Bull Math Biol (2015) 77:2354–2365
DOI 10.1007/s11538-015-0127-z

ORIGINAL ARTICLE

A Stochastic Model for the Interbreeding of Two Populations Continuously Sharing the Same Habitat

Maurizio Serva[1]

Received: 20 July 2015 / Accepted: 12 November 2015 / Published online: 19 November 2015

Abstract We propose and solve a stochastic mathematical model of general applicability to interbreeding populations which share the same habitat. Resources are limited so that the total population size is fixed by environmental factors. Interbreeding occurs during all the time of coexistence until one of the two population disappears by a random fluctuation. None of the two populations has a selective advantage. We answer the following questions: How long the two populations coexist and how genetically similar they become before the extinction of one of the two? how much the genetic makeup of the surviving population changes by the contribution of the disappearing one? what it is the number of interbreeding events given the observed introgression of genetic material? The model was originally motivated by a paleoanthropological problem concerning the interbreeding of Neanderthals and African modern humans in Middle East which is responsible for the fraction of Neanderthal genes (1–4 %) in present Eurasian population.

Keywords Interbreeding · Wright–Fisher diffusion · Stochastic trajectories · Ito calculus

1 Introduction

The present paper concerns a stochastic mathematical model of general applicability to populations which share the same habitat and which may interbreed; nevertheless,

✉ Maurizio Serva
serva@univaq.it

[1] Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica, Università dell'Aquila, L'Aquila, Italy

it was introduced for the first time in Neves and Serva (2012) to give an answer to a specific paleoanthropological problem.

Until recently, the majority of the scientific community favored a scenario where anatomically modern Africans completely replaced archaic populations in other continents. Things changed radically in 2010 with a paper by Green et al. (2010) which provided the first indisputable evidence of African/Neanderthal interbreeding which likely occurred in the Middle East where the two population cohabited for a long time. This new scenario prompted our model Neves and Serva (2012) which was mainly conceived to give an answer to the following question: Given the fraction of Neanderthal genes (1–4 %) in present Eurasian population, how many successful mating events have occurred?

Later on, evidence of interbreeding of modern Africans emigrants with other archaic populations was also clearly highlighted. It was found that 4–6 % of the genome in Melanesians and Australians derives from Denisovans Reich et al. (2010) and that Denisovan genome can be also detected in other populations at a limited extent. Moreover, roughly 2 % of the genetic material found in some sub-Saharan African populations was inserted into the human genome approximately 35,000 years ago from archaic hominids that broke away from the modern human lineage around 700,000 years ago Hammer et al. (2011).

Therefore, while we will refer to the African/Neanderthal interbreeding for reference, in principle the model apply also to these events as well as to analogous phenomena in the animal world.

Although Neanderthals disappeared approximately 40,000 years ago, they contributed 1–4 % of the DNA of all living humans with the exception of sub-Saharan populations. Since the proportion of Neanderthal genes is the same in Europeans and Asians, it was suggested Green et al. (2010) that interbreeding might have occurred in the Middle East before expansion of Africans into Eurasia, at a time in which both populations coexisted there. Indeed, according to Bar-Yosef (1998), the Skhul and Kafzeh caves in Israel have been occupied both by anatomically modern Africans and by Neanderthals, changing hands between one group and the other several times over a period of thousands of years.

Conversely, recently it was found that 6–9 % of the genome of a 37,000- to 42,000-year-old modern individual from Oase (Romania) is of Neanderthal origin Fu et al. (2015). Nevertheless, it seems that the Oase population did not contribute substantially to later humans in Europe.

Although we do not know exactly when and for how long the two groups interacted, it seems reasonable that some interaction did occur in the Middle East and that it may have lasted for a long time, until one of the two populations went extinct there. On the contrary in Sankararaman et al. (2012), it is assumed that Middle East interbreeding was limited in time and it was probably an exceptional event. This seems in contradiction with Fu et al. (2015) which suggests that interbreeding occurred likely many times in different places. In fact, given the limited amount of available genetic material, the detection of a second exceptional interbreeding event with no output in present population, is very unlikely. For this reason, we posit that interbreeding occurred whenever the two populations where in contact.

In our model, we assume that the two populations share the same habitat with limited resources so that the total population size is fixed by environmental factors. We also assume that interbreeding occurs during all the time of coexistence until one of the two population extinguishes and that none of the two populations is selectively favored with respect to the other.

Assuming that the size of the total population is known as well as the fraction of the loser population genes in the winner (1–4 % for the Neanderthal case), we answer the following questions: How long the two populations coexist and how genetically similar they become before the extinction of one of the two? How much the genetic makeup of the surviving population changes by the contribution of the disappearing one? What it is the number of interbreeding events given the observed introgression of genetic material?

The model was initially proposed in Neves and Serva (2012) where, by means of numerical simulations of the stochastic dynamics, we gave partial answers to those questions. In this paper, we give a complete answer solving the stochastic model and applying the results, as an example, to the Neanderthal/African case.

2 The Model

Starting at generation $g = 0$, the two populations meet and start to share a common environment. The model consists of the following three stochastic steps:

2.1 Sizes of the Populations

We assume that the size of the meta-population (composed by the two interbreeding populations) is constant and equal to N being fixed by environmental factors (limited resources). The sizes $N_A(g)$ and $N_B(g)$ of the two component populations change at any generation g although they are constrained by $N_A(g) + N_B(g) = N$.

If one assumes that the average number of offspring is the same for individuals of the two populations (none of the two has a selective advantage) the dynamics, when N is large, is governed by the well-known *neutral* Wright–Fisher diffusion Serva (2005), Traulsen et al. (2005) and Durrett (2008) independently of the microscopic rule one may choose. Since the random number of generations until extinction of one population is proportional to N, one has to use *time* defined as $t = g/N$ instead of generation g and also one has to use the fractions $x(t) = N_A(g)/N$ and $1 - x(t) = N_B(g)/N$.

In this way, the large N limit Serva (2005) can be correctly performed and one gets the Wright–Fisher diffusion

$$\mathrm{d}x(t) = \sqrt{x(t)(1 - x(t))}\,\mathrm{d}w(t) \tag{1}$$

where $w(t)$ is a standard Brownian motion. The initial conditions is $x(0) = x_0$ which is the initial fraction of individuals of population A.

We remark that a Wright–Fisher trajectory always ends in a random finite time τ in one of the two frontiers in 0 or in 1. This means that, although the model is neutral

(individual of both populations have the same fitness), random fluctuations always wipe out one of the two populations.

2.2 Mating

At each generation a number α of individuals from population A migrate to population B and vice-versa, i.e., the same number of individuals from population B migrate to population A. In other words, α *pairs* of individuals per generation are exchanged, passing from one population to the other. Once they are exchanged, they become part of the host population, and they mate with its members and their genes introgress.

If α is smaller than 1, the fraction $1/\alpha$ can be interpreted as the number of generations between the exchange of a pair of individuals.

Let us considers only those alleles which, before mating, were different for the two populations and assume that their number is large. Let us call type A (B) those alleles which, before mating, were typical of individual of the population A (B). Therefore, by our isolation assumption, at $g = 0$ all individuals of population A have a fraction equal to 1 of type A alleles and all individuals of population B have a fraction 0 of such alleles.

At any time $t > 0$ any individual will be characterized by his/her fractions of type A alleles and, since the fraction of type A alleles in an individual is the average of its parent's fractions, given interbreeding, it will be different from the initial value.

We define then $z_A(t)$ as the *mean fraction of type A alleles in population A at time t* and $z_B(t)$ *as the mean fraction of type A alleles in population B at time t*. The isolation assumption implies $z_A(0) = 1$ and $z_B(0) = 0$.

The mean allelic fraction $z_A(t + 1/N)$ equals $z_A(t)$ plus the contribution of type A alleles from the immigrating individuals of population B and minus the loss of type A alleles due to emigration. These loss and gain terms are both proportional to α and inversely proportional to the number $Nx(t)$ of individuals in population A. Similar considerations apply to $z_B(t + 1/N)$.

In symbols:

$$\begin{cases} z_A\left(t+\frac{1}{N}\right) = \left(1-\frac{\alpha}{Nx(t)}\right) z_A(t) + \frac{\alpha}{Nx(t)} z_B(t) \\ z_B\left(t+\frac{1}{N}\right) = \frac{\alpha}{N(1-x(t))} z_A(t) + \left(1-\frac{\alpha}{N(1-x(t))}\right) z_B(t). \end{cases} \tag{2}$$

In deriving the above equations, we implicitly made the assumption that all exchanged individuals have the same mean allelic fractions, i.e., that one can neglect the fluctuations in genomic composition of the cross-breeding individuals, because each subpopulation reaches full genetic mixing in isolation between one cross-breeding event and the next one. This is a very strong assumption, and it is not strictly true, but it is a very good approximation if the interbreeding rate α is small. To be more precise, the assumption is substantially correct if α is smaller than $1/\log_2 N$. In fact, $1/\alpha$ is the mean number of generations between two cross-breeding events and $\log_2 N$ is the typical number of generations for genetic homogenization in a population of N individuals with diploid reproduction and random mating Derrida et al. (1999) and

Chang (1999). Therefore, the condition that α is smaller than $1/\log_2 N$ makes sure that both subpopulations are homogeneous at the exchange times.

The above equations, after taking the $N \to \infty$ limit, become a system of linear ordinary differential equations

$$\begin{cases} \frac{\mathrm{d}z_A(t)}{\mathrm{d}t} = -\frac{\alpha}{x(t)}\,(z_A(t) - z_B(t)) \\ \frac{\mathrm{d}z_B(t)}{\mathrm{d}t} = \frac{\alpha}{1-x(t)}\,(z_A(t) - z_B(t)) \end{cases} \tag{3}$$

which should be solved given a solution $x(t)$ of the stochastic differential Eq. (1) and with initial conditions $z_A(0) = 1$, $z_B(0) = 0$.

Ultimately, the model consists in three coupled equations: the two ordinary Eq. (3) and the stochastic Eq. (1).

3 Formal Solution and Goals

By introducing the auxiliary functions $u(t) = [z_A(t) - z_B(t)]/2$ and $v(t) = [z_A(t) + z_B(t)]/2$ we obtain the equations

$$\begin{cases} \frac{\mathrm{d}u(t)}{\mathrm{d}t} = -\alpha \frac{1}{x(t)(1-x(t))}\,u(t) \\ \frac{\mathrm{d}v(t)}{\mathrm{d}t} = \alpha \frac{2x(t)-1}{x(t)(1-x(t))}\,u(t) \end{cases} \tag{4}$$

which replace Eq. (3) and should be solved according to the initial conditions $u(0) = v(0) = 1/2$. They have the formal solution for $t \le \tau$

$$u(t) = \frac{1}{2}\exp\left(-\alpha\int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right), \quad v(t) = u(t) + \alpha\int_0^t \frac{2}{1-x(s)}\,u(s)\,\mathrm{d}s \tag{5}$$

where $x(t)$ is a realization of the Wright–Fisher process (1) with initial condition $x(0) = x_0$.

The function $u(t)$, which starts at $u(0) = 1/2$ and later decreases, measures the genetic difference between the two populations, while the function $v(t)$ is the average fraction of genes of type A over both populations

The evolution (5) ends at the final random time τ when the Wright–Fisher trajectory hits one of the two boundaries. The values of $z_A(\tau) = v(\tau) + u(\tau)$ and $z_B(\tau) = v(\tau) - u(\tau)$ depend on all the Wright–Fisher (random) trajectory.

Our aim is:

- to show that population A wins with probability x_0, which is a trivial and well-known consequence of (1) and which we re-obtain here for the sake of completeness;
- to show that $u(\tau) = 0$ which implies $z_A(\tau) = z_B(\tau) = v(\tau)$. This has the important meaning that the two populations become genetically equal just before one of the two extinguishes and, consequently, that $v(\tau)$ is the average fraction of genes of type A in the winner population;

- to find the average $E[v(\tau)]$ which depends on x_0 but not on α. This is not astonishing since the final average genetic makeup of the population only depends on its initial composition and not on the interbreeding rate.

Then after having defined $\tilde{v}(t) = x(t)v(t) + (1-x(t))(1-v(t))$, one can easily get convinced that $\tilde{v}(\tau)$ is the fraction of winner genes in the winner population (which is *a priori* unknown). In fact, if population A wins one has $x(\tau) = 1$ and $\tilde{v}(\tau)$ coincides with $v(\tau)$, vice-versa, if population B wins one has $x(\tau) = 0$ and $\tilde{v}(\tau)$ coincides with $1 - v(\tau)$. Therefore, $\tilde{v}(\tau)$ is the quantity that should be compared with the fraction 0.96–0.99 of African genes in Eurasian population.

Thus, our final goal is:

- to compute the average $E[\tilde{v}(\tau)]$ which depends both on x_0 and α;
- to infer the value of α, as a function of x_0 and of the observed value of $\tilde{v}(\tau)$ (0.96–0.99 in the example case). This is the main goal of present paper, and given that the result is weakly dependent on x_0, the known value of $\tilde{v}(\tau)$ alone allows us to infer the rate of successful Neanderthal/African interbreeding events.

4 Genetic Convergence ad Average Final Genetic Composition

Consider the Wright–Fisher Eq. (1), one has

$$\frac{d}{\mathrm{d}t}E[x(t)] = 0 \quad \rightarrow \quad E[x(\tau)] = x_0 \tag{6}$$

Since we know that at $t = \tau$ all realizations $x(t)$ have reached one of the barriers, i.e., $x(\tau) = 1$ with probability p and $x(\tau) = 0$ with probability $1 - p$. Therefore, the above equality implies $p = x_0$ as it is well-known. In other words, population A survives with probability x_0 and population B survives with probability $1 - x_0$.

Consider again the Wright–Fisher Eq. (1) and define the variable

$$\gamma(t) = g(x(t)) \exp\left(-\int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right), \tag{7}$$

where $g(x)$ is a function to be determined later. Using Ito calculus, one has

$$d\gamma(t) = \gamma \frac{g'(x(t))}{g(x(t))}\mathrm{d}x(t) + \gamma\left[x(t)(1-x(t))\frac{g''(x(t))}{2g(x(t))} - \frac{1}{x(t)(1-x(t))}\right]\mathrm{d}t \tag{8}$$

where $g' = dg/\mathrm{d}x$ and $g'' = d^2g/\mathrm{d}x^2$. Since $E[\mathrm{d}x(t)] = 0$, $\gamma(t)$ has a constant expected value if $g(x)$ is a solution of

$$x^2(1-x)^2\, g''(x) = 2g(x). \tag{9}$$

As we will see, we need that the solution diverges at the boundaries $x = 0$ and $x = 1$, so that among its solutions we choose:

$$g(x) = [x(1-x)]^{-1} - 3 \tag{10}$$

which is strictly positive for $x \in [0, 1]$ since $x(1-x) \leq 1/4$. According to our choice, $\gamma(t)$ has a constant expected value for all $t \in [0, \tau]$, so that

$$E\left[g(x(t))\,\exp\left(-\int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right)\right] = g(x_0) \tag{11}$$

where for all non-trivial values $x_0 \in (0, 1)$ one has that the right-hand side is finite and strictly positive. Therefore, since $g(x(t))$ diverges for all trajectories in the limit $t \to \tau$, we must have

$$\int_0^\tau \frac{\mathrm{d}s}{x(s)(1-x(s))} = \infty \tag{12}$$

for almost all realizations of $x(s)$.

The immediate important consequence is

$$u(\tau) = \frac{1}{2}\exp\left(-\alpha\int_0^\tau \frac{\mathrm{d}s}{x(s)(1-x(s))}\right) = 0 \tag{13}$$

for almost all realizations of the process $x(t)$ which, in turn, implies that $z_A(\tau) = z_B(\tau) = v(\tau)$, i.e., the genetic difference between the two population vanishes at the moment that one of the two extinguishes.

This phenomenon can be understood as follows: When the two populations have a comparable size, their genetic makeup converges in a symmetrical way, but when a population is much larger then the other, the effect on the genetic composition of the smaller is very high since the number of exchanged individuals is the same. At the end of the process, the effect is so high that the extinguishing population rapidly genetically converges to the surviving one.

Indeed, the exact result depends on our continuous approach, In a real case the meta-population size N may be large but not infinite; therefore, all intensive quantities $x(t)$, $v(t)$, $u(t)$ only may assume a finite set of values. Thus, the last survivors of the extinguishing population, although very similar, will be not completely identical to the winners.

The result in (13) also implies the following useful equality:

$$\int_0^\tau \frac{\alpha}{x(t)(1-x(t))}\,\exp\left(-\alpha\int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right)\mathrm{d}t = 1 \tag{14}$$

which can be easily understood if one considers that the function in the integral is the time derivative of $-2u(t)$.

We have shown that at the final time the genetic composition of the two populations coincides ($u(\tau) = 0$), then we have from (5)

$$v(\tau) = \alpha\int_0^\tau \frac{x(t)}{x(t)(1-x(t))}\,\exp\left(-\alpha\int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right)\mathrm{d}t \tag{15}$$

which, given that $u(\tau) = 0$, can be transformed by integration by parts in

$$v(\tau) = x_0 + \int_0^\tau \exp\left(-\alpha \int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right) \mathrm{d}x(t) \tag{16}$$

where, according to (1), $\mathrm{d}x(t) = \sqrt{x(t)(1-x(t))}\mathrm{d}w(t)$. Taking the expectation, we immediately obtain

$$E[v(\tau)] = x_0, \tag{17}$$

This is not surprising since the final fraction of population A genes in the surviving population must equal the initial proportion x_0 of population A individuals.

5 Genetic Composition of the Surviving Population and Rate of Interbreeding

The fact that the expectation (17) does not depend on α gives the illusion that the final average fraction of genes of type A is independent from the interbreeding rate. This is, indeed, an illusion since the quantity we are interested in is the fraction of winner genes in the winner population. In other words, we are interested in $\tilde{v}(\tau)$, in fact, if population A wins one has $x(\tau) = 1$ and $\tilde{v}(\tau)$ coincides with $v(\tau)$, vice versa, if population B wins one has $x(\tau) = 0$ and $\tilde{v}(\tau)$ coincides with $1 - v(\tau)$.

Let us consider this quantity at a generic time:

$$\tilde{v}(t) = x(t)v(t) + (1-x(t))(1-v(t)), \tag{18}$$

the initial condition is $\tilde{v}(0) = 1/2$ and the final ($t = \tau$) distribution and average both depend on α and x_0.

The differential of $\tilde{v}(t)$ is

$$\mathrm{d}\tilde{v}(t) = (2v(t)-1)\mathrm{d}x(t) + (2x(t)-1)\mathrm{d}v(t). \tag{19}$$

Then, using (4) one immediately gets

$$\mathrm{d}\tilde{v}(t) = (2v(t)-1)\mathrm{d}x(t) + \alpha \frac{(1-2x(t))^2}{x(t)(1-x(t))} u(t)\mathrm{d}t. \tag{20}$$

We integrate now this last expression, and using the explicit expression for $u(t)$, the initial condition $\tilde{v}(0) = 1/2$ and the equality (14), we get:

$$\tilde{v}(\tau) = 1 - 2\alpha \int_0^\tau \exp\left(-\alpha \int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right) \mathrm{d}t + \int_0^\tau (2v(t)-1)\mathrm{d}x(t) \tag{21}$$

which is the fraction of winner genes in the winner population. Notice that the winner population is not known a priori, but it is decided by the random output of the Wright–Fisher dynamics.

Afterward, one can take the expectation and, given (1), obtain

$$E\left[\tilde{v}(\tau)\right] = 1 - 2\alpha\, E\left[\int_0^\tau \exp\left(-\alpha \int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right) \mathrm{d}t\right] \tag{22}$$

In order to compute the average $E[\tilde{v}(\tau)]$, let us consider a generic test function $h(x)$ with $h(0) = h(1) = 0$. Ito calculus gives: $\mathrm{d}h(x(s)) = h'(x(s))\mathrm{d}x(s) + x(t)(1 - x(t))[h''(x(s))/2]\mathrm{d}s$ where $h' = \mathrm{d}h/\mathrm{d}x$ and $h'' = d^2h/\mathrm{d}x^2$, then integration by parts gives

$$\begin{aligned} &\alpha\, E\left[\int_0^\tau \frac{h(x(t))}{x(t)(1-x(t))} \exp\left(-\alpha \int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right) \mathrm{d}t\right] \\ &= h(x_0) + \frac{1}{2} E\left[\int_0^\tau x(t)(1-x(t))h''(x(t)) \exp\left(-\alpha \int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right) \mathrm{d}t\right] \end{aligned} \tag{23}$$

Assume now that $h(x)$ is the solution of the equation

$$2\alpha h(x) - [x(1-x)]^2 h''(x) = 2\alpha x(1-x) \tag{24}$$

with boundary conditions $h(0) = h(1) = 0$. In this case, Eq. (23) rewrites:

$$\alpha\, E\left[\int_0^\tau \exp\left(-\alpha \int_0^t \frac{\mathrm{d}s}{x(s)(1-x(s))}\right) \mathrm{d}t\right] = h(x_0) \tag{25}$$

In conclusion, from (22) we have

$$E[\tilde{v}(\tau)] = 1 - 2h(x_0) \tag{26}$$

where $h(x)$ is the solution of (24) with boundary conditions $h(0) = h(1) = 0$.

We were unable to solve (24) in the general case, but it can be easily used for expansion for small α and for large α. For example, assuming that $h(x) = a_0(x) + \frac{1}{\alpha}a_1(x) + o(\frac{1}{\alpha})$ we easily get from (24) and (26)

$$E[\tilde{v}(\tau)] = 1 - 2x_0(1-x_0) + \frac{2}{\alpha}[x_0(1-x_0)]^2 + o\left(\frac{1}{\alpha}\right) \tag{27}$$

which is the expansion for large value of α.

In particular, when $\alpha \to \infty$, this result simply states that in extreme interbreeding case the two populations mix immediately and completely so that the fraction of genes of type A is x_0 in population A and the of genes of type B is $1 - x_0$ in population B. Then, population A wins with probability x_0 and population B with probability $1 - x_0$ so that $\lim_{\alpha\to\infty} E[\tilde{v}(\tau)] = x_0x_0 + (1 - x_0)(1 - x_0) = 1 - 2\,x_0(1 - x_0)$.

On the contrary, for small value of α we may assume $h(x) = b_0(x) + \alpha b_1(x) + o(\alpha)$ and solve Eq. (24) (with border conditions $h(0) = h(1) = 0$). Using also (26), we easily get

$$E[\tilde{v}(\tau)] = 1 + 4\alpha[x_0 \ln x_0 + (1 - x_0)\ln(1 - x_0)] + o(\alpha) \tag{28}$$

In particular, when $\alpha \to 0$, this result simply states that in absence of interbreeding populations do not mix so that the final fraction of genes of type A is 1 in population A and the final fraction of genes of type B is also 1 in population B, so that $\lim_{\alpha\to 0} \tilde{v}(\tau) = x_0 + (1 - x_0) = 1$.

We can find again the result (28) in the form of an inequality. From (22), we immediately have

$$E[\tilde{v}(\tau)] \geq 1 - 2\alpha\, E[\tau] \tag{29}$$

which is obtained replacing the exponential in the integral by unity. This estimate holds for any value of α but it is accurate only for small α. The random time τ, as usual, is the exit time of the simple Wright–Fisher diffusion starting in x_0. Since the expected value $E[\tau]$ equals $-2x_0 \ln x_0 - 2(1 - x_0)\ln(1 - x_0)$ one also has

$$E[\tilde{v}(\tau)] \geq 1 + 4\alpha[x_0 \ln x_0 + (1 - x_0)\ln(1 - x_0)] \tag{30}$$

which is in agreement with (28).

If we need to estimate the value of α for the Modern/Neanderthal case ($\tilde{v}(\tau) \simeq 1 - 0.025$), we may assume that α is small but we face the problem that the initial x_0 is unknown. Nevertheless, the estimates weakly depends on x_0 if the initial relative sizes of the populations were not extremely different. Assuming that the final genetic composition corresponds to $\tilde{v}(\tau) \simeq 1 - 0.025$ (Modern/Neanderthal), one can simply use (30) and get

$$[-4x_0 \ln x_0 - 4\,(1-x_0)\ln(1 - x_0)]\alpha \simeq 0.025 \quad \to \quad 1/110 \leq \alpha \leq 1/80 \tag{31}$$

which is obtained choosing $0.2 \leq x_0 \leq 0.8$.

The interbreeding rate is extremely small, since about one pair of individuals is exchanged by the two populations every one hundred generations.

In order to obtain this result, we compared the average $E[\tilde{v}(\tau)]$ (which depends strongly on α and weakly on x_0) with the observed data and we extrapolated α. One could argue that what one observes is the outcome of a single realization and not an average. In some cases, a more correct procedure would be to compare the observed data with the value of $\tilde{v}^*(\tau)$ corresponding to the maximum of the probability density of $\tilde{v}(\tau)$. In any case, the best one could do is to determine the distribution of $\tilde{v}(\tau)$ (technically almost impossible) in order to assign a probability density to each value of α (given x_0). The true value of α, determined by the single realization, would anyway remain unknown. Nevertheless, the distribution of $\tilde{v}(\tau)$ is continuous and bounded; therefore, we may assume that the average is close or, at least. of the same order of magnitude of the single and, unfortunately, unrepeatable realization.

Finally, we would like to mention that one of the authors (Neves) of the original work Neves and Serva (2012) also considered the case in which one of the two populations had a larger fitnesses Neves (2012). He concluded that even a very small fitness difference drastically reduces the time of cohabitation of the two populations. Concerning our African/Neanderthal example, this result, as highlighted by Neves,

seems incompatible with the known data. From a technical point of view, he had to introduce two new parameters and all main equations resulted modified. A mathematical treatment of this more general case, on the same ground of the treatment of the neutral case in this paper, would need a new study.

6 Conclusions

We have introduced a stochastic mathematical model of general applicability to populations which share the same habitat with limited resources, which may mate and with no selective advantage of one of the two.

Our main results are: (1) We have shown that just before extinction, the genetic makeup of the two populations become identical, (2) We have computed the average fraction of succumbing population genes in the survived one as a function of the relative initial size of the two populations and of the mating rate, and (3) We have shown that it is possible to infer the rate of interbreeding from the observed fraction of introgressed genes alone since the relative initial size of the two populations weakly influences the result.

Since the model was initially motivated by a paleoanthropological problem concerning the interbreeding of Neanderthals and African modern humans in Middle East, we have tested the model results against this problem. The result confirms that Neanderthals and Moderns made successful sex at an extremely low rate; nevertheless, this allowed the introgression of a small but not negligible fraction of Neanderthal genes in modern Eurasian populations.

Acknowledgments We are very grateful to Josè Fontanari, Armando Neves and Simone Pigolotti for many illuminating discussions, suggestions and encouragement.

References

Bar-Yosef O (1998) The chronology of the middle paleolithic of the levant. In: Akazawa T, Aoki K, Bar-Yosef O (eds) Neandertals and modern humans in western Asia. Springer, New York, p 3956

Chang JT (1999) Recent common ancestors of all present-day individuals. Adv Appl Probab 31:10021026

Derrida B, Manrubia SC, Zanette DH (1999) Statistical properties of genealogical trees. Phys Rev Lett 82:19871990

Durrett R (2008) Probability models for DNA sequence evolution. In: Asmussen S, Gani J, Jagers P, Kurtz ThG (eds) Probability and its applications, 2nd edn. Springer, New York, pp 252–253

Fu Q, Hajdinjak M, Moldovan OT, Constantin S et al (2015) An early modern human from Romania with a recent Neanderthal ancestor. Nature 524:216219

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U et al (2010) A draft sequence of the Neandertal genome. Science 328:710722

Hammer MF, Woerner AE, Mendez FL, Watkins JC, Wall JD (2011) Genetic evidence for archaic admixture in Africa. Proc Natl Acad Sci 468(37):512315128

Neves AGM (2012) Interbreeding conditions for explaining Neandertal DNA in living humans: the non-neutral case. In: Mondaini RP, Dilão R (eds) BIOMAT 2011. World Scientific, Singapore

Neves AGM, Serva M (2012) Extremely rare interbreeding events can explain Neandertal DNA in living humans. PLoS One 7(10):e47076

Reich D, Green RE, Kircher M, Krause J, Patterson N et al (2010) Genetic history of an archaic hominin group from Denisova cave in Siberia. Nature 468:10531060

Sankararaman S, Patterson N, Li H, Pääbo S, Reich D, Akey JM (2012) The date of Interbreeding between Neandertals and modern humans. PLoS Genet 8(10):e1002947

Serva M (2005) On the genealogy of populations: trees, branches and offspring. J Stat Mech: Theory Exp P07011

Traulsen A, Claussen JC, Hauert C (2005) Coevolutionary dynamics: from finite to infinite populations. Phys Rev Lett 95:238701