \Phi(\Gamma) = F(x_n,t_n) - F(x_0,t_0) + G(\Gamma) = \Delta F + G(\Gamma)$.

In the first two F is called a landscape rather than a seascape. When F is a seascape, the cumulative flux is not conservative because Equation (5.1) does not include a sum over Δt which is necessary to obtain the ΔF term and generates an extra term $G(\Gamma)$.

What's more important: these last example shows, in addition, that fitness differences between populations at different times are in general ill-defined. It is not difficult to see: since fitness seascape depends both on x and t, we could imagine a situation in which $F(x_0,t_0) > F(x_n,t_0)$ but $F(x_0,t_n) < F(x_n,t_n)$, that is, the original state is fitter than the final one at the initial time, but less fit at the final time. Thus, fitness differences among the same population seem to be ambiguously defined and, in general, there is no reason to believe the opposite. This is nothing but a mental experiment which shows the non-validity of Fisher's theorem. On the other side, fitness flux, being only a function of x, is well defined at every step of the trajectory, hence measurable and possible to infer from genetic data. This gives strong evidence in favour of the fitness flux, rather than the fitness, as the appropriate quantity to define a generally valid version of Fisher's theorem. We only need to answer: when does Φ increase?

5.2 The fitness flux theorem

For extended calculations on this section the reader should refer to Appendix D.

Our objective is to apply the generalised second law of thermodynamics (4.9) to the Kimura-Ohta Equation (3.33) to obtain a quantity that, on average, is always increasing. Later on, we will discuss whether this quantity is physical or not and a good substitute for the fitness in Fisher's original theorem when drift, mutations and time-dependent selection act together[15].

The total entropy is given by Equation (4.4), for which one needs the equilibrium distribution of the K-O equation. By inspecting the K-O equation (5.2), one notices that the only way a time-independent solution is possible is by imposing also time-independent selection. This condition is therefore necessary in order to compute the calculations for the entropy, but, as we shall see, the final result will not depend on whether the selection is time-independent or not, and the assumption will be dropped again. The equilibrium distribution is given by

$$\mathcal{P}^{eq}(x) = \mathcal{P}_0(x)\, e^{NF(x)}. \tag{5.2}$$

Here, $F(x)$ is the fitness landscape (time-independent selection) and $\mathcal{P}_0(x)$ is the equilibrium distribution for the K-O Equation when selection is excluded, i.e, the neutral equilibrium distribution. Using this expression, the entropy reads

$$S(\Gamma) = N\Phi(\Gamma) - \big(\mathcal{H}(x_n,t_n) - \mathcal{H}(x_0,t_0)\big) = N\Phi(\Gamma) - \Delta\mathcal{H}(\Gamma). \tag{5.3}$$

Finally, we can apply Jarzynski's equality (4.8) to get

$$\langle\, e^{-N\Phi+\Delta\mathcal{H}} \,\rangle = 1, \tag{5.4}$$

where $\mathcal{H}(x,t) = \log\left(\mathcal{P}(x,t)/\mathcal{P}_0(x)\right)$ is the relative log likelihood at a given point. Note that this expression is independent of the seascape chosen for the derivation. Moreover, the result depends only on the probability distributions for neutral equilibrium and K-O out of equilibrium, but not on the equilibrium distribution used during the derivation. Hence, result (5.4) will still be valid for arbitrary time-dependent seascapes, for which no proper equilibrium distribution for the K-O Equation can be found and the computations cannot be done according

[15]For example, if S where to coindide with $F(x,t)$, the result would not be interesting, because as we have seen the fitness seascape is not well defined nor measurable.

to the prescription introduced in Section 4. This is the general expression of the fitness flux theorem: any evolutionary process with mutations, drift, and selection given by an arbitrary time-dependent seascape must satisfy identity (5.4). Using now Equation (4.9) one gets:

$$\langle \Phi \rangle \geq \Delta H \Longrightarrow \frac{d}{dt}\langle \Phi \rangle \geq 0. \tag{5.5}$$

where $\Delta H = \langle \Delta \mathcal{H} \rangle$ stands for the relative entropy difference between the initial and final frequency configurations.

Equation (5.4) shows that adaptation can be overcome by genetic drift and $\Phi < \Delta \mathcal{H}$ only in a exponentially small subset of populations. In general, $\Phi > \Delta \mathcal{H}$ and, therefore, $d\Phi/dt > 0$. Hence, increase of (cumulative) fitness flux is an almost universal result.

The left side of Equation (5.5) indicates that $\langle \Phi \rangle$ is a bounded quantity. Moreover, although negative contributions ($\Delta H < 0$) are not excluded, they are limited to rather unlikely processes which occur in very short time intervals and cannot take place continuously in time without threatening the population, so average fitness flux will be typically not only lower bounded but also a positive quantity.

On the other hand, the right side of Equation (5.5) implies that, on average, fitness flux is an always increasing quantity, with independence of the path followed by the evolutionary process. In other words, evolution always takes place in a way such that the mean cumulative fitness flux increases, as well as it happened with the mean fitness for natural selection alone. Note also that the mean fitness flux, unlike the fitness seascape or fitness itself, is a physical quantity, for it is well-defined for different times among the same population. Thus, Equation (5.5) is essentially the statement of adaptation we were looking for in this more global, complex evolutionary context. To understand the significance of this result, note that all the derivation has been done following the assumption that, even with non-adaptive forces involved in the dynamics, evolution should still be adaptive for selection governing the process. Hence, not only have we proved the adaptive nature of this kind of evolution, but we have also provided a quantity able to account for it.

Fitness flux theorem is true for generic systems away from equilibrium and it can be shown that holds in analogous versions for variable populations, for diploid populations with multiple k alleles per gene (recall the K-O was obtained for haploid bi-allelic organisms) or when the implicit assumption of the existence of neutral equilibrium in (5.2) is dropped. In addition, the same concept can be developed for complex phenotype evolution, and not just for genotype ones as we have shown here. Thus, slightly modified versions of the fitness flux theorem can immediately be found for more realistic populations and more complex evolutionary dynamics.

Finally, the fitness flux theorem contains Fisher's fundamental theorem. As it is shown in Appendix D.2 for the case of two alleles, when selection is constant, i.e, time and frequency independent, and selection pressure is not only dominant but strong enough for the other evolutionary forces to be neglected (strong selection limit), fitness flux theorem reduces to Fisher's fundamental theorem, which is in agreement with our expectations.

6 Conclusions

In this work we have presented the basic elements of the theory of evolution based on population genetics and we have defined the cumulative fitness flux Φ as a measure of adaptation in evolution under mutations, genetic drift and time-dependent selection.

We have started by developing simple theoretical models (space-independence, no epistasis, two allele genes) for three of the most important evolutionary forces: mutations, genetic drift and natural selection, thus analysing the way they act on populations through variations in

allele frequency over generations. As a result, we have derived Fisher's fundamental theorem of natural selection, discussing the role of fitness as a gauge to account for the adaptive nature of evolution under natural selection.

Fisher's theorem has proven to be non-valid within forces other than selection, hence fitness has been discarded as a universal parameter to describe adaptation. Our goal has been then to reach a similar result for a generalised fitness-based quantity. For that, we have integrated all three forces into a global simple model, the Kimura-Ohta Equation, thanks to which we have been able to evidence the stochastic nature of such evolutionary process. Subsequently, we have introduced some necessary notions on stochastic thermodynamics, like the generalised definition of entropy and Jarzynski's inequality.

Ultimately, we have defined the (cumulative) fitness flux in terms of the selection coefficients and the history of the population and we have shown that, on average, it is an always increasing quantity.

In conclusion, the fitness flux theorem provides an accurate picture of adaptation in evolution under natural selection, mutations and genetic drift, for variable populations and out-of-equilibrium configurations, which remains valid when further implicit assumptions are dropped or more complex dynamics are considered. Moreover, the fitness flux theorem proves to be a generalisation of Fisher's fundamental theorem, to which it reduces in the limit of strong time-independent selection. Thus, cumulative fitness flux based on population genetics is a well-defined extension of fitness to evolutionary processes including forces other than natural selection and, as such, must be considered a much more fundamental measure of adaptation.

References

[1] Daniel L Hartl and Andrew G Clark. *Principles of population genetics*, volume 116. Sinauer associates Sunderland, 1997.

[2] John H Gillespie. *Population genetics: a concise guide*. JHU press, 2004.

[3] James Franklin Crow and Motoo Kimura. *An introduction to population genetics theory*. The Blackburn Press, 1970.

[4] Steven A Frank and Montgomery Slatkin. Fisher's fundamental theorem of natural selection. *Trends in Ecology & Evolution*, 7(3):92–95, 1992.

[5] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[6] Ville Mustonen and Michael Lässig. Fitness flux and ubiquity of adaptive evolution. *Proceedings of the National Academy of Sciences*, 107(9):4248–4253, 2010.

[7] David T Suzuki, Anthony JF Griffiths, Jeffrey H Miller, Richard C Lewontin, et al. *An introduction to genetic analysis*. WH Freeman and Company, 1986.

[8] Gavin E Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, 90(5):1481–1487, 1998.

A Proof of Fisher's Fundamental Theorem of Natural Selection

It is convenient to start by introducing a more general approach to the concept of mean fitness in a diploid organism. In general, given a set of genotype (relative) fitnesses, ω_{ij}, and allele frequencies, p_i, one can define the set of allele fitnesses as

$$\omega_i = \sum_j p_j \, \omega_{ij}. \tag{A.1}$$

The mean fitness of the population is then defined using these allele fitnesses as

$$\langle \omega \rangle = \sum_i p_i \, \omega_i = \sum_{ij} p_i p_j \, \omega_{ij}. \tag{A.2}$$

For any statistical variable X with an associated set of values x_i one can define its variance as a measure of their distance with respect to the mean behaviour of the variable. If we take the allele fitness ω to be our variable, then

$$\mathrm{Var}(\omega) = \sum_i p_i \, (\omega_i - \langle \omega \rangle)^2. \tag{A.3}$$

Although it can be proven more generally, let's restrict in the following to a gene with two allele as we have been doing all over the main text, but keeping, for simplicity, the general notation here introduced. To avoid any suspicion that both prescriptions might not be equivalent, one can check that with the original notation from the main text: $\omega_{ij} = \{\omega_{11}, \omega_{12}, \omega_{22}\} = \{1, 1 - hs, 1 - s\}$, Equation (A.2) provides $\langle \omega \rangle = 1 - 2pqhs - q^2$.

For two alleles with frequencies $\{p_1, p_2\} = \{p, q\}$, one easily obtains

$$\mathrm{Var}(\omega) = pq \, (\omega_1 - \omega_2)^2 \tag{A.4}$$

by using (A.2): $\langle \omega \rangle = p\omega_1 + q\omega_2$. Similarly, the difference in allele fitness can be related to the derivative of the mean fitness just noticing that $dq/dp = -1$ and using Equations (A.1) for ω_i and (A.2) for $\langle \omega \rangle = p^2\omega_{11} + 2pq\omega_{12} + q^2\omega_{22}$:

$$\begin{aligned} \frac{d\langle \omega \rangle}{dp} &= \frac{\partial \langle \omega \rangle}{\partial p} - \frac{\partial \langle \omega \rangle}{\partial q} = \\ &= 2 \, (p \, \omega_{11} + q \, \omega_{12}) - 2 \, (p \, \omega_{12} + q \, \omega_{22}) = \\ &= 2 \, (\omega_1 - \omega_2) \end{aligned} \tag{A.5}$$

Finally, combining Equations (A.5) and (A.4) we deduce

$$\mathrm{Var}(\omega) = \frac{pq}{4} \left(\frac{d\langle \omega \rangle}{dp} \right)^2, \tag{A.6}$$

which inserted in Equation (3.25) yields Equation (3.26) from the text:

$$\Delta \langle \omega \rangle = 2 \frac{\mathrm{Var}(\omega)}{\langle \omega \rangle}. \tag{A.7}$$

B The diffusion limit of the Moran Equation

We start from Equation (3.27) and Equation (3.28) for the transition probabilities. We also change from time steps to generations and apply the continuous limit. This yields

$$\begin{aligned}\mathcal{P}(x,\tau+d\tau) = (1-x+dx)(x-dx)\mathcal{P}(x-dx,\tau)+ \\ + \left(x^2+(1-x)^2\right)\mathcal{P}(x,\tau) + (x+dx)(1-x-dx)\mathcal{P}(x+dx,\tau). \quad \text{(B.1)}\end{aligned}$$

We expand up to first order in $d\tau$ and second order in dx. The motivation for this will be clear in a few steps.

$$\begin{aligned}&\mathcal{P} + \frac{\partial \mathcal{P}}{\partial \tau}\, d\tau = \\ &= (x - x^2 - (1-2x)dx - dx^2)\left(\mathcal{P} - \frac{\partial \mathcal{P}}{\partial x}\, dx + \frac{1}{2}\frac{\partial^2 \mathcal{P}}{\partial x^2}\, dx^2\right) + \\ &+ \left(1 + 2(x^2 - x)\right)\mathcal{P} + \\ &+ (x - x^2 + (1-2x)dx - dx^2)\left(\mathcal{P} + \frac{\partial \mathcal{P}}{\partial x}\, dx + \frac{1}{2}\frac{\partial^2 \mathcal{P}}{\partial x^2}\, dx^2\right).\end{aligned} \quad \text{(B.2)}$$

Now we develop the right hand side keeping only up to second order terms. We obtain

$$\frac{\partial \mathcal{P}}{\partial \tau}\, d\tau = -dx^2\left(2\mathcal{P} + 2\frac{\partial \mathcal{P}}{\partial x}(1-2x) + \frac{\partial^2 \mathcal{P}}{\partial x^2}(x-x^2)\right), \quad \text{(B.3)}$$

and recalling $d\tau = dx = 1/N$, the expression results in

$$\frac{\partial \mathcal{P}}{\partial \tau} = -\frac{1}{N}\left(2\mathcal{P} + 2\frac{\partial \mathcal{P}}{\partial x}(1-2x) + \frac{\partial^2 \mathcal{P}}{\partial x^2}(x-x^2)\right), \quad \text{(B.4)}$$

which is a first order Equation in (1/N). Note that this is due to the second order expansion in allele frequency. Since the first order cancels out, a second order expansion for an accurate non-trivial description of genetic drift it is necessary. Finally, grouping the right terms in (B.4) we obtain Equation (3.29):

$$\frac{\partial \mathcal{P}}{\partial \tau}(x,\tau) = -\frac{1}{N}\frac{\partial^2}{\partial x^2}\Big(x(1-x)\mathcal{P}(x,\tau)\Big). \quad \text{(B.5)}$$

C Entropy production in small systems at equilibrium

We start by recalling the definitions of $\Gamma = (C_0, ..., C_n)$ and $\Gamma_R = (C_0^R, ..., C_n^R)$, with $C_k^R = C_{n-k}$; and their associate protocols $\lambda = (\lambda_0, ..., \lambda_n)$ and $\lambda^R = (\lambda_0^R, ..., \lambda_n^R)$, with $\lambda_k^R = \lambda_{n-k}$. It is clear that $\mathcal{P}_F^{eq}(\Gamma)$ ($\mathcal{P}_R^{eq}(\Gamma_R)$) will be a function of the probabilities of finding the system at different C_k (C_k^R) for different control parameters λ_k (λ_k^R). Our aim is to deduce how this is done.

At this point, it is important the detailed balance condition (4.3) for the equilibrium distribution. In combination with the equilibrium condition, i.e $\mathcal{P}_{\lambda_{k+1}}^{eq}(C) = \mathcal{P}_{\lambda_k}^{eq}(C)$, this states that

$$\mathcal{P}_{\lambda_{k+1}}^{eq}(C_{k+1}) = \frac{W_{\lambda_k}(C_k \to C_{k+1})}{W_{\lambda_k}(C_{k+1} \to C_k)}\, \mathcal{P}_{\lambda_k}^{eq}(C_k). \quad \text{(C.1)}$$

Even if the $W(C \to C')$ might be different in both senses, this Equation tells something very important: the state at $k+1$ depends only on the state of the system at k. In other words, the

transition that takes place at k depends only on the state of the system at k. This allows us to treat each step of the path independently from the others, and write

$$\mathcal{P}_F^{eq}(\Gamma) = \prod_{k=1}^{n} \mathcal{P}_{\lambda_k}^{eq}(C_k) = \prod_{j=0}^{n-1} \mathcal{P}_{\lambda_{j+1}}^{eq}(C_{j+1}) = \prod_{j=0}^{n-1} \mathcal{P}_{\lambda_j}^{eq}(C_{j+1}). \tag{C.2}$$

Note that the product starts at $k = 1$ because the probability of finding the state at C_0 for λ_0 is trivially one (the starting point is fixed externally). In the last two steps we have just redefined the product index as $k - 1 = j$ and used again the equilibrium condition.

We can do the same for the inverse path. This yields

$$\mathcal{P}_R^{eq}(\Gamma_R) = \prod_{k=1}^{n} \mathcal{P}_{\lambda_k^R}^{eq}(C_k^R) = \prod_{k=1}^{n} = \mathcal{P}_{\lambda_{n-k}}^{eq}(C_{n-k}) = \prod_{j=0}^{n-1} \mathcal{P}_{\lambda_j}^{eq}(C_j), \tag{C.3}$$

where in the last step we have, again, redefined $n - k = j$ as the index over the product. We can now divide Equation (C.2) and Equation (C.3) to obtain

$$\begin{aligned} \frac{\mathcal{P}_F^{eq}(\Gamma)}{\mathcal{P}_R^{eq}(\Gamma_R)} &= \prod_{j=0}^{n-1} \frac{\mathcal{P}_{\lambda_j}^{eq}(C_{j+1})}{\mathcal{P}_{\lambda_j}^{eq}(C_j)} = \\ &= \prod_{j=0}^{n-1} \exp \log \left(\frac{\mathcal{P}_{\lambda_j}^{eq}(C_{j+1})}{\mathcal{P}_{\lambda_j}^{eq}(C_j)} \right) = \exp \left(\sum_{j=0}^{n-1} \log \frac{\mathcal{P}_{\lambda_j}^{eq}(C_{j+1})}{\mathcal{P}_{\lambda_j}^{eq}(C_j)} \right). \end{aligned} \tag{C.4}$$

Finally, recalling the definition of the entropy production introduced in (4.5) this last expression reads

$$\frac{\mathcal{P}_F^{eq}(\Gamma)}{\mathcal{P}_R^{eq}(\Gamma_R)} = e^{S_p(\Gamma)} \tag{C.5}$$

D The fitness flux theorem

D.1 Total entropy of the Kimura-Ohta Equation

We start by obtaining the K-O equilibrium distribution. Since the equilibrium distribution is time-independent, we have to solve:

$$0 = \frac{\partial}{\partial x} \Big[- \big(m(x) + g(x)s(x)\big)\mathcal{P}^{eq}(x) + \frac{1}{N} \frac{\partial}{\partial x} \big(g(x)\mathcal{P}^{eq}(x)\big) \Big]. \tag{D.1}$$

Since x is a finitely-bounded variable, the probability flux has to be zero, and condition (D.1) is equivalent to solve

$$\begin{aligned} 0 = &- \left(m(x) + g(x)s(x)\right) \mathcal{P}^{eq}(x) + \frac{1}{N} \frac{\partial}{\partial x} \left(g(x)\mathcal{P}^{eq}\right)(x) \Longrightarrow \\ \Longrightarrow &\frac{1}{g(x)\mathcal{P}^{eq}(x)} \frac{\partial}{\partial x} \left(g(x)\mathcal{P}^{eq}(x)\right) = N \left(\frac{m(x)}{g(x)} + s(x) \right). \end{aligned} \tag{D.2}$$

This Equation is easily integrable and yields

$$\begin{aligned} \mathcal{P}^{eq}(x) &= \frac{(g\mathcal{P}^{eq})(0)}{g(x)} \exp \int dx\, N \left(\frac{m(x)}{g(x)} + s(x) \right) = \\ &= \frac{(g\mathcal{P}^{eq})(0)}{g(x)} \left(\exp N \int dx\, \frac{m(x)}{g(x)} \right) \left(\exp\, N \int dx\, s(x) \right) = \\ &= \mathcal{P}_0(x) \exp \left(N \int dx\, s(x) \right), \end{aligned} \tag{D.3}$$

where $\mathcal{P}_0(x)$ is the equilibrium distribution in absence of selection, that is, $\mathcal{P}_0(x) = \mathcal{P}^{eq}(x, s = 0)$.

Now, we recall the definition of the selection coefficient in terms of the fitness landscape: $s(x) = \nabla F(x)$. This provides

$$\int dx\, s(x) = \int dx\, \nabla F(x) = F(x) - F(0). \tag{D.4}$$

The constant term $F(0)$ is just a normalisation factor that can be absorbed by the neutral equilibrium distribution. The final result is

$$\mathcal{P}^{eq}(x) \simeq \mathcal{P}_0(x)\, e^{NF(x)}. \tag{D.5}$$

This expression can be used now to obtain the total dissipation or entropy of the system. From Equation (4.4), we have (recall that the control parameter is time in this case and frequency is the state):

$$S(\Gamma) = \sum_{k=0}^{n-1} \log \frac{\mathcal{P}^{eq}(x(t_k + \Delta t))}{\mathcal{P}^{eq}(x(t_k))} - \log\left(\frac{\mathcal{P}(x(t_n), t_n)}{P(x(t_0), t_0)}\right). \tag{D.6}$$

Introducing $x(t_k) = x_k$ and $x(t_k + \Delta t) \approx x(t_k) + \Delta x$, one can apply Equation (D.5) to obtain

$$S(\Gamma) \approx \sum_{k=0}^{n-1} \log \exp\Big(N\big(F(x_k+\Delta x) - F(x_k)\big)\Big) - \left(\log\left(\frac{\mathcal{P}(x_n, t_n)}{P_0(x)}\right) - \log\left(\frac{\mathcal{P}(x_0, t_0)}{P_0(x)}\right)\right). \tag{D.7}$$

One can expand $F(x_k+\Delta x) - F(x_k) \approx \nabla F(x_k)\Delta x_k$. Introducing the same definition for $\mathcal{H}(x,t)$ as in Sec. 5, Equation (D.7) reads

$$S(\Gamma) = N \sum_{k=0}^{n-1} \nabla F(x_k)\Delta x_k - \mathcal{H}(x_n, t_n) + \mathcal{H}(x_0, t_0). \tag{D.8}$$

Finally, with definition (5.1) for the cumulative fitness flux, the total entropy for the K-O Equation adopts the expression (5.3) of the text:

$$S(\Gamma) = N\Phi(\Gamma) - \Delta\mathcal{H}(\Gamma). \tag{D.9}$$

D.2 Strong selection limit of the fitness flux theorem

We start by recalling Equation (A.4) for the variance in fitness of a two-allele gene,

$$\mathrm{Var}(\omega) = pq(\omega_1 - \omega_2)^2. \tag{D.10}$$

Note that this expression does not specify if the individual is diploid or haploid: no assumption is made other than a set of two values $\{\omega_1, \omega_2\}$, so the same expression must hold for any other generic variable. Concretely, for the (normalised) fitness f introduced in the text, one has

$$\mathrm{Var}(f) = x(1-x)(f_1 - f_2)^2 = x(1-x)s^2(x,t), \tag{D.11}$$

where we have introduced the selection coefficient as defined in (3.32) and the continuous notation for the allele frequency.

We aim now to calculate the variation in time of the mean cumulative flux in the limit when selection is constant, $s(x,t) = s$, and dominant over drift and mutations, $s \gg \mu,\, 1/N$.

The first assumption yields a fitness landscape in which the cumulative fitness flux just equals the selection coefficient, $\Phi(\Gamma) = s$. As a consequence, the mean cumulative fitness flux reads

$$\langle \Phi \rangle = x s_1 + (1 - x) s_2. \tag{D.12}$$

Definition (3.32) for the selection coefficients can be generalised to k alleles (each having allele fitness f_i) as

$$s_i = f_i - f_k. \tag{D.13}$$

It is clear this way that $s_k = 0$, a condition that simply reproduces the fact that in haploid organisms having k alleles, due to normalisation, only $k - 1$ of the allele fitnesses are linearly independent. With this more general picture, it is then obvious that, for the two allele case, $s_1 = s$ and $s_2 = 0$. Equation (D.12) can be written as

$$\langle \Phi \rangle = sx, \tag{D.14}$$

and taking its time derivative yields

$$\frac{d}{dt}\langle \Phi \rangle = s \frac{dx}{dt}. \tag{D.15}$$

The second assumption regards the time evolution of the allele frequencies. When selection is the dominance force, drift and mutations can be neglected and the K-O Equation reduces to Equation (3.31) for selection, that is,

$$\frac{dx}{dt} = x(1 - x)(f_1 - f_2) = x(1 - x)s. \tag{D.16}$$

Finally, by introducing Equation (D.15) into (D.14) one gets

$$\Delta\langle \Phi \rangle \simeq \frac{d}{dt}\langle \Phi \rangle = x(1 - x)\, s^2, \tag{D.17}$$

and, recalling expression (D.11) for the variance,

$$\Delta\langle \Phi \rangle = \mathrm{Var}(f) \tag{D.18}$$

which is an analogous of expression (3.26) for Fisher's fundamental theorem of selection.