This document is published in:

Newsletter of the European Mathematical Society, nº 76 (june 2010), pp. 28-37

Mathematics of evolution

José A. Cuesta

> A mathematician is a blind man in a dark room looking for a black cat which isn't there.
> *Charles Darwin*

It is somehow strange to read an article on Darwin and evolution in a mathematical journal. Both are always associated, obviously, with biology. However, evolutionary theory, like every other theory deserving that name, admits quantitative formulations of many of its aspects. Darwin himself, not very skilled in mathematics but nevertheless an educated scientist, acknowledged that "every new body of discovery is mathematical in form, because there is no other guidance we can have". For Darwin, mathematicians were people with a sixth sense that allowed them to "see" in places of the mind where everybody else is "blind". Hence his quotation at the beginning of this article which, funny as it may seem, expresses Darwin's admiration for mathematicians.[1]

Darwin's great contribution, the theory of evolution by natural selection [3], received a fundamental input when the laws of inheritance, discovered by Gregor Mendel in 1865 [15] and rediscovered by de Vries, Correns and von Tschermak in 1900, were incorporated into the theory. From that moment on, and in a way that later became known as *population genetics,* a group of mathematicians, amongst whom Fisher, Haldane, Wright and later Kimura are prominent examples, laid the foundations of the mathematical theory of evolution. Nowadays, this theory has a status of its own within the field of applied mathematics, and it has developed and diversified, allowing us to understand the subtle mechanisms that evolution operates with, not just in biology but in many other disciplines sharing similar principles, like linguistics, economics, sociology and computer science.

Last year we celebrated Darwin's bicentennial as well as the sesquicentennial of the publication of *The Origin of Species,* and these anniversaries provide an appropriate excuse to revise evolutionary theory from a mathematical point of view. This is the goal of this article. It should be clarified at this point that mathematical contributions to evolutionary theory are so many and so diverse that only a few of them can be sampled here. Besides, the divulgatory aim of this article discourages any attempt at deeply reviewing them, so the interested reader is referred to the excellent texts on the matter that are referenced in this article [5, 7, 16]. Another clarification is also needed. There are two kinds of reproductive mechanisms in living beings: asexual, by which an organism can replicate by itself alone, and sexual, by which the intervention of more than one organism (almost always two) is necessary for reproduction. The former is typical (but not exclusive) of simple organisms, like viruses, and is the subject matter of most mathematical models; the latter implies combining genetic material coming from at least two parents, which leads to particular complications that are also the subject matter of more elaborate models. This article will almost exclusively deal with asexual reproduction (which is already complicated enough), although sexual reproduction will occasionally be mentioned at specific points.

1 Fundamental mechanisms of evolution

Upon reflection of what is necessary, at an abstract level, for an evolutionary process to occur, no matter what the context is, one realizes that the necessary condition is the concurrence of three fundamental mechanisms: *replication,* the mechanism by which entities can create copies of themselves; *mutation,* the mechanism that generates small variations within those copies; and *selection,* the mechanism by which the "best" copies are able to eliminate all the others generation after generation. Let us consider these three mechanisms one by one.

Replication

A typical bacteria divides every 20 minutes, generating two copies of itself; 20 minutes later there will be four bacteria; after one hour there will be eight... Bacterial population in a generation n_t is related to that of the previous generation by the simple equation $n_t = 2n_{t-1}$, whose solution, assuming $n_0 = 1$, is $n_t = 2^t$. A replicative process like this one leads to exponential growth. It was Malthus who first proposed this law of growth in his book *An Essay on the Principle of Population* [13]. According to this law, assuming a situation in which generations are intertwined, if a population $n(t)$ reproduces at a constant replication rate per individual r, i.e. $\dot{n} = rn$, then $n(t) = n(0)e^{rt}$. This is Malthus' model for human population growth and the continuum version of the bacterial reproduction law that we have just obtained.

Malthus was one of the most important influences on Darwin's thoughts because of what has been referred to as "Malthusian catastrophe". It can be illustrated with the example of bacteria. According to the law we have obtained, after only

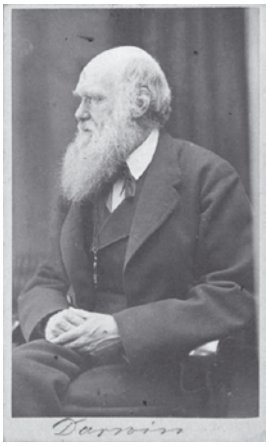

Figure 1. A photograph of Charles Darwin, circa 1871, by Oscar Gustave Rejlander (1813–1875). Source: Wikimedia Commons.

two days (144 divisions) we will have $2^{144} \approx 2 \times 10^{43}$ bacteria. More precisely, if the bacterial diameter is $\sim 1\,\mu$m and the bacterial density is that of water ($1\,\mathrm{g\,cm^{-3}}$) then 2×10^{43} bacteria have a mass of approximately 2×10^{28} kg, i.e. about 3000 times the mass of the Earth! Obviously such a growth cannot be sustained and should drive most individuals to extinction. In fact, the lack of resources can be effectively added to Malthus' law by replacing the constant replication rate per individual with a decreasing function of the population that vanishes at the point where the population reaches its sustainability limit. In its simplest form this leads to the *logistic* law

$$\dot{n} = r\left(1 - \frac{n}{K}\right)n, \tag{1}$$

whose solution

$$n(t) = \frac{Kn_0e^{rt}}{K + n_0(e^{rt} - 1)}, \qquad \lim_{t\to\infty} n(t) = K \tag{2}$$

is a curve that increases monotonically in time (when $n_0 < K$) up to a saturation value K, referred to as the "carrying capacity" of the environment. Therefore we can consider that populations undergoing such a growth process reach an equilibrium with their environment that keeps the population constant.[2] In practice, the population will f uctuate around that value due to stochastic effects that the kind of laws we are considering simply neglect.

Selection

In a situation where resources are scarce, as we have just described, individuals struggle to obtain them, survive and reproduce. Competition with like individuals yields the saturation predicted by the logistic law but when there are individuals of different types (species), their differences, no matter how small, play a crucial role in deciding who survives and who perishes. The key evolutionary parameter here is *fi ness,* def ned as the mean number of adults that an individual yields in the next generation. In case of Malthusian populations, f tness is measured by the parameter r. If logistic, the f tness, $f = r(1 - n/K)$, will depend on the total population. In the general case, f will be a function of the total population.

Suppose the population is made up of n different species with fit ess f_i, $i = 1, \ldots, n$. By definition their respective populations grow according to $\dot{n}_i = f_i n_i$. The total population $N = \sum_{i=1}^n n_i$ will thus grow as

$$\dot{N} = \sum_{i=1}^n \dot{n}_i = \sum_{i=1}^n f_i n_i = N\phi, \qquad \phi = \sum_{i=1}^n f_i x_i, \tag{3}$$

$x_i = n_i/N$ being the fraction of the population corresponding to species i and ϕ being the population *mean fi ness.* The total population is thus Malthusian with replication rate ϕ (remember that f_i can depend on N). We can now obtain a growth law for x_i,

$$\dot{x}_i = \frac{\dot{n}_i}{N} - x_i\frac{\dot{N}}{N} = x_i(f_i - \phi). \tag{4}$$

This evolutionary law is commonly known as the *replicator equation* [16] and not only describes the evolution of biological systems but also plays a prominent role in game theory [8]. The law followed by the population fractions is similar to Malthus' law, only that now the replication rate is measured with respect to its mean over the population. This means that the population fraction of a species will increase only if its fitne s is above that mean and will otherwise decrease. We can envisage here the principle of "survival of the f ttest". In fact, it is very easy to derive this principle from equation (3). Let us assume that species k is fitte than the rest of them, i.e. $f_k > f_i\ \forall i \neq k$. The evolution equation for x_k can be rewritten as

$$\begin{aligned}\dot{x}_k &= x_k\Big(f_k - \sum_{i=1}^n f_i x_i\Big) = x_k \sum_{i=1}^n (f_k - f_i)x_i \\ &= x_k \sum_{i\neq k,\, i=1}^n (f_k - f_i)x_i,\end{aligned} \tag{5}$$

where we have used $\sum_{i=1}^n x_i = 1$. It follows from (5) that the sum on the right side will be positive as long as there is at least one species $i \neq k$ with $x_i > 0$, and in that case x_k will increase in time at the expense of the population fractions of the remaining species. In other words,

$$\lim_{t\to\infty} x_k(t) = 1, \qquad \lim_{t\to\infty} x_i(t) = 0, \quad \forall i \neq k, \tag{6}$$

which expresses mathematically the principle of survival of the f ttest.

In the case where the f tness is constant we have a result, due to Fisher, referred to as the *fundamental theorem of natural selection* [5]. Its derivation amounts to computing

$$\begin{aligned}\dot{\phi} &= \sum_{i=1}^n f_i\dot{x}_i = \sum_{i=1}^n f_i x_i(f_i - \phi) = \sum_{i=1}^n x_i(f_i - \phi)^2 \\ &= \sigma_f^2 \geq 0.\end{aligned} \tag{7}$$

Put in a different way, the mean f tness never decreases in time and its growth rate is the f tness variance over the population. Thus it will increase as long as there is variability in the population and will do so by increasing the population of the f ttest.

Mutation

Replication is not error-free. In general, replication errors lead to non-viable individuals that cannot survive. These errors can be accounted for by adjusting the replication rate. Occasionally, however, a mutation can produce an offspring of a different and viable type. Thus mutation can be regarded as a stochastic process by which individuals of species i produce individuals of species j with a probability q_{ij} ($\ll 1$). This mechanism introduces variability in an otherwise homogeneous population. The replicator equation (3) must be modif ed to account for this new process:

$$\dot{x}_i = x_i\Big(f_i - f_i\sum_{j\neq i} q_{ij} + \sum_{j\neq i} f_j q_{ji} - \phi\Big), \tag{8}$$

where the f rst new term accounts for mutations transforming individuals of species i into individuals of any other species and the second one accounts for mutations transforming individuals of any other species into individuals of species i. Defi ing $q_{ii} = 1 - \sum_{j\neq i} q_{ij}\ (\geq 0)$ and introducing the stochastic matrix $Q = (q_{ij})$ (which we will refer to as the *mutation matrix),* the equation above can be rewritten in vector form as

$$\dot{\mathbf{x}} = \mathbf{x}FQ - \phi\mathbf{x}, \tag{9}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ and $F = \mathrm{diag}(f_1, \ldots, f_n)$. This equation is termed the *replicator-mutator equation* [16]. The stochastic character of Q can be expressed in matrix form as

$Q\mathbf{1}^\mathsf{T} = \mathbf{1}^\mathsf{T}$, where $\mathbf{1} = (1, \ldots, 1)$. Because of this, as $\phi = \mathbf{x}F\mathbf{1}^\mathsf{T} = \mathbf{x}FQ\mathbf{1}^\mathsf{T}$, it follows immediately that $\mathbf{x}\mathbf{1}^\mathsf{T} = 1$ is a constraint preserved by equation (9).

Since FQ is a non-negative matrix, Perron-Frobenius theory [18] tells us that the rest points of the dynamical system (9) are all given by the left-eigenvectors of matrix FQ corresponding to its largest eigenvalue, all of whose components are non-negative. Furthermore, if Q is irreducible then there is only one such eigenvector. Every irreducible matrix has at least one non-diagonal element in each row, so the equilibrium vector necessarily has more than one non-zero component. Therefore, when all species can mutate, the equilibrium population cannot be homogeneous. This is an interesting observation because it implies that ϕ is not at its maximum in equilibrium, as it was when there were no mutations; there is competition between selection, which pushes ϕ towards its maximum value, and mutation, which tends to decrease it. We can try to see what happens to the fundamental theorem in this case. A similar reasoning to the one leading to equation (7) yields

$$\dot{\phi} = \mathbf{x}(FQ - \phi I)^2\mathbf{1}^\mathsf{T}. \tag{10}$$

Despite its similarity to σ_f^2, the non-negativity of this term is no longer guaranteed.

2 The problem of reversion

Despite Mendel publishing his work almost simultaneously with Darwin, it seems that Darwin was never aware of its existence. This put Darwin in serious trouble. Reading *The Origin of Species* one realizes how Darwin stumbles once and again over it: the inheritance theory in sexually reproducing species that was accepted in his time (formulated by Galton [2] in 1875) led to the phenomenon of reversion. According to this theory, every progenitor contributes f fty-f fty to a given trait, e.g. height. We can then formulate a simple stochastic law for the quantitative value of the trait of an offspring given those of its parents: $X_{n+1} = \frac{1}{2}(X_n^{(1)} + X_n^{(2)}) + Z_n$, where $X_n^{(1)}$ and $X_n^{(2)}$ are two stochastic variables, identically distributed according to $P_n(x) = \Pr(X_n \leq x)$, representing the parents' traits in generation n, and where Z_n is a noise, which we will assume to be normally distributed $N(0, \sigma)$ for all n. If $F_n(q) = \int_{-\infty}^{\infty} e^{iqx}\, dP_n(x)$ denotes the characteristic function of distribution P_n then $F_{n+1}(q) = F_n(q/2)^2 e^{-\sigma^2 q^2/2}$. The solution to this equation is

$$\log F_n(q) = 2^n \log F_0(2^{-n}q) - \frac{\sigma^2}{2}q^2 \sum_{k=0}^{n-1} 2^{-k} \tag{11}$$

and therefore

$$\lim_{n\to\infty} F_n(q) = e^{qF_0'(0) - \sigma^2 q^2}. \tag{12}$$

Since $-iF_0'(0) = \mu$, the mean value of the initial distribution $P_0(x)$, we conclude that $P_n(x)$ approaches the normal distribution $N(\mu, \sqrt{2}\sigma)$ as $n \to \infty$. So, regardless of the initial distribution (provided it has a fi ite mean), the distribution of the trait approaches a normal distribution with the same mean value as the initial distribution. This means that, even if the latter was bimodal, in the end all the population ends up being of homogeneous type (up to some noise). The difficulty this poses to Darwin's theory is to suppress the variability introduced by mutations. This is the reason why Darwin often resorted to the argument that mutant populations must remain isolated for quite some time in order for a new species to emerge.

3 Mendel, or the "quantum" theory of inheritance

Mendel's crucial contribution was to discover that traits are transmitted in "quanta" of inheritance (what we nowadays call *genes)* that do not admit gradations.[3] Traits such as height, which seem to violate this principle, are but complex traits resulting from the combined effect of several simple traits, all of which are "quantum" (either they are present or absent). Every sexed individual carries two of those quanta per trait, one from its father and another from its mother, and in its turn transmits one of them (at random) to each of its offspring.

Consider a trait (e.g. the red colour of a rose) determined by a variant (an *allele* in genetic parlance) A of the corresponding gene. Suppose that this allele has a muted variant a, which does not produce colour. According to Galton's inheritance model, hybrid descendants should show a pink colour grading, generation after generation, until its eventual return to the original red colour. According to Mendel's laws, if the parent generation is made of p A-alleles and q a-alleles, if mating is random and if the population is sufficiently large then there will be p^2 AA-individuals, $2pq$ Aa-individuals and q^2 aa-individuals in the next generation. Furthermore, the distribution will remain stable in successive generations. This result is known in genetics as the Hardy-Weinberg law [5]. It implies that there is no reversion to the wild type: pure aa mutants remain in the population, generation after generation, in a ratio which depends on the initial amount of mutants. Mendelian genetics therefore is responsible for the maintenance of the variability in the populations introduced by mutations.

4 The fourth element: genetic drift

In section 1, we mentioned that replication, selection and mutation form the basic triad of evolutionary dynamics. Although strictly speaking this is true, there is a fourth ingredient that is unnecessary in principle and yet becomes crucial in understanding the mechanisms of adaptation and speciation undergone by evolving entities: *genetic drift.* All the previous discussion presumes that evolving populations are inf nitely large; hence the deterministic dynamics we have introduced so far and the validity of principles like the Hardy-Weinberg law. But when it comes to f nite populations some noise appears due to the statistical sampling inherent to replicative processes. This noise is what we refer to as genetic drift. Its effects can be dramatic in small populations; that is why it becomes crucial when populations traverse an evolutionary "bottleneck", i.e. a situation in which the population gets strongly reduced, because of epidemics, climate change, geographical isolation, etc.

There are two basic mathematical models of genetic drift: Fisher-Wright and Moran [5]. Both assume a constant population, limited by the carrying capacity of the environment. The f rst one describes situations in which every generation

replaces the previous one (such as in stational plants); the second one describes the case in which generations can overlap.

Fisher-Wright model

Assume that we deal with a population of N individuals and focus on a particular trait (e.g. the red colour of a rose again). Suppose that this character is determined by the presence of an allele A, of which there is a mutant a producing a different colour (e.g. white). Assume further that the population reproduces asexually (i.e. A begets A and a begets a) and that mutations A $\to$ a occur with probability μ and mutations a $\to$ A occur with probability ν. If initially there are k A-individuals and there is no selective difference between both alleles (both have the same fit ess) then the probability that an individual in the next generation is of type A will be $\psi_k = (k/N)(1-\mu) + (1-k/N)\nu$. Imagine that every individual yields a large number of offspring to a pool and that we randomly extract N individuals from that pool to form the next generation. If $X_t \in \{0, 1, \ldots, N\}$ is a random variable describing the number of A-alleles in generation t then the Fisher-Wright model is defi ed as a Markov chain with transition probability

$$P_{kj} = \Pr(X_{t+1} = j \mid X_t = k) = \binom{N}{j} \psi_k^j (1-\psi_k)^{N-j}. \qquad (13)$$

The chain is ergodic for all $0 < \mu, \nu < 1$ because $P_{kj} > 0$ $\forall k, j \in \{0, 1, \ldots, N\}$, thus there is a unique stationary probability distribution $\mathbf{w} = (w_0, w_1, \ldots, w_N)$ that can be obtained as the solution to the equation $\mathbf{w} = \mathbf{w}P$ [10, 18]. There is no analytic expression for $\mathbf{w}$ except for a very particular case: $\mu + \nu = 1$ (of no biological interest because mutation rates are too large). In this case $\psi_k = \nu = 1 - \mu$ and

$$(\mathbf{w}P)_j = \binom{N}{j} \nu^j (1-\nu)^{N-j} \qquad (14)$$

for every $\mathbf{w}$, so the right side of the previous equation describes the stationary distribution.

What can be calculated in general is the mean value of $\mathbf{w}$. To obtain it let us denote $X_\infty = \lim_{t\to\infty} X_t$ and defin the vector $\boldsymbol{\xi} = (0, 1, \ldots, N)$. We can then write $E(X_\infty) = \mathbf{w}\boldsymbol{\xi}^\mathsf{T} = \mathbf{w}P\boldsymbol{\xi}^\mathsf{T}$. Now,

$$\begin{aligned}(P\boldsymbol{\xi}^\mathsf{T})_k &= \sum_{j=0}^{N} j \binom{N}{j} \psi_k^j (1-\psi_k)^{N-j} = N\psi_k \\ &= k(1-\mu) + (N-k)\nu, \qquad (15)\end{aligned}$$

so $E(X_\infty) = (1-\mu)E(X_\infty) + [N - E(X_\infty)]\nu$; hence $E(X_\infty) = N\nu/(\mu+\nu)$. By a similar, albeit more tedious, procedure we can obtain the variance $\sigma^2 = N^2\mu\nu/[(\mu+\nu)^2(1+2N\mu+2N\nu)] + \epsilon$, where ϵ contains terms of smaller order (for instance, if $\mu, \nu = O(N^{-1})$ then $\epsilon = O(N)$).

The most interesting case to be considered is that in which there is only genetic drift ($\mu = \nu = 0$, thus $\psi_k = k/N$). The Markov chain is not ergodic anymore because it has two absorbing states: $k = N$ (all individuals are of type A) and $k = 0$ (all individuals are of type a). This means that, regardless of the initial population and in spite of the lack of selective factors, eventually one type invades the whole population. The interesting magnitude now is the probability $\pi_j = \Pr(X_\infty = N \mid X_0 = j)$ that the population ends up being of type A given that there were initially j individuals of that type. Denoting $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_N)$, it can be shown that $\boldsymbol{\pi}^\mathsf{T} = P\boldsymbol{\pi}^\mathsf{T}$ and that $\pi_0 = 0$ and $\pi_N = 1$. However, the simplest way to fi d $\boldsymbol{\pi}$ is by showing that this Markov chain is a martingale, i.e. that $E(X_t \mid X_{t-1}) = X_{t-1}$ (the proof is simple: it is the mean value of a binomial distribution). This means that $E(X_\infty) = j$. But this mean value can also be obtained as $E(X_\infty) = N\pi_j + 0(1-\pi_j)$, whereby $\pi_j = j/N$. An interesting by-product of this result is the probability that a single mutant invades the population: $\pi_1 = 1/N$. This probability is small in large populations but non-negligible during evolutionary bottlenecks, so the f xation of a new allele is something that surely has occurred more than once in the past.

Moran model

The Moran model is more interesting from a theoretical point of view because it is more amenable to analytic treatment than the Fisher-Wright model. It describes the same situation: a population of N individuals, k of which are A-alleles and $(N-k)$ a-alleles. The difference is that now individuals reproduce at a constant rate in time τ. The offspring will be a mutant with the same probabilities μ and ν as the Fisher-Wright model. After the reproduction event the newborn will replace a random individual of the population (even its parent!) chosen with uniform probability. The Markov process is specifie by the conditional probabilities

$$\Pr(X(t+dt) = j \mid X(t) = k) = \tau T_{kj}\, dt, \qquad \forall j \neq k, \qquad (16)$$

where $X(t)$ represents the population of A individuals at time t. The magnitude of interest in this stationary process is $P_{ij}(t) = \Pr(X(t+s) = j \mid X(s) = i)$, the solution to either the equation resulting from multiplying (16) by $P_{ik}(t)$ and summing for $0 \le k \le N$

$$\dot{P}_{ij}(t) = \tau \sum_{k=0}^{N} [P_{ik}(t)T_{kj} - P_{ij}(t)T_{jk}], \qquad (17)$$

or the equation resulting from multiplying (16) by $P_{jl}(t)$ and summing for $0 \le j \le N$ (with an appropriate change of indices),

$$\dot{P}_{ij}(t) = \tau \sum_{k=0}^{N} [T_{ik}P_{kj}(t) - T_{ik}P_{ij}(t)]. \qquad (18)$$

In both cases the initial condition is, of course, $P_{ij}(0) = \delta_{ij}$. Equations (17) and (18) are, respectively, the forward and backward forms of the *master equation* of the process.

In any inf nitesimal time interval $(t, t+dt)$ there can be at most one reproduction event in a Moran process, so if the state at time t is k, at time $t+dt$ it can be k, $k+1$ or $k-1$. Denoting $\tau T_{k\,k+1} = \lambda_k$ and $\tau T_{k\,k-1} = \mu_k$, equation (17) becomes the *forward Kolmogorov equation*

$$\dot{P}_{ij}(t) = \mu_{j+1}P_{i\,j+1}(t) - (\mu_j + \lambda_j)P_{ij}(t) + \lambda_{j-1}P_{i\,j-1}(t), \qquad 0 \le j \le N, \qquad (19)$$

and equation (18) the *backward Kolmogorov equation*

$$\dot{P}_{ij}(t) = \lambda_i P_{i+1\,j}(t) - (\mu_i + \lambda_i)P_{ij}(t) + \mu_i P_{i-1\,j}(t), \qquad 0 \le j \le N, \qquad (20)$$

where $P_{i\,-1}(t) = P_{-1\,i}(t) = P_{i\,N+1}(t) = P_{N+1\,i}(t) = \mu_0 = \lambda_N = 0$. This kind of Markov process is referred to as a *birth-death process* [10]. For the Moran model $\lambda_k = \tau\psi_k(1-k/N)$ (the probability that an A-allele is created and replaces an a-allele)

and $\mu_k = \tau(1-\psi_k)k/N$ (the probability that an A-allele is created and replaces an a-allele), with ψ_k as in the Fisher-Wright model.

In the steady state $w_j = \lim_{t\to\infty} P_{ij}(t)$ can be obtained by setting $\dot{P}_{ij}(t) = 0$ in (19), which yields $\mu_{j+1}w_{j+1} - \lambda_j w_j = \mu_j w_j - \lambda_{j-1}w_{j-1}$ for all $1 < j < N$. For $j = 0$ we have $\mu_1 w_1 - \lambda_0 w_0 = 0$, thus $\mu_j w_j - \lambda_{j-1}w_{j-1} = 0$, a simple difference equation whose solution is

$$w_j = w_0 \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1\mu_2\cdots\mu_j}, \qquad 0 < j \le N, \tag{21}$$

or alternatively

$$w_j = w_N \frac{\mu_{j+1}\mu_{j+2}\cdots\mu_N}{\lambda_j\lambda_{j+1}\cdots\lambda_{N-1}}, \qquad 0 \le j < N. \tag{22}$$

The values w_0 or w_N are determined through the normalization $\sum_{j=0}^{N} w_j = 1$.

For the Moran model with mutations a closed expression for $\mathbf{w}$ can be obtained in the limit $N \to \infty$, $k \to \infty$ with $k/N = x \in [0,1]$, $N\mu \to \gamma$ and $N\nu \to \kappa$ [10]. For that we write $\lambda_j = A(N-j)j[1+a/j]$ and $\mu_j = A(N-j)j[1+b/(N-j)]$, where $A = \tau(1-\mu-\nu)/N^2$, $a = N\nu/(1-\mu-\nu)$ and $b = N\mu/(1-\mu-\nu)$. Then,

$$w_k = \frac{w_0 Na}{k(N-k)[1+b/(N-k)]} \prod_{j=1}^{k-1} \frac{1+a/j}{1+b/(N-j)}. \tag{23}$$

Now, using the Taylor expansion for $\log(1+x)$, the asymptotic behaviour $\sum_{j=1}^{k} j^{-1} \sim \log k$ when $k \to \infty$ and the fact that $\sum_{j=1}^{\infty} j^{-p} < \infty$ for all integers $p > 1$,

$$\begin{aligned} \sum_{j=1}^{k-1} \log\left(1+\frac{a}{j}\right) &\sim \log(k^a) + c, \\ \sum_{j=1}^{k-1} \log\left(1+\frac{b}{N-j}\right) &\sim \log\left(\frac{N^b}{(N-k)^b}\right) + d \end{aligned} \tag{24}$$

when $k \to \infty$, for certain constants c and d. Therefore (considering that in this limit $b/(N-k) \to 0$),

$$w_k \sim Ck^{\kappa-1}\left(1-\frac{k}{N}\right)^{\gamma-1}, \tag{25}$$

C being a normalization constant. In this limit

$$\begin{aligned} \sum_{k=0}^{N} w_k = CN^\kappa \sum_{k=0}^{N} \frac{1}{N}\left(\frac{k}{N}\right)^{\kappa-1}\left(1-\frac{k}{N}\right)^{\gamma-1} \\ \sim CN^\kappa \int_0^1 x^{\kappa-1}(1-x)^{\gamma-1}\,dx, \end{aligned} \tag{26}$$

therefore the stationary probability distribution $w_k \sim x^{\kappa-1}(1-x)^{\gamma-1}\,dx/B(\kappa,\gamma)$ has the shape of a beta distribution with parameters κ and γ.

When there are no mutations ($\mu = \nu = 0$ and $\psi_k = k/N$), $\lambda_0 = \mu_N = 0$. Then (21) and (22) lead, respectively, to the two vectors $\mathbf{w} = (1,0,\ldots,0)$ and $\mathbf{w} = (0,\ldots,0,1)$, thus expressing the fact that, with probability 1, the system ends up absorbed either in state $k = 0$ or in state $k = N$. The magnitude that characterizes this absorption is $\pi_k = \lim_{t\to\infty} P_{kN}(t)$, i.e. the probability that if the process starts with k A-alleles it ends up with N A-alleles. The equation for $\boldsymbol{\pi}$ can be obtained by setting $\dot{P}_{ij}(t) = 0$ in (20). The resulting equation can be written as $\lambda_k(\pi_{k+1} - \pi_k) = \mu_k(\pi_k - \pi_{k-1})$, where $0 < k < N$. Taking into account that $\pi_0 = 0$, this equation implies that $\pi_k - \pi_{k-1} = q_{k-1}\pi_1$, for all $0 < k \le N$, where

$$q_0 = 1, \qquad q_k = \frac{\mu_1\mu_2\cdots\mu_k}{\lambda_1\lambda_2\cdots\lambda_k}. \tag{27}$$

Thus $\pi_k = \pi_1 \sum_{j=0}^{k-1} q_j$ for all $0 < k \le N$ and π_1 follows from the condition $\pi_N = 1$. The shape of π_k for the Moran process is very simple because $\mu_j = \lambda_j$ for all $0 \le j \le N$. Therefore $\pi_k = k/N$, just as in the Fisher-Wright process.

Karlin and McGregor proved [9] that the solution to (20) can be expressed in the form

$$P_{ij}(t) = \frac{w_j}{w_0}\int_0^\infty e^{-xt} R_i(x)R_j(x)\,d\varphi(x), \tag{28}$$

where w_j is given by (21), $R_j(x)$ is the (f nite, because $\lambda_N = 0$) system of polynomials def ned by the three-term recurrence

$$\begin{aligned} -xR_j(x) = \lambda_j R_{j+1}(x) - (\lambda_j+\mu_j)R_j(x) + \mu_j R_{j-1}(x), \\ 0 \le j < N, \end{aligned} \tag{29}$$

with $R_{-1}(x) = 0$ and $R_0(x) = 1$, and $\varphi(x)$ is a unique measure with unit mass and with increments in $N+1$ points, with respect to which the family of polynomials is orthogonal.

And a bonus of the Moran model is that it permits the inclusion of selection. If f_{A} and f_{a} denote the f tness of the two alleles, the probability that, given a reproduction event, the individual that reproduces is of type x will be proportional to f_{x}. This yields a new expression for ψ_k, namely

$$\psi_k = \frac{kf_{\mathsf{A}}(1-\mu) + (N-k)f_{\mathsf{a}}\nu}{kf_{\mathsf{A}} + (N-k)f_{\mathsf{a}}}. \tag{30}$$

Similar arguments to those employed, in the limit of large populations and small mutations, to f nd the stationary distribution in the absence of selection now lead to

$$w_k \sim (N\log r^{-1})^{\kappa/r} x^{\kappa/r-1} r^{Nx}\,dx/\Gamma(\kappa/r) \text{ if } r < 1$$

and

$$w_k \sim (N\log r)^{\gamma r}(1-x)^{\gamma r-1} r^{N(x-1)}\,dx/\Gamma(\gamma r) \text{ if } r > 1,$$

where $r = f_{\mathsf{A}}/f_{\mathsf{a}}$ is the f tness of the A-alleles relative to that of the a-alleles.

In the absence of mutations $\mu_j/\lambda_j = r^{-1}$, thus $\pi_k = (1-r^{-k})/(1-r^{-N})$. Therefore, the f xation probabilities of a mutant allele ($\rho_{\mathsf{A}} = \pi_1$ or $\rho_{\mathsf{a}} = 1-\pi_{N-1}$) become

$$\rho_{\mathsf{A}} = \frac{1-r^{-1}}{1-r^{-N}}, \qquad \rho_{\mathsf{a}} = \frac{1-r}{1-r^N}. \tag{31}$$

As expected, if $r > 1$ then the probability to fi a mutant A-allele increases and the probability to fi a mutant a-allele decreases, and vice versa if $r < 1$.

Diffusion approximation

In the limit $N \to \infty$, $i, j \to \infty$ with $i/N \to y$, $j/N \to x$, if $NP_{ij}(t) \to f(x,t \mid y,0)$, $\lambda_j - \mu_j \to m(x)$ and $\lambda_j + \mu_j \to s(x)$ then equations (19) and (20) become, respectively,

$$\begin{aligned} \frac{\partial}{\partial t}f(x,t\mid y,0) = -\frac{\partial}{\partial x}[m(x)f(x,t\mid y,0)] \\ + \frac{1}{2N}\frac{\partial^2}{\partial x^2}[s(x)f(x,t\mid y,0)], \end{aligned} \tag{32}$$

$$\begin{aligned} \frac{\partial}{\partial t}f(x,t\mid y,0) = -m(y)\frac{\partial}{\partial y}f(x,t\mid y,0) \\ + \frac{s(y)}{2N}\frac{\partial^2}{\partial y^2}f(x,t\mid y,0). \end{aligned} \tag{33}$$

In both cases the initial condition is $f(x, 0 \mid y, 0) = \delta(x - y)$. These are two diffusion equations that describe the Moran process in this special regime. Notice that the diffusion term is proportional to N^{-1}, which means that the deterministic approximation is valid when populations are large. Thus, the finiteness of real populations adds a "noise" to the deterministic behaviour.

The diffusion approximation can also be obtained starting from the Fisher-Wright process. The result is similar but the time-scale is N times faster than in the Moran process. This approximation allows calculations that cannot be achieved for the discrete processes, and it can be easily generalized to situations in which there are multiple alleles or individuals have more than just one gene. Exploring all its possibilities would lead us too far from the scope of this article, so the interested reader is urged to consult the extensive literature on this topic (see, e.g. Refs. [1, 5, 11]).

5 The role of genetic drift in evolution

Genetic drift is present in the first studies on population genetics as an additional mechanism that accounts for the "noise" that small populations introduce in evolutionary dynamics. This noise acquires a special relevance in those times in which species go through evolutionary bottlenecks. However, the breakthroughs achieved in the field of molecular biology have put forward the very relevant role that this mechanism plays even in ordinary situations, so much so that we are at the onset of a change of paradigm: neutral evolution by pure genetic drift seems to be not only the most common way in which life forms evolve but also the mechanism that *de facto* allows them to adapt and speciate. Therefore genetic drift is one of the keys to the origin of species.

But in order to investigate the true implications of neutral evolution we first need to take a small walk through biology.

Biology is built on sequences and networks

At its most fundamental level, life is written in a DNA molecule consisting of a long sequence of four kinds of bases: adenine (A), thymine (T), guanine (G) and cytosine (C). We could say that this sequence of bases is a coding of all information on the building of a cell, and eventually of a life being.[4]

DNA is made of introns (pieces of the chain that do not code for proteins) and exons (the coding elements). In the first stages of transcription DNA transfers its information into RNA molecules (replacing thymine with a new base: uracil, U) in which exons get isolated through a splicing mechanism of the chain and are combined to form genes. These genes are the true pieces of code that translate into proteins in the cell rhybosomes (kind of reading-translating "machines"). At this level the DNA chains that build up chromosomes can be regarded as sequences of genes.

RNA gets transcribed into proteins by translating *codons* (sequences of three consecutive base pairs) into amino acids. This translation forms a universal code[5] known as *genetic code.* Transcription produces a new kind of sequence – proteins, this time made of 20 different types of amino acid. As a consequence of the interaction between amino acids, proteins get folded into three-dimensional structures, sometimes rigid and sometimes including mobile elements, just as if they were a sort of small machines. This three-dimensional structure determines its function insofar as a change of its conformation can make the protein lose its biological function or acquire a new one. The set of proteins of a cell (its proteome) forms a complex network; proteins interact with each other and with genes in very varied ways, activating or inhibiting the production of other proteins, catalyzing reactions, etc. The result forms a metabolic and regulatory network of interactions, a kind of protein ecosystem, whose result is cell activity.

We can still go up the scale and consider multicellular organisms as a new complex network of different types of cells that interact with each other. And in their turn, these organisms (animals, plants, etc.) entangle their life activity, competing for resources, eating each other, cooperating, etc., to give rise to the top biological scale: ecosystems.

All these biological organizations – sequences or networks – share a common property: they are made of a well-defined set of elements whose modification in any way (by changes, eliminations or additions) can induce drastic changes in the whole organization. At the most basic level – DNA chains – these modifications are commonly known as mutations. Mutations can be just replacements of one base by another, or addition or removal of some bases, or more drastic changes like inversions of whole pieces of the DNA sequence, duplications, etc. Drastic mutations are normally lethal: the resulting organism is not viable anymore (imagine, for instance, the effect of removing a single base if we take into account that protein transcription occurs through codon reading). Nonetheless many other mutations can be innocuous, and some of them can even give rise to viable modifications; the genetic code contains 64 different codons that only codify for 20 amino acids – plus a stop sequence; the redundancy is such that there are some amino acids that are coded by up to six different codons. This implies that there will be many base substitutions that have no effect whatsoever on the transcription into proteins. They are therefore innocuous.

Regarding proteins, we have already mentioned that they can fold into three-dimensional (tertiary) structures because of amino acid interactions, and these structures determine their functions. It turns out that most amino acids of the chain have little or no influence in the tertiary structure because the folding is determined by a small set of them, placed at strategic positions. Substituting one of these key amino acids will modify the tertiary structure, hence the protein function. However, substituting any of the other amino acids will either not change the tertiary structure or change it only slightly, so little that the protein maintains its function. This means that even mutations that lead to amino acid substitutions may have no biological effect, enormously enhancing the redundancy already existing in the genetic code.

Also, it often happens that adding, eliminating or replacing a protein in the metabolic and regulatory network of a cell has little influence on the global dynamics of the system. Thus, even at this level some changes are innocuous, hence not subject to selection. And not just at this level – a similar thing happens at the ecological level with the species forming an ecosystem. In summary, many changes can occur at all scales of a biosystem producing little or no effect at all. Selection is thus blind to these changes. But, which fraction of the set of possible changes do they represent?

Most evolution is neutral

In 1968, Kimura surprised the scientif c community with the idea that most genome mutations are neutral [12]. His argument goes as follows. Comparative studies of some proteins lead to the conclusion that chains 100 amino acids long undergo one substitution every 28 million years. The length of DNA chains in the two chromosome sets of mammals is about 4×10^9 base pairs. Every three base pairs code for one amino acid and, due to the existing redundancy, only 80 % of base pair substitutions lead to an amino acid substitution. Therefore there are 16 million substitutions in the whole genome every 28 million years, i.e. almost one substitution every 2 years! Kimura's conclusion is that organisms can only afford such a mutational load provided the great majority of mutations are neutral.

Recent studies on RNA molecules reach similar conclusions [6]. RNA molecules fold as a result of the interaction between the bases forming their sequences. This folding can be regarded as the molecule phenotype because it determines the function of these chains. Thus natural selection acts directly on it being blind to the actual sequences. The number of different sequences folding into the same structure is huge. This implies, once more, that a large number of mutations in the chain leave the phenotype intact, thus avoiding selection.

Neutrality seems the rule rather than the exception, at least at the molecular level. The consequences of this fact are far reaching but to see how much, we need to introduce a new concept: *adaptive landscapes,* and resort again to mathematics.

6 Adaptive landscapes

Perhaps Sewall Wright's most relevant contribution to evolutionary theory is his metaphor of *adaptive landscape* [19]. From a formal viewpoint, an adaptive landscape is a mapping $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X}$ denotes a conf guration space equipped with some notion of adjacency, proximity, distance or accessibility, and whose image is the f tness associated to a particular conf guration in $\mathcal{X}$. This structure, together with the fact that the set of all adaptive landscapes forms the vector space $\mathbb{R}^{|\mathcal{X}|}$, allows the development of a rich theory of combinatorial landscapes that covers not only the adaptive landscapes of biology but also the energy landscapes of physics or the combinatorial optimization problems arising in computer science [17].

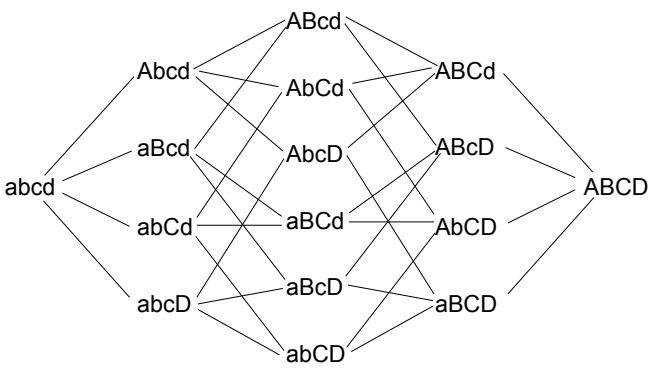

Figure 2. Graph for a sequence of length 4 and 2 alleles per locus

In our particular case, $\mathcal{X}$ is made of the set of underlying sequences or networks. When talking about sequences we will generically use loci to refer to particular genetic traits and alleles to describe the different variants of that trait. Depending on the context, by allele we may refer to bases (DNA), amino acids (proteins) or genes (chromosomes). Therefore, for a DNA chain of length L, $\mathcal{X} = \{\mathrm{A, T, C, G}\}^L$, for a protein of the same length, $\mathcal{X} = \{\mathrm{Phe, Leu, Ile,}\ldots, \mathrm{Gly}\}^L$ and for a chromosome $\mathcal{X} = \{A_1, \ldots, A_{n_a}\} \times \{B_1, \ldots, B_{n_b}\} \times \cdots$, where letters represent the different alleles of a given locus. On $\mathcal{X}$ we can introduce Hamming distance, $d_{\mathrm{H}}(x, y)$, which represents the number of different loci of sequences $x, y \in \mathcal{X}$.

The structure of $\mathcal{X}$ is determined by the allowed transitions between its sequences, which we will refer to generically as mutations. We will speak about point mutations to refer to changes at a given locus of the chain, i.e. substitutions of a base pair in a DNA chain, of an amino acid in a protein or of a different allele in a chromosome. If all mutations are of this type we can build the graph $\mathcal{G} = \{\mathcal{X}, \mathcal{L}\}$, where the set of links $\mathcal{L}$ is made of all pairs of sequences $x, y \in \mathcal{X}$ with $d_{\mathrm{H}}(x, y) = 1$ (Figure 2 illustrates the case for $L = 4$ and two alleles per locus). We can also consider mutations that amount to adding or eliminating a given locus of the sequence. If $\mathcal{X}_L$ denotes the set of sequences of length L then $\mathcal{X} = \bigcup_L \mathcal{X}_L$, and the graph will contain links between sequences of different length.

The probability that a mutation of those def ning $\mathcal{G}$ occurs need not be uniform. The general way of describing the evolution of a given sequence is by introducing a transition probability matrix T whose element T_{xy} is the probability to mutate from one sequence $x \in \mathcal{X}$ to another $y \in \mathcal{X}$. The zeros of the adjacency matrix of the graph $\mathcal{G}$ are also the zeros of matrix T. The evolution of the sequence is therefore the random walk across $\mathcal{X}$ described by the Markov process associated to T.

But we began this section talking about a metaphor. And indeed, beyond the formal description and mathematical treatment of adaptive landscapes, it is the mental picture they provide that leads our intuitions. As a matter of fact, in the development of population genetics there are three metaphors that have been widely used: Fisher's Fujiyama landscape, Wright's rugged landscape and Kimura's f at landscape. Let us examine these three models in more detail.

Fujiyama landscape: quasispecies and the error catastrophe

Fisher imagined that species were in a situation of optimal adaptation to their environment. Hence each species should sit at one of the many tops of the adaptive landscape. According to this picture, a sequence would be maximally adapted and as sequences get away from it in Hamming distance, their f tness should decrease. Fisher did not have in mind sequences when he elaborated this metaphor because molecular biology was in its early stages. It was Eigen [4] who used it to elaborate his theory of *quasispecies* and discover the *error catastrophe.* If the Markov process define on the set of sequences $\mathcal{X}$ is ergodic, there will be a stationary probability distribution. When the landscape is of Fujiyama type, this distribution will be localized around the optimally adapted (or master) sequence. In spite of this being the most probable se-

quence, close to it (i.e. a few mutations away) there will be a "cloud" of less adapted sequences coexisting with the master sequence. Quasispecies is the term that Eigen introduced to describe this ensemble of sequences.

In order to study the behaviour of a quasispecies more closely we will resort to equation (9). For simplicity we shall assume that sequences have finite length $L \gg 1$, that there are just two alleles per locus, that the master sequence has fitness $f_1 = f > 1$ and that all other sequences have fitness $f_2 = \cdots = f_{2^L} = 1$. We shall also assume that the probability that a point mutation occurs is $\mu \ll 1$, independent of the sequence. Let $x_1 = x$ denote the population fraction of the master sequence; thus $x_2 + \cdots + x_{2^L} = 1 - x$ and $\phi = fx + 1 - x$. Equation (9) then becomes

$$\dot{x} = x[f(1-\mu)^L - 1 - (f-1)x] + O(\mu). \qquad (34)$$

The term $O(\mu)$ accounts for the transitions from the L nearest neighbour sequences of the master sequence that revert to the master sequence. Neglecting these terms and approximating $(1-\mu)^L \approx e^{-L\mu}$ we can see that if $fe^{-L\mu} > 1$ then x asymptotically approaches $x^* = (e^{-L\mu}f - 1)/(f-1)$, whereas if $fe^{-L\mu} < 1$ the bracket in equation (34) becomes negative and therefore $x = O(\mu)$. The threshold $\mu_{\text{err}} = \log f/L$ defines the error catastrophe. When $\mu < \mu_{\text{err}}$ the quasispecies is well-defined because the master sequence is the most probable one. However, when $\mu > \mu_{\text{err}}$ the identity of this master sequence gets lost in the cloud of mutants and the quasispecies disappears as such.

Experimental studies performed in the '90s seem to confirm [16] that indeed the length of the genome of different species – ranging from virus to Homo sapiens – and the mutation rate per base are related as $\mu L \leq O(1)$. Hence an increase in the mutation rate is a mechanism that this theory puts forward to fight viral infections. We will come back to this point later.

Rugged landscape

Although locally the adaptive landscape can be well described by the Fujiyama model, Wright visualized it as a rugged landscape, full of high peaks separated by deep valleys. The reason is that mutations that change the sequence minimally may induce large variations in the fitness of individuals. In addition, there exists the well known phenomenon of *epistasis,* according to which some genes interact, either constructively or destructively, amplifying these large variations in response to small changes in the sequences.

According to the rugged landscape metaphor, species evolve by climbing peaks and sitting on the summits. Different peaks correspond to different species with different fitnesses. This picture seems to fit well with our idea of evolution by natural selection. However, it has a serious drawback: species that are at a summit can only move to a higher one by going through an unfit valley. In the most favourable case this valley will consist of a single intermediate state. Formula (31) tells us that if a population is small, it is not impossible that an unfit allele replaces a fitter one. Nevertheless, the probability that this happens is very small, i.e. adaptation times should be very large. And this is only the most favourable case. The high speed of adaptation to rapidly changing environments that viruses exhibit seriously challenges this model. What is then wrong in our picture of adaptive landscapes?

Holey landscape: neutral networks

Let us review the most extreme case of a rugged landscape: the random landscape. In this case every sequence of $\mathcal{X}$ has a random fitness, independent of the other sequences. In general, rugged landscapes are not that extreme because there is some degree of correlation between the fitness of neighbouring sequences. However, beyond the correlation length, fitness values become uncorrelated. The random landscape is the extreme case in which the correlation length is smaller than 1. Suppose now that the length of the sequences is large, and that every locus can host A independent alleles. The degree of graph $\mathcal{G}$ will thus be $g = (A-1)L$ and its size $|\mathcal{X}| = A^L$. With $L = 100$ and $A = 2$ (a rather modest choice), $g = 100$ and $|\mathcal{X}| = 2^{100} \approx 10^{30}$. To all purposes such a graph can be locally approximated by a tree, the more so the larger the degree (see Figure 3). Imagine an extreme assignment of fitness: 1 if the sequence is viable and 0 if it is not. Let p be the fraction of viable sequences. Evolution can only proceed by jumping between consecutive viable nodes. According to Figure 3, which illustrates what this landscape looks like locally in a particular graph, it becomes clear that if p is small, the number of viable nodes a distance d apart from the initial node is well approximated by a branching process where, except for the first generation, the number of offspring (viable nodes) is given by $p_k = \binom{g-1}{k} p^k (1-p)^{g-1-k}$, with an expected value of $(g-1)p$. The theory of branching processes [10] tells us that, with a finite probability, the process never ends provided $(g-1)p > 1$. Translated to our graphs this implies that whenever $p \gtrsim 1/g$ (with $g \gg 1$) there is a connected subgraph of viable nodes containing a finite fraction of all nodes of $\mathcal{G}$. This kind of subgraph is called a *neutral network* [7].

If we consider a more general model in which $\mathcal{P}(f)$ describes the probability density that a node fitness is between f and $f + df$, if $\int_{f_1}^{f_2} \mathcal{P}(f)\, df \gtrsim 1/g$ then there will be a quasineutral network whose node fitnesses will all lie in the interval (f_1, f_2). As g is usually very large (proportional to the sequence length), the existence of neutral networks becomes

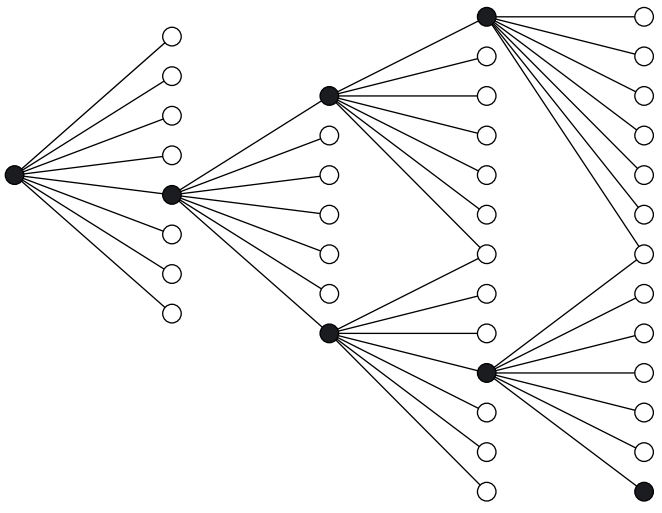

Figure 3. Local section of a configuration graph with $A = 2$ and $L = 8$. Black nodes are viable, whereas white nodes are not viable.

the rule rather than the exception. Wright's metaphor of a landscape full of mountains and valleys becomes utterly inappropriate to describe the existence of these neutral networks in rugged adaptive landscapes. A more appropriate metaphor would be that of a f at landscape (*á la* Kimura) with holes. Evolution moves sequences across this neutral network, transforming them into completely different sequences without ever decreasing their f tness. Undoubtedly, this mechanism dramatically speeds up not only adaptation of species to the environment but even speciation.

The fact that the f tness depends on phenotype and not on genotype favours the appearance of neutral networks. This is what is observed in RNA [6]. The properties of these networks have an enormous inf uence in evolutionary dynamics, an inf uence that we are now only beginning to understand. Just as an example, if we reconsider Eigen's model under the viewpoint of this new metaphor, we will realize that its main hypothesis, namely that locally the landscape is Fujiyama, is completely wrong. There is no such a thing as a master sequence. Instead there is a master network (or phenotype) that contains a huge number of sequences. Accordingly, the probability that a mutation recovers the optimal fit ess is much larger than what Eigen's theory assumes because it can be recovered by hitting any of the sequences of the network, not necessarily the initial one. When this probability is not negligible the error catastrophe goes away [14].

7 Conclusion

This article tries to provide an overview of the contributions of mathematics to evolutionary theory. From population genetics to the theory of complex networks, going through the theory of stochastic processes, many relevant results about the evolutionary mechanisms driving life have been obtained thanks to their mathematical descriptions. We still cannot say that the theory of evolution is a fully mathematically formulated scientifi theory, like Darwin would have liked it to be, but it is unquestionable that we are getting closer and closer to such an achievement. Nowadays we could say that the theoretical studies of evolutionary processes are at least as important as the experimental ones and that, as the opening sentence by Darwin states, it is those that shed light in the darkness.

Given the divulgatory nature of this article many interesting topics have been left out. Some of them provide new insights into evolutionary mechanisms and some illustrate further contributions of mathematics to evolutionary theory. Among them we can mention the infi ite allele model [5], which is currently employed in analysing evolutionary divergence of DNA or protein sequences, or the coalescent process [5], which is an interesting and practical backward formulation of genetic drift. We have not mentioned the important contributions of game theory to evolution either. This theory is currently being used to deal with the problem of the evolutionary emergence of cooperation [16]. Instead, the focus has been on the subject of adaptive landscapes and neutral networks because, in the author's opinion, this is the area where a new reformulation of the evolutionary paradigm can emerge in the following years. Understanding evolution requires, against all expectations, an understanding of the role of genetic drift on neutral networks. And due to the complexity of this problem, it is one of the topics in which mathematics can be our eyes in the dark room that help us fin the black cat that is not there.

I want to acknowledge Susanna Manrubia for her critical reading and her valuable remarks. This work is supported by projects MOSAICO (Ministerio de Educación y Ciencia, Spain) and MODELICO-CM (Comunidad de Madrid, Spain).

This article is a translation of an article originally published (in Spanish) in *Gaceta de la RSME*, Vol. 12, no. 4, 2009, pp. 667–686. Published with permission.

Notes

1. Admiration not unrelated to Darwin's care for his progeny: Sir George Howard Darwin, the f fth of Darwin's children, became an astronomer and mathematician.
2. Human beings have so far not reached this equilibrium with the environment because this environment is the whole planet. In spite of that, the law provides a reasonable description of isolated populations in low resource environments. The permanent famine suffered by many African countries provides an illustration of what our situation will be when we reach that equilibrium with the resources of the planet.
3. It is most remarkable that de Vries rediscovered Mendel's laws for the scientif c world in 1900, the same year that Planck proposed the quantum hypothesis for the f rst time.
4. This classic dogma is not quite true because in cell division every daughter cell inherits not only an exact copy of the parent cell's DNA but also half of its cytoplasm. This makes the two daughter cells slightly different in composition and this difference induces different gene expressions. This lies at the heart of cell differentiation and becomes the core of what is currently known as *epigenetics*.
5. In 1979 it was discovered that this code is not quite universal: mitochondria, as well as some bacteria and yeasts, use codes that differ slightly from the "universal" one.

References

[1] R. A. Blythe and A. J. McKane, Stochastic Models of Evolution in Genetics, Ecology and Linguistics, *J. Stat. Mech.* **P07018** (2007), 1–58.

[2] M. Bulmer, Galton's law of ancestral heredity, *Heredity* **81** (1998), 579–585.

[3] C. R. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, John Murray, 1859. Available at http://darwin-online.org.uk/

[4] M. Eigen, Self-organization of Matter and the Evolution of Biological Macromolecules, *Naturwissenschaften* **58** (1971), 465–523.

[5] W. J. Ewens, *Mathematical Population Genetics. I. Theoretical Introduction*, 2.ª ed., Springer, 2004.

[6] W. Fontana, Modelling 'evo-devo' with RNA, *BioEssays* **24** (2002), 1164–1177.

[7] S. Gavrilets, *Fitness Landscapes and the Origin of Species*, Princeton University Press, 2004.

[8] J. Hofbauer and K. Sigmund, *Evolutionary Games and Replicator Dynamics*, Cambridge University Press, 1998.

[9] S. Karlin and J. McGregor, On a genetic model of Moran, *Proc. Cambridge Philos. Soc.* **58** (1962), 299–311.

[10] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2.ª ed., Academic Press, 1975.

[11] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes*, Academic Press, 1981.

[12] M. Kimura, Evolutionary rate at the molecular level, *Nature* **217** (1968), 624–626.
[13] T. R. Malthus, *An Essay on the Principle of Population*, J. Johnson, 1798. Available at http://www.econlib.org/library/Malthus/malPopCover.html
[14] S. C. Manrubia, E. Domingo, and E. Lázaro, Pathways to extinction: beyond the error threshold, *Phil. Trans. R. Soc. B,* in press DOI: 10.1098/rstb.2010.0076
[15] J. G. Mendel, Versuche über Pf anzenhybriden, *Verhandlungen des naturforschenden Vereines in Brünn* **IV/1865** (1866), 3–47. English translation available at http://www.esp.org/foundations/genetics/classical/gm-65.pdf
[16] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life*, Belknap Press, 2006.
[17] C. M. Reidys and P. F. Stadler, Combinatorial landscapes, *SIAM Rev.* **44** (2002), 3–54.
[18] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer, 2006.
[19] S. Wright, The roles of mutation, inbreeding, crossbreeding and selection in evolution, *Proceedings of the Sixth International Congress on Genetics* (Austin, Texas, 1932), Vol. I, 356–366. Available at http://www.blackwellpublishing.com/ridley/classictexts/wright.pdf

José Cuesta [cuesta@math.uc3m.es] was born in Girona, Spain, in 1964 and moved two years later to Madrid, where he has been living since then. He graduated in physics from the Universidad Complutense de Madrid in 1987 and got a physics PhD at the same university in 1992. Since then he has worked in the mathematics department of the Universidad Carlos III de Madrid, where he was appointed as an associate professor in 1995. His research f eld is statistical physics and his current interests focus on the theory of complex systems, both biologically and socially motivated, and on evolutionary dynamics, very much related to the former. He is a member of the Interdisciplinary Group of Complex Systems (GISC), a research group that has brought together people from departments of physics and mathematics at several universities with an interest in the physics of complex systems.