Applications of the Galton-Watson process to human DNA evolution and demography

Armando G. M. Neves and Carlos H. C. Moreira

UFMG, Departamento de Matematica, Caixa Postal 702, 30123-970, B. Horizonte - MG, Brazil

Abstract

We show that the problem of existence of a mitochondrial Eve can be understood as an application of the Galton–Watson process and presents interesting analogies with critical phenomena in Statistical Mechanics. In the approximation of small survival probability, and assuming limited progeny, we are able to find for a genealogic tree the maximum and minimum survival probabilities over all probability distributions for the number of children per woman constrained to a given mean. As a consequence, we can relate existence of a mitochondrial Eve to quantitative demographic data of early mankind. In particular, we show that a mitochondrial Eve may exist even in an exponentially growing population, provided that the mean number of children per woman $\overline{N}$ is constrained to a small range depending on the probability p that a child is a female. Assuming that the value $p \approx 0.488$ valid nowadays has remained fixed for thousands of generations, the range where a mitochondrial Eve occurs with sizeable probability is $2.0492 < \overline{N} < 2.0510$. We also consider the problem of joint existence of a mitochondrial Eve and a Y chromosome Adam. We remark why this problem may not be treated by two independent Galton–Watson processes and present some simulation results suggesting that joint existence of Eve and Adam occurs with sizeable probability in the same $\overline{N}$ range. Finally, we show that the Galton–Watson process may be a useful approximation in treating biparental population models, allowing us to reproduce some results previously obtained by Chang and Derrida et al..

Key words: Branching processes, Mitochondrial Eve, Mitochondrial DNA, Y Chromosome, Evolutionary Genetics
PACS: 87.23.Kg, 02.50.-r, 05.40.-a

Email addresses: aneves@mat.ufmg.br (Armando G. M. Neves), cmoreira@mat.ufmg.br (Carlos H. C. Moreira).

1 Introduction

Francis Galton and the reverend H. W. Watson formulated and solved in 1874 the model named after them to explain the phenomenon of disappearance of family names. They considered that family names are passed by men to further generations only by their male children. Also, the number of sons of each man is considered as a random variable taking value $r \in \{0, 1, 2, \ldots\}$ with probabilities $q_0, q_1, q_2, \ldots$ respectively. The number of sons of two men are considered as independent random variables and the values of the q_r identical for any man at any time.

With these hypotheses, a probability of occurrence can be assigned to each genealogic tree. The basic result of Galton and Watson, reviewed in section 2, is that if $\sum_{r=1}^{\infty} r q_r > 1$, the probability of the genealogic tree being infinite is positive. In other words, if the mean number of sons is greater than 1, a positive fraction of all family names existing at some time will survive after any finite number of generations.

The Galton–Watson (GW) process has been exhaustively studied since then, together with some of its generalizations. The field came to be known in Mathematics as the theory of branching processes [1]. Applications of the GW process in Evolutionary Genetics date at least back to the 1920's in a work by Haldane [2] on the survival of mutant genes.

In section 4 we shall introduce the basics on mitochondrial DNA inheritance and the mitochondrial Eve. The idea of applying the GW model to study these issues seems to have appeared for the first time in [3]. That paper has the importance of having shown that a strict population bottleneck is not necessary for the existence of a mitochondrial Eve. On the other hand, it explicitly claims that the existence of a mitochondrial Eve in an exponentially increasing population is impossible.

Our main result is the possibility of existence of a mitochondrial Eve in an exponentially growing population. Using realistic data on the number of humans living at the time of the mitochondrial Eve and the time elapsed since then, we show that a mitochondrial Eve occurs with sizeable probability, up to 37%, if the mean number of children per woman lies between 2.0492 and 2.0510 and the maximum number of children per woman is 10. These figures are obtained without any further knowledge on the probability distribution for the number of children per woman by using an optimization argument over all possible probability distributions.

Despite the simplicity of the arguments and the idea of using a well-known model, as far as we know these results have never been explicitly stated before. We have seen no improvements on the results in [3] in the Biology literature. A

different path is to produce Monte Carlo simulations of interesting and more complicated population models and see that these models exhibit a mitochondrial Eve [4–7]. Although much can be learned from these simulations, results are restricted to a finite number of probability distributions for the number of children and run into serious problems for quickly growing populations. As a simple exactly solvable model for mitochondrial DNA inheritance, the GW process may be useful in better understanding such simulations, its role being analogous to the one of the two-dimensional Ising model in Statistical Mechanics.

An important consequence of our results is that the existence of a mitochondrial Eve in an increasing population shows that the Multiregional Evolution hypothesis in Anthropology cannot be discarded.

Although the same GW model might be used to study existence of the male analogue of the mitochondrial Eve - the Y chromosome Adam - we will explain why the question of joint existence of a mitochondrial Eve and a Y Adam is more difficult. We present two results of small simulations of reasonable models and see that the numbers of surviving mitochondrial DNA and Y chromosome lineages are correlated and joint existence of a mitochondrial Eve and a Y Adam occurs with a non-negligible probability if demographic parameters are chosen consistently.

Finally, we further analyze a model of a biparental population. In the GW model, individuals receive the relevant information from a single parent. In the family name problem, the information is the family name and the parent which transmits it is the father. In the mitochondrial DNA inheritance problem, it is the mother who transmits the information. Chang [8] and Derrida, Manrubia and Zanette [9–11] independently presented results on the generalization of the GW model to the biparental case. Both groups consider a population of fixed size N with the number of children per individual obeying a Poisson distribution with mean 2. They show that after a small number of generations of order $\log N$ and with large probability, a fraction of about 20% of the initial population will have no descendants at all. More surprisingly, with large probability, the remaining 80% of the initial population will be ancestors to the whole population also after a number of generations of order $\log N$. We will reproduce the above results in a much simpler way by using again the GW model and a reasonable approximation.

Our paper is organized as follows. In section 2 we shall review the basic results on the GW model. In section 3 we shall deal with the time scales related to two events: the convergence to the infinite generations limit and the extinction of trees. Although the results in this section are probably not original, we feel that scarce interest has been devoted in the applications literature to these important issues. In section 4, we briefly introduce the relevant genetic and

anthropologic background and proceed to apply the GW results to the problems of the mitochondrial Eve and Y Adam. In section 5 we study applications to a biparental model similar to the ones studied by Chang and Derrida et al.. Some concluding remarks are presented in section 6.

2 The Galton–Watson model

Consider the set $\mathcal{T}$ of all rooted tree graphs, finite and infinite. The root of each tree will be referred to as generation 0. The vertices immediately linked to the root are vertices at generation 1 and, in general, vertices linked to the vertices at generation k are at generation $k-1$ if one is going towards the root and at generation $k+1$ if one is going away from it. Let also numbers $q_0, q_1, \ldots$ be given such that $q_r \geq 0,\ r = 0, 1, 2, \ldots$ and $\sum_{r=0}^{\infty} q_r = 1$.

The GW process may be thought of as the assignment of a probability measure P to subsets of $\mathcal{T}$ according to the following rule. If $T \in \mathcal{T}$ is a tree with a vertex v at generation k, we define $c(v)$ as the number of vertices at generation $k+1$ stemming from v. In other words, $c(v)$ is the number of "children" of v. The probability of a tree is

$$P(T) = \prod_{v \in T} q_{c(v)} \ . \tag{1}$$

In the family name problem of Galton and Watson, the root of a tree is a man bearing some family name and each tree is some possible male genealogy of his, i.e. his sons, the sons of his sons and so on, each son being linked to his father. The numbers q_r mean the probability that a man has r sons, $P(T)$ is the probability of genealogy T and the way P was constructed reflects the properties of statistical independence of men and time independence of the progeny distribution built in the model.

Let $E_n \subset \mathcal{T}$ be the set of all trees which come to an end, but not after generation n, and $E = \cup_{n=0}^{\infty} E_n$. In the family name problem, E means of course the set of all genealogies which lead to extinction of the family name of the man in the root. Define also the generating function $S(x)$ of the probability distribution defined by the q_r as

$$S(x) = \sum_{r=0}^{\infty} q_r x^r \tag{2}$$

and let $\overline{\theta}_n = P(E_n)$ be the probability of extinction in at most n generations. We also denote $\theta_n = 1 - \overline{\theta}_n$. As any tree in E_n can be thought of as formed

by its root, its first generation, if present, and, attached to each vertex in the first generation, a tree in E_{n-1}, then

$$\overline{\theta}_n \;=\; P(E_n) \;=\; \sum_{r=0}^{\infty} q_r \, P(E_{n-1})^r \;=\; S(\overline{\theta}_{n-1}) \;, \tag{3}$$

where the rth term in the sum corresponds to the possibility that the root has r children. Taking limits at both sides,

$$\overline{\theta} \;=\; S(\overline{\theta}) \;, \tag{4}$$

where $\overline{\theta} = \lim_{n\to\infty} \overline{\theta}_n$ is the extinction probability, a fixed point of S according to (4). The initial condition to be used in conjunction with (3) in order to determine $\overline{\theta}_n$ is $\overline{\theta}_0 = q_0$.

The normalization of the q_r's implies by (2) that 1 is a fixed point of S. Since S is non-decreasing with all derivatives non-decreasing in $[0,1]$, it will have a second fixed point in $[0,1)$ if and only if $S'(1) > 1$. There may be no other fixed points of S in $[0,1]$. Regarding number and attractiveness [12] of the fixed points of S in $[0,1]$, we may have three different regimes:

(i) If $S'(1) < 1$, then 1 is the only fixed point and it is attractive.
(ii) If $S'(1) > 1$, then 1 is a repulsive fixed point, whereas the other fixed point is attractive.
(iii) If $S'(1) = 1$, then 1 is again the only fixed point in $[0,1]$ and it is *weakly* attractive.

These results imply that $\theta = 1 - \overline{\theta}$ is 0 if $S'(1) \leq 1$ and positive otherwise. As

$$S'(1) \;=\; \sum_{r=1}^{\infty} r q_r \;=\; m \;,$$

where m is the mean number of children of the vertices, we have recovered the classical result [1] that the probability θ that a family name survives is positive only if $m > 1$ and zero otherwise.

The survival probability may be obtained by finding the solution for (4) in $[0,1)$. If $m > 1$ and m close enough to 1, then $\overline{\theta} \approx 1$. Replacing S by its Taylor polynomial of degree 2 around 1 we find

$$\theta \;\approx\; \theta_{\mathrm{a}} \;=\; \frac{2(m-1)}{\sum_{r=2}^{\infty} r(r-1)\, q_r} \;=\; \frac{2(m-1)}{v + m(m-1)} \;, \tag{5}$$

where $v = \sum_{r=0}^{\infty} (r-m)^2 q_r$ is the variance of the number of children of the

vertices. From the non-negativity of S''', it follows that $\theta_{\rm a}$ is actually a lower bound for θ. We shall refer to the above approximation for θ as the *small survival probability approximation.*

A last fact we quote without proof regards the typical number Z_n of branches of a tree at generation n. Of course, as trees are regarded as genealogical histories to which we assign probabilities, Z_n is a random variable. When trees do not terminate at a finite number of generations, it is expected that Z_{n+1} is approximately $m\,Z_n$ because in the average each vertex at generation n should produce m children. If this simple reasoning is indeed true, we should have $Z_n \sim m^n$. This fact is rigorously proved in [1], chapter I, section 8.1. In simple words, populations in the GW process grow in the average exponentially according to the Malthus law.

3 Time scales

In the applications to follow, we shall need to consider two time scales related to the GW process. The first is the number of generations it takes for the $\overline{\theta}_n$ in (3) to approach $\overline{\theta}$ in (4). The second is the typical number of generations in a finite tree.

By subtracting (3) from (4) and using the mean value theorem of Calculus, there must then exist $\alpha \in (\overline{\theta}_{n-1}, \overline{\theta})$ such that $\overline{\theta} - \overline{\theta}_n = S'(\alpha)\,(\overline{\theta} - \overline{\theta}_{n-1})$. Iterating this argument, we obtain $\overline{\theta} - \overline{\theta}_{n_0+k} = \prod_{j=1}^{k} S'(\alpha_j)\,(\overline{\theta} - \overline{\theta}_{n_0})$. For large enough n_0, we may approximate each α_j by $\overline{\theta}$, from which it follows, if $S'(\overline{\theta}) \neq 1$, that

$$\theta_n - \theta \sim {\rm e}^{-n/\xi} \tag{6}$$

with the *correlation time* being

$$\xi = -1/\ln[S'(1-\theta)]\ . \tag{7}$$

In the critical regime $S'(1) = 1$, the correlation time diverges. In this case, instead of the exponential decay in (6), the convergence of θ_n to $\theta = 0$ is polynomial if $S'''(1) < \infty$:

$$\theta_n \sim \frac{2}{S''(1)\,n}\ . \tag{8}$$

A proof of this is given in [1], chapter 1, section 10.2.

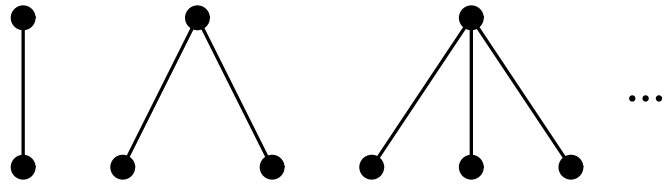

Fig. 1. Trees with exactly one generation.

A curious fact which will be useful later is that in the small survival probability approximation, although θ in (5) depends on the variance v of the probability distribution for the number of children of the vertices, ξ does not. Indeed, the reader may check that if m is close to 1, either larger or smaller,

$$\xi \approx \frac{1}{|m-1|} \, . \tag{9}$$

Let us now consider the typical number of generations in a finite tree. In order to calculate that, define p_n as the probability that a tree terminates *exactly* at the nth generation. Then $p_0 = q_0$ and for $n \geq 1$

$$p_n \, = \, P(E_n \setminus E_{n-1}) \, . \tag{10}$$

The trees in $E_1 \setminus E_0$ are shown in figure 1 and their probabilities add up to p_1. By using (1) we see that

$$p_1 \, = \, q_1 q_0 \, + \, q_2 q_0^2 \, + \, q_3 q_0^3 \, + \, \ldots \, = \, S(q_0) \, - \, q_0 \, = \, S(p_0) \, - \, p_0 \, .$$

If we want now to calculate p_n, $n \geq 2$, we may consider similar trees, but instead of terminating at generation 1, we think that appended to each vertex at generation 1 there is a tree with at most $n-1$ generations rooted at that vertex, i.e. a tree in E_{n-1}. If the number of vertices at generation 1 is k, then at least one of the k trees appended to these vertices must not be in $E_{n-2} \subset E_{n-1}$. Otherwise, the tree would terminate in less than n generations. Thus we calculate the contribution of the trees with k vertices at generation 1 to p_n by allowing trees in E_{n-1} appended to all vertices and subtracting the case of all such those trees being in E_{n-2}. We mean that for $n \geq 2$,

$$p_n \, = \, \sum_{k=1}^{\infty} q_k \, \left[P(E_{n-1})^k - P(E_{n-2})^k \right] \, = \, S(P(E_{n-1})) \, - \, S(P(E_{n-2})) \, .$$

Using now that $P(E_n) = \sum_{r=0}^{n} p_r$, we obtain

$$p_n \, = \, S(\sum_{r=0}^{n-1} p_r) \, - \, S(\sum_{r=0}^{n-2} p_r) \, . \tag{11}$$

Of course, $\overline{\theta} = \sum_{r=0}^{\infty} p_r$. Then the probability p_n that a finite tree has exactly n generations must tend to 0 as $n \to \infty$. We may calculate the speed of this convergence by using again the mean value theorem trick in (11). For large enough n, we get $p_n \sim \mathrm{e}^{-n/\xi}$ with ξ being given by (7). It turns out that the two time scales we wanted to consider are equal.

4 Mitochondrial Eve and Y Adam

4.1 Mitochondrial DNA and human evolution

Although most of the genetic information in higher animals is located at cells' nuclei, some DNA may be found in the subcellular organelles called mitochondria. These structures are present in large numbers in nearly every cell and play a key role in metabolism. Mitochondrial DNA (mtDNA) is very short; in humans it consists in only 16,569 base pairs carrying the information for only 37 genes. Despite that, mtDNA of humans and many other species has been the object of recent intensive research [13], both for its availability (because mitochondria are so numerous) and peculiar inheritance. Unlike nuclear DNA, which is inherited in equal parts from mother and father, mtDNA is inherited only from the mother. So, in the absence of mutations, the mtDNA of an individual would be identical to the mtDNA of a single ancestor out of his or her up to [10] 2^n ancestors n generations before, namely the mother of the mother ... of his or her mother. Nuclear and mitochondrial DNAs also differ because most biologists believe [13] that the latter is neutral from the natural selection point of view.

This simple feature allows one to use mtDNA comparisons among living individuals to look far back in time and draw conclusions about the separation of subpopulations in one species or the speciation process, in which several extant species may descend from one extinct species [13].

An experiment performed in the late 80's brought press popularity to mtDNA. By examining mtDNA of 147 living humans, Cann, Stoneking and Wilson asserted in [14] that mtDNA of all living humans could be described as arising from mutations in the mtDNA of a single woman. As we would all be her descendants, this woman was called the *mitochondrial Eve*. By using known mutation rates and exploiting geographical correlations, it could be inferred that the mitochondrial Eve has lived in Africa more or less 200,000 years ago.

The mitochondrial Eve should not be confused with the biblical Eve. Unlike the latter, the mitochondrial Eve is not supposed to be the only woman living at her time. As we will see in section 5, most men and women contemporary

to her with large probability have left traces of their nuclear DNA in modern humans, but not of their mtDNA.

The time and the place in which the mitochondrial Eve would have lived are considered a strong evidence for the *Out of Africa* model for the origin of our own species [15]. This model proposes that modern humans, *Homo sapiens*, evolved once in Africa and subsequently colonized the rest of the world replacing, *without mixing with them*, earlier forms such as *Homo neanderthalensis* and *Homo erectus*, which had already colonized all the other continents, except for America.

The competing *Multiregional evolution* model [16] suggests instead that modern humans evolved from earlier forms concurrently in different regions of the world, with occasional genetic flow among regions, necessary to preserve uniqueness of our species. According to this view, neanderthals, for example, would not be a species different from ours, but the group of modern humans which inhabited Europe and West Asia.

Defenders of the Out of Africa Model claim that if we had been descendants of the neanderthals or of other early hominids, then some of us would have to carry mtDNA very different from the existing types reported in [14]. In other words, a mitochondrial Eve would not exist. Controversy was stirred by fossil child bones found in Portugal in 1998 [17] and dated of 25,000 years ago, 5,000 years after extinction of the neanderthals. This finding suggests there might have been mixing among neanderthals and modern humans which came from Africa and arrived in Europe around 40,000 years ago. Despite some dispute, the majority of contemporary paleoanthropologists believe in the Out of Africa model, having decided for the genetic evidence in favour of the morphologic one.

In order to explain why all mtDNA lineages stemming from women contemporary to the mitochondrial Eve were extinct, Brown proposed in [18] that a severe bottleneck, in which human population dropped to only a few individuals, must have existed after the mitochondrial Eve. This is a very strong hypothesis, considering that human population, at least in historical times, has been steadily growing, and that achievement of important technological developments in prehistory made it possible for humans to spread all over the world. In [3] Avise, Neigel and Arnold argued that a prehistorical population bottleneck is not strictly necessary by showing that stochastic mtDNA lineage extinction can be sufficiently rapid even in *stable-sized* populations. In the next subsection we shall quantitatively show that a mitochondrial Eve may exist even if the population grows exponentially.

4.2 The mitochondrial Eve in a growing population

The assumptions in our model for mtDNA inheritance are the following:

(A1) Generations are nonoverlapping.
(A2) The numbers of children for each woman are statistically independent and identically distributed random variables assuming value $r \in \{0, 1, 2, \ldots\}$ with probability Q_r. The values for the Q_r's are time- and population size-independent.
(A3) A newborn child is a female with probability p and a male with probability $1 - p$. The value for p is also time- and population size-independent.
(A4) There always exist males enough to mate with all females.

Assumption (A1) grants some formal simplicity to the model. (A2) disregards interactions that might come from a number of different sources such as fertility correlations among members of a family, competition (for food supplies, mating partners, etc), cooperation and geographical aspects. One natural attempt to improve the model would be to make the Q_r's dependent on the total population, thus accounting for saturation effects. Although feasible in computer simulations, that would ruin the linearity on which our theoretical analysis relies. The time independence of the Q_r's is also questionable because it disregards the effects for example of climate changes. However this assumption may be adequate over an initial period of time long enough to produce most of the lineage extinctions. (A3) adds some generality to the model with respect to the one in [3], in which $p = 1/2$. Usually p varies from species to species and for modern humans $p \approx 0.488$, strictly less that $1/2$ [19]. This fact will be of quantitative relevance in our results. (A4) is assumed, since we do not keep track of the male population. It is a reasonable assumption because males from all concurrent mitochondrial lineages may participate in any single one, without interfering in the mitochondrial inheritance. Also, even for p only slightly less than $1/2$, the results in subsection 4.3 will show that male extinction is much less probable than female extinction for humans.

Consider now the genealogic tree of an ancestral woman constructed according to the above assumptions. By genealogic tree we mean the rooted tree graph obtained by considering as vertices the ancestral herself, located at the root, and all her descendants after an *infinite* number of generations, drawing edges joining each father or mother to their children of either sex. Define as *open* any edge linking a mother to her children and *closed* all other edges. It is clear that the mtDNA lineage of the woman at the root will survive if and only if there is an infinite path of open edges beginning at the root. In other words, if the configuration of open edges starting at the root percolates [20]. Therefore it is instructive to view mtDNA inheritance as a problem of edge percolation in a tree graph, in which edges are open with probability p and closed with

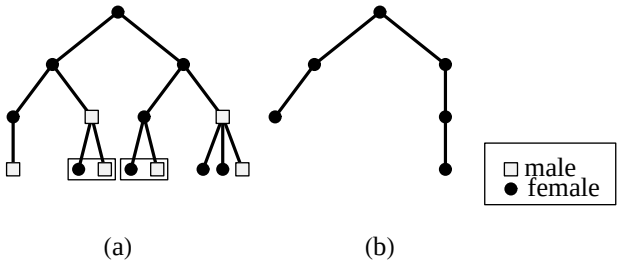

Fig. 2. (a) Example of four generations in a genealogic tree of the type considered in this paper. The individuals enclosed in the rectangle have both their father and mother in the tree, exemplifying the possibility of statistically dependent edges. (b) The female genealogic tree corresponding to the complete tree in (a).

probability $1 - p$. Unfortunately, unlike most percolation models, edges may not be statistically independent because any child with both parents in a tree will appear twice, separately linked to each of the parents. An example is shown in figure 2(a).

In order to overcome the dependence problem, we define the female genealogic tree (FGT) as the tree obtained after stripping the genealogic tree of all male individuals, see figure 2(b). As every woman has a single mother, in the FGT all edges are open and statistically independent. Percolation in the complete genealogic tree is equivalent to the corresponding FGT being infinite. It should be also clear by our assumptions that FGTs have probabilities assigned by a GW process with

$$q_r \;=\; \sum_{k=r}^{\infty} Q_k \begin{pmatrix} k \\ r \end{pmatrix} p^r \,(1-p)^{k-r} \;, \tag{12}$$

where q_r, $r = 0, 1, 2, \ldots$, is the probability that an individual has r children of female sex.

By using (12) and the results in section 2, we may express the survival probability θ of a mtDNA lineage in terms of the demographic parameters Q_r. If

$$\overline{N} \;=\; \sum_{r=1}^{\infty} r\, Q_r \tag{13}$$

is the mean number of children of either sex per woman, we see that there exists a percolation threshold or critical probability

$$p_{\rm c} \;=\; \frac{1}{\overline{N}} \;, \tag{14}$$

such that the survival probability is positive only if $p > p_c$.

If we take $p = 1/2$, the critical case $p = p_c$ corresponds to $\overline{N} = 2$, i.e. a population on the average with a fixed size. We may then use (8) to find that the expected number of surviving mtDNA lineages decays with the number of generations exactly as n^{-1}. In our model, such a power law is thus a characteristic of constant population and absence of any mechanism of genetic selection for mtDNA, which could possibly speed up the lineage extinction process. In [4] authors consider a constant on average population in a variation of the Penna model for biological ageing. They observe that a mitochondrial Eve appears in their model and claim that the number of surviving lineages decays as n^{-z} with z roughly equal to 1, in accordance with our results. On the other hand, a more recent work [7] on the standard asexual Penna model displayed an initial n^{-1} behavior gradually changing to n^{-2} in very long simulations. According to the authors, the initial behavior is a consequence of the population being limited in their model, whereas the n^{-2} follows from accumulated genetic selection associated to an ageing mechanism not present in our model.

If we take p as analogous to the inverse temperature, p_c as the critical value for the inverse temperature and θ as an order parameter, then our model for mtDNA inheritance behaves as a critical system in Statistical Mechanics. If $p > p_c$, there is a non-zero probability that trees are infinite. This fact resembles long range order, such as spontaneous magnetization arising in a magnet at the ferromagnetic phase. Trees being infinite does not mean that populations are infinite - they tend to infinity as the number of generations tends to infinity, but are finite at any finite time.

Up to now we have studied survival probability for a single mtDNA lineage. We may now proceed to the study of the number of surviving mtDNA lineages. Let W be the number of women contemporaneous to the mitochondrial Eve. As in our assumptions, different lineages may be considered as independent. Then the number r of lineages remaining after n generations is a binomially distributed random variable with

$$\text{Prob}\{r = l\} = \binom{W}{l} \theta_{n-1}^{l} (1 - \theta_{n-1})^{W-l} , \tag{15}$$

where $\theta_{n-1} = 1 - \overline{\theta}_{n-1}$ is the probability that a tree does not terminate in $n - 1$ or less generations.

This same approach for calculating the number of mtDNA lineages surviving after n generations was used in [3], although authors considered only probability distributions for which they were able to sum explicitly series (2). They concentrated on the probability Π_n for the survival of two or more lineages

after n generations.

They found out that in the supercritical regime $p > p_c$, Π_n tends to a positive value as $n \to \infty$, which agrees with our results. As this means that there is a finite probability that a mitochondrial Eve does not exist, in case more than one lineage survives, they discarded this solution as incompatible with a mitochondrial Eve. The subcritical regime $p < p_c$ was also discarded because it of course leads to exponential extinction. Instead, the critical regime $p = p_c$ was selected because it is the only regime in which $\Pi_n \to 0$ slowly. With $W = 1,000 \sim 10,000$ they showed it leads to Π_n approaching zero in $n \approx 10^4$ generations. Such values for W and n are well within the range expected by geneticists and paleontologists. For $p = 1/2$, as used in [3], the critical regime yields $\overline{N} = 2$ and can only account for a stable-sized population. In other words, although showing the possibility of existence of a mitochondrial Eve in a stable-sized population, authors in [3] missed the possibility of a mitochondrial Eve in a growing population.

We shall now argue that for a range of values of $\overline{N}$, the supercritical regime also provides a biologically plausible solution for the existence of a mitochondrial Eve in a growing population. The human population at the time of the mitochondrial Eve may be estimated by using data from variability in the nuclear DNA of living people [21]. A reasonable estimate is $W = 5,000$. Assuming 20 years per generation, mtDNA [14], and fossil findings [22] agree in estimating the number of generations from the mitochondrial Eve to nowadays as $n \approx 10^4$. We also use $p = 0.488$, obtained from the standard figure of 105 male births per 100 female births, the modern human *sex ratio at birth* [19], assuming that this ratio can be extrapolated to the times of early mankind.

From (15) we get that the expected number of surviving lineages after n generations is $W\theta_n$. To be consistent with the existence of a mitochondrial Eve as an event with a not too small probability, this number must not be much smaller nor much larger than 1. For illustration purpose, we take it between $1/2$ and 2, implying θ_n between $1/(2W)$ and $2/W$. In order to estimate the range of values for $\overline{N}$ consistent with that, we first approximate θ_n by

$$\theta_a = \frac{2(p - p_c)}{p_c\, p^2 \sum_{r=2}^{\infty} r(r-1)\, Q_r} , \tag{16}$$

which may be deduced from (5) and (12). This approximation is valid as long as two conditions are fulfilled. First, the number of generations n must be so large that θ_n is close to θ. Finally, the small survival probability approximation must hold in order that θ is close to θ_a. The latter is true, because θ is of order W^{-1}, a small number if $W = 5,000$. The former, $\theta_n \approx \theta$, will be justified soon.

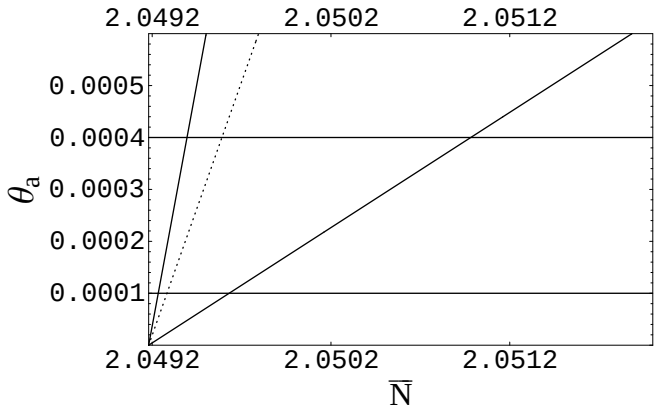

Fig. 3. Continuous lines correspond to the maximum and minimum values for $\theta_{\rm a}$ as functions of the mean number of children per woman $\overline{N}$ with $M = 10$. The dotted line is the value of θ as a function of $\overline{N}$ for a Poisson distribution for the progeny. Horizontal lines correspond to the values $1/(2W)$ and $2/W$ for $\theta_{\rm a}$.

Also, for practical purposes we may truncate the series in the right-hand side of (16) at some order M (limited progeny assumption). The problem of estimating the maximum and minimum of $\theta_{\rm a}$ for given $\overline{N}$ then becomes a linear optimization problem, exactly solvable by standard methods: we must respectively minimize or maximize the linear function $\sum_{r=2}^{M} r(r-1)\,Q_r$ under the constraints $\sum_{r=1}^{M} r\,Q_r = \overline{N}$, $\sum_{r=1}^{M} Q_r = 1$ and $0 \leq Q_r \leq 1$ for $r = 0, 1, \ldots, M$.

In figure 3 we show a plot of the maximum and minimum values for $\theta_{\rm a}$ as functions of $\overline{N}$, taking $M = 10$. The values for $\overline{N}$ consistent with existence of a mitochondrial Eve as a not very rare event must thus obey two conditions. The first is that the maximum $\theta_{\rm a}$ lies over $1/(2W)$, which yields $\overline{N} > 2.0492$. Otherwise, the expected number of surviving mtDNA lineages is much smaller than 1. The second condition is, analogously, that the minimum $\theta_{\rm a}$ lies below $2/W$, which yields $\overline{N} < 2.0510$. Otherwise, the expected number of surviving mtDNA lineages is much larger than 2.

The value $M = 10$ for the maximum number of children is of course arbitrary. However it turns out that the maximum values for $\theta_{\rm a}$ become independent of M for $M \geq 3$ and are obtained with $Q_r \neq 0$ only for $r = 2$ and 3. On the other hand, the minimum values for $\theta_{\rm a}$ do depend on M and tend to 0 as $M \to \infty$ This is why we must impose a maximum value for the number of children. Nonetheless, the minimum value for $\theta_{\rm a}$ is attained when $Q_r = 0$, $r = 0, 1, 2, \ldots, M-1$ and $Q_M = \overline{N}/M$, which is quite an unrealistic probability distribution for the number of children. Imposing realistic conditions on the Q_r's would constrain the average number of children consistent with existence of a mitochondrial Eve more than the use of $M = 10$. This is also illustrated in figure 3, where the dotted line corresponds to the survival probability in case the distribution probability for the progeny is the Poisson distribution, a reasonably realistic distribution.

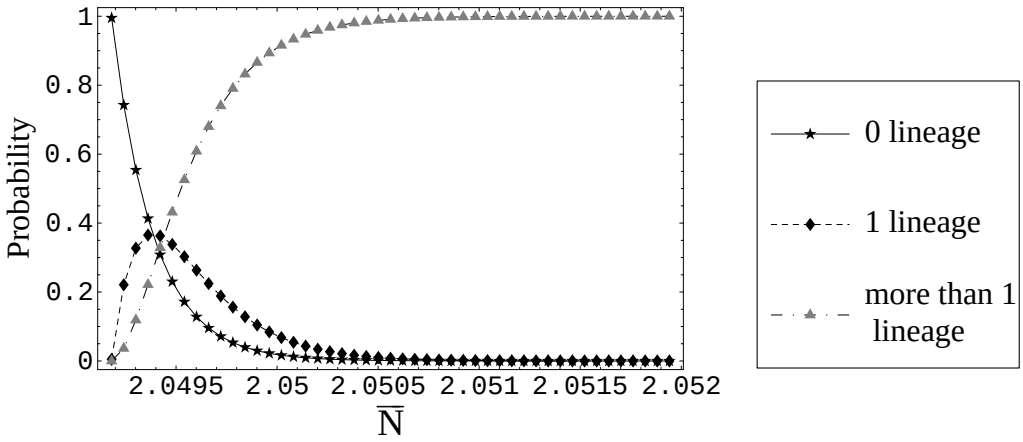

Fig. 4. Probability of survival for l mtDNA lineages as a function of the mean number of children per woman $\overline{N}$, for $l = 0$, $l = 1$ and $l > 1$. We considered here that the progeny distribution is a Poisson distribution.

For each progeny distribution, the number of surviving mtDNA lineages is random with probability given by (15). In the range of values we consider for θ and W, in which $W \gg 1$ and $\theta \ll 1$, this binomial distribution may be very well approximated by a Poisson distribution with mean $W\theta$. This means that the probability of only one lineage surviving is of the form $x\mathrm{e}^{-x}$ with $x = W\theta$. So, the maximum probability for the existence of a mitochondrial Eve is approximately equal to $\mathrm{e}^{-1} \approx 0.37$ for any progeny distribution, provided $W \gg 1$ and $\theta \ll 1$. This is illustrated at figure 4, where we considered that the progeny distribution is also Poisson and we also showed graphs for the probability of no lineages surviving and more than one lineage surviving.

Results plotted in figure 3 assume, as already mentioned, that θ_n can be approximated by its limit θ when $n \to \infty$. In other words, the results will be valid provided $n \gg \xi$, where the correlation time ξ is given by (7). We already know by (9) that for fixed $\overline{N}$, in the small survival probability approximation ξ does not depend on the variance of the progeny distribution. In figure 5 we show the plot of ξ as a function of $\overline{N}$ for a Poisson distribution for the progeny, but, as a consequence of the independence of the variance, the graph is indistinguishable from the analogous ones for binomial distributions or the maximum and minimum survival probability distributions. In particular, for the largest value $\overline{N} = 2.0510$ compatible with the existence of the mitochondrial Eve, it can be seen that ξ is of the order of 1,100 generations almost independently of the progeny distribution. As geneticists assume that the mitochondrial Eve lived more or less 10,000 generations ago, the approximation $\theta_n \approx \theta$ is well justified for $\overline{N} = 2.0510$. On the other hand ξ diverges when $\overline{N}$ approaches $1/p$. This means that the range of values for $\overline{N}$ compatible with the Eve should be extended down to $1/p$.

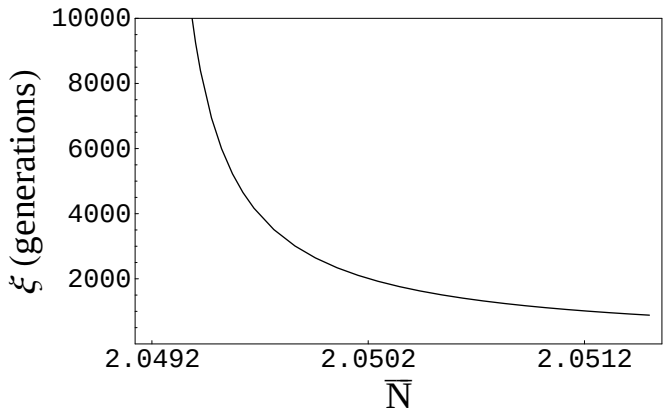

Fig. 5. Correlation time ξ as a function of the mean number of children per woman $\overline{N}$. The curve shown is for a Poisson distribution, but the corresponding curve for other distributions is indistinguishable from this.

Finally, although our data do not directly support the Multiregional Evolution model, they show that it cannot be discarded. In fact, if there had been some genetic mixing among modern humans coming from Africa and other humans living in other continents, our results show that this mixing probably would have completely disappeared from mtDNAs.

4.3 The Y Adam

The Y chromosome is transmitted from fathers to sons and not to daughters. Furthermore, there is a portion in the Y chromosome that seems not to be subject to recombination. Most geneticists also believe that there is no natural selection acting on this portion of the Y chromosome - it is sort of genetic trash which does not code any protein or has some other visible function. In this sense, the non-recombining portion of the Y chromosome would be the male analogue of mtDNA. But experiments sampling worldwide variation in Y chromosomes of living people seem to be much more difficult than their mtDNA analogues, such as the one in [14]. In the experiment reported in [23], authors are still cautious about having discovered the majority of human Y chromosome polymorphisms. Despite that, they affirm the existence of a most recent common ancestor to the Y chromosomes of their sample, and even calculate that he would have lived between 162,000 to 186,00 years ago, about the same period of the mitochondrial Eve.

Let us now try to understand the question of existence of a Y Adam in terms of our previous results on Eve. Of course, all calculations for survival of female trees must hold for male trees if we exchange p by $1-p$ and take $\overline{N}$ as the mean number of children per man, provided that the male analogues of assumptions (A2) and (A4) at the beginning of subsection 4.2 still apply.

If we assume these hypotheses, a first naive guess would be to consider that the mean numbers of children per man and per woman are the same and simply exchange p by $1-p=0.512$. As survival probabilities increase very rapidly with p when p is slightly larger than $p_{\rm c}$, the effect of that exchange is dramatic. Starting with a male population of 5,000 and using the smallest value $\overline{N}\approx 2.0492$ in the allowed range for existence of a mitochondrial Eve in figure 3, we find that the expected number of Y chromosome lineages surviving is 101.7. Accordingly, the probability of only one such lineage surviving is of order 10^{-44}, rendering existence of a Y Adam virtually impossible.

A serious problem with this simplistic scenario is that the rate of exponential growth for populations of each sex would be different. Whereas women population would grow as $(\overline{N}p)^n$, where n is the number of generations after the beginning, male population would grow as $(\overline{N}(1-p))^n$. With $\overline{N}=2.0510$ and $p=0.488$, male population would be 10^{208} times larger than female population after 10,000 generations, which would violate the male analogue of hypothesis (A4). We thus abandon the simplistic assumption of equal mean numbers of children per woman and per man.

Still dealing with mean numbers, a necessary condition for the tuning of male and female populations is that the growth rate be the same for both genders. As there would be more men than women living at each generation, it is also necessary that men have in the average less children than women. Quantitatively, if we define $\overline{N}_{\rm W}$ as the mean number of children per woman and $\overline{N}_{\rm M}$ as the mean number of children per man, then we must have

$$\overline{N}_{\rm W}\, p \;=\; \overline{N}_{\rm M}\,(1-p)\;, \tag{17}$$

which gives us $\overline{N}_{\rm M}<\overline{N}_{\rm W}$ if $p=0.488$. In this scenario, if the initial population of women is $W=5,000$, then the total population (men and women) T_n at the nth generation is

$$T_n \;=\; \frac{W}{p}\,(\overline{N}_{\rm W}p)^n\;.$$

By using $\overline{N}_W=2.0510$, $p=0.488$ and $n=10,000$, we would have a human population nowadays of 73 million people, which is not too far from the actual value of 6.4 billion, considering roughness of the model and uncertainties in W and n.

Another interesting consequence of this second scenario is that the condition $p>p_{\rm c}$ of survival for female trees may be written as $\overline{N}_{\rm W}\,p>1$ by using (14). By (17), we have that the probability of survival for male trees is also positive. The expected number of surviving Y chromosome lineages will depend however not only on the value of the parameter $\overline{N}_{\rm M}\,(1-p)$, but rather on the

progeny distribution for men, still to be specified. If we further suppose that this progeny distribution is some reasonable one, such as a Poisson distribution, then the probability of existence of a Y Adam will be not negligible.

A more realistic treatment for the problem of joint existence of a mitochondrial Eve and a Y Adam would require a model where (17) holds and some assumption on the progeny distribution for men. Such a model is at present beyond our possibilities. In fact, as it is not possible that only one gender becomes extinct, the number of surviving male and female lineages are of course correlated random variables. Although we have no exact result on this realistic scenario, in order to stimulate further research, we shall present some simulation results, in which the previously considered demographic aspects are implemented.

We simulated two situations. In both cases, we supposed that hypothesis (A4) holds, so that the number of children at generation $n+1$ is determined by the number of females at generation n. Equation (17) is incorporated in both cases because, after we generate a random number of children for each woman at generation n according to some progeny distribution, we choose the sex of each child by independent Bernoulli trials with probability $p = 0.488$ for females.

Our two simulations differ in how we choose the progeny distribution for the male parents. In one case, each child at generation $n+1$ "chooses" at random with equal probability its father in generation n. We will call this the *panmictic* case. In the other case, which we will call *monogamous* case, each woman at generation n chooses at random with equal probability one man at the same generation to be the father of all her children.

Results are presented in table 1. For simplicity, the progeny distribution for women was assumed to be Poisson in both simulations. In order that the correlation time ξ is not too large, we chose the mean of this distribution to be 2.07, so that $\xi \approx 40$ generations. We took initial populations of 49 women and 51 men, each with a different label meaning his/her Y chromosome/mtDNA lineage and counted the number of surviving lineages after 80 generations. The expected number of surviving mtDNA lineages at this time for the given progeny distribution and initial population was 1.12. For each initial population as described we ran the simulation process 100 times. If m denotes the number of surviving male lineages and f the number of surviving female lineages, in table 1 we present for each pair (m, f) the frequency of its occurrence.

In both cases, we see that complete extinction is the most frequent outcome, followed by joint existence of a mitochondrial Eve and a Y Adam. As the maximum value for the probability that only one lineage (Y or mtDNA) survives

Table 1
Frequencies of possible outcomes for the numbers f of surviving mtDNA lineages and m of surviving Y chromosome lineages in a run of 100 simulations for an initial population of 49 women and 51 men along 80 generations and two different assumptions on the male progeny distribution. Details given in the main text.

		$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
$f=0$	panmictic	36	0	0	0	0	0
	monogamous	35	0	0	0	0	0
$f=1$	panmictic	0	24	4	0	0	0
	monogamous	0	24	2	0	0	0
$f=2$	panmictic	0	11	10	5	0	0
	monogamous	0	21	2	0	0	0
$f=3$	panmictic	0	2	2	2	0	0
	monogamous	0	6	8	0	0	0
$f=4$	panmictic	0	0	2	0	0	1
	monogamous	0	1	0	0	0	0
$f=5$	panmictic	0	0	0	0	1	0
	monogamous	0	0	1	0	0	0

is e^{-1}, see figure 4, if m and f were independent, then joint existence of Eve and Adam would occur with probability not greater than $\mathrm{e}^{-2} \approx 0.14$. Data at table 1 clearly contradict this, showing that m and f are correlated. We also see that with large probability m is not much larger or much smaller than f.

5 Biparental model results

Let us now consider a population model with biparental reproduction, i.e. each individual at generation $k+1$ has two parents at generation k. Of course human reproduction is biparental, but in the models considered so far for mtDNA, family name and Y chromosome inheritance, it works as if it were asexual, because individuals inherit the relevant character from a single parent. Consider also that the initial population is N and the mean number of children for any individual at any generation is 2, so that population is on average constant. In order to compare our results with others, we shall suppose that the probability distribution for the number of children is Poisson, i.e. the probability that an individual has r children, $r = 0, 1, 2, \ldots$, is $q_r = \mathrm{e}^{-2}\, 2^r / r!$. Chang proved rigorously some interesting results [8] on a similar model and Derrida et al. [9–11] also studied variants of this model. Of course, as the

number of ancestors of any individual doubles as we proceed one generation to the past and population is of a fixed size, then some ancestors must appear repeatedly as we ascend genealogic trees.

The first interesting result obtained by both groups is that approximately 20% of the individuals in the initial population will have no descendants after some generations. The second is that the remaining 80% of the initial population not only will have descendants at any generation in the future, but after a very small number of generations of order $\log N$, with large probability, these individuals will be ancestors to any individual at future generations. In this section we shall derive again these results in a simpler way.

Before proceeding, we should say that these results do not contradict the ones at the preceding section, because biparental genealogic trees will have roughly twice the number of branches of the corresponding monoparental ones of section 4. As a consequence, survival probabilities are much larger and correlation times much smaller. Another remark is that although individuals have two parents, in the models considered by ourselves in this section and by Derrida et al. and by Chang, gender questions are not addressed. For example, the difficult question, as we mentioned before, of calculating the probability for joint existence of a mitochondrial Eve and Y chromosome Adam is not answered. Notice that the results are related to questions of ancestorship, but do not deal with the gender of ancestors.

We have already touched on the difficulty of applying the GW process to such a biparental model when referring to figure 2. The fact is that whenever repetitions start to occur at the trees, the assumption of statistical independence of the vertices will not hold anymore. In the cases considered so far, we eliminated the problem by considering trees of individuals of a single sex. For such trees, statistical independence is restored as already commented.

We may then suppose that the GW process is *approximately* accurate to describe our biparental model as long as the number of generations is small enough so that repetitions occur seldom at genealogic trees. In case the q_r are given by the Poisson distribution, we may exactly sum the series for the generating function obtaining $S(x) = \mathrm{e}^{2x-2}$.

The survival probability for one such tree may be found by numerically solving (4) with the above expression for S. The result is $\theta \approx 0.7968$, which means that approximately 20.32% of the trees will not survive. Of course, we may only trust this number if the time it takes for a tree to be extinct is small enough so that repetitions do not occur.

By the argument in section 3 we already know that trees which will eventually be extinct live a number of generations of order ξ. By using the result for θ in (7), we get $\xi \approx 1.11$, a very small number of generations. So, as long as

$N \gg 1$, the above GW approximation is valid.

In order to obtain the second result, consider p_n defined before (10). Let A_k be the number of individuals at generation 0 whose genealogic trees are not extinct at generation k and B_k be the number of individuals at generation 0 whose genealogic trees are not extinct at generation k but will be extinct sometime. The probability that $A_k = n$ is given by the binomial distribution

$$\binom{N}{n} (1 - (p_0 + p_1 \, + \, \ldots p_{k-1}))^n (p_0 + p_1 \, + \, \ldots p_{k-1})^{N-n} \, .$$

The expected value of A_k is then $\mathcal{E}(A_k) \, = \, N \, (1 - (p_0 + p_1 \, + \, \ldots p_{k-1}))$. By subtracting $N\theta$ from this number we get

$$\begin{aligned} \mathcal{E}(B_k) &= N \, (1 - (p_0 + p_1 \, + \, \ldots p_{k-1})) \, - \, N \, \theta \, = \, N \sum_{i=k}^{\infty} p_i \\ &\approx N \, p_k \, (1 + S'(\overline{\theta}) + S'(\overline{\theta})^2 \, + \, \ldots) \, = \, \frac{N p_k}{1 - S'(\overline{\theta})} \, . \end{aligned}$$

By the reasoning at the end of section 3 and the smallness of ξ, we have, even for n close to 1, $p_n \approx c \, \mathrm{e}^{-n/\xi}$, where c is some constant. Using this expression, we obtain

$$\mathcal{E}(B_k) \, \approx \, \frac{N \, c \, \mathrm{e}^{-k/\xi}}{1 - S'(\overline{\theta})} \, . \tag{18}$$

As $\mathcal{E}(B_0)$ is $O(N)$, then for $N \gg 1$ the number of generations τ such that $\mathcal{E}(B_\tau) = 1$ is given by

$$\tau \, = \, \xi \, (\ln N + \ln c) \approx \xi \ln N \, . \tag{19}$$

This is the typical time it takes for extinguishing all trees that will be extinct at some time. At this time, each non-extinct tree will have grown to a number of branches of order $2^\tau \, = \, O(N)$. Thus, at time τ, with large probability, all individuals at the initial population whose trees were not extinct will be ancestors to the whole population. As the trees which will be extinct sometime never grow too large, then the approximation of using GW to calculate τ is justified.

6 Conclusions and perspectives

We have shown that the GW process is useful in understanding phenomena in human evolution in a setting very similar to the Statistical Mechanics of critical phenomena, with the advantage of being exactly solvable. Within this model we found out that a mitochondrial Eve may exist even in an exponentially growing population and her existence constrains the mean number of children per woman to a narrow range. We also showed that independent GW processes are not a good model for joint existence of a mitochondrial Eve and a Y chromosome Adam. We provided some simulation results in this case, showing that with a correct choice of parameters, joint existence of Eve and Adam may occur with a sizeable probability.

One tacit assumption in our analysis, supported by biologists, is that all W original mtDNA lineages are equally fit, i.e. there is no natural selection acting on lineage sorting. It also follows from our results that within the range of values for $\overline{N}$ in which a mitochondrial Eve is likely, there is a probability of at least 63% that the number of surviving mtDNA lineages is different from 1. As the probability that this number is 0 is not negligible, see figure 4, we may explain extinction of other hominid species which have existed for some periods, if they had demography similar to our own.

Returning to the two competing models on human origins, we should say that the existence of an African mitochondrial Eve only proves that a sizeable part of early mankind did originate in that continent, but not the whole of it. mtDNA lineages originated somewhere else may have simply become spontaneously extinct, which in our analysis appears as a highly probable event even in a growing population.

While preparing this paper, we discovered another application for our methods. Unlike most other animals, in which the sex of the offspring is genetically determined (X and Y chromosomes in mammals, for example), the sex of the offspring in some living reptile species depends on the temperature during egg incubation. Although a definitive proof still lacks, Miller, Summers and Silber [24] conjecture that the same sex determination by temperature might hold for dinosaurs. Should that be true, a predominance of males could have arisen as a consequence of worldwide climate change after the impact of a large meteor. By numerically solving a mathematical model based on differential equations, they show that in this case dinosaur populations would decrease with time. If the sex skew were not too severe or lasted only for a short time, populations would start growing again.

We notice here that the GW model can also easily account for this phenomenon. In fact, an increase in male proportion is equivalent to lowering the

value of p, while p_c is held constant. On the other hand, the analysis carried out in [24] seems to disregard the fact that even in the worst cases, their model foresees population increase some time after an initial decrease. Here we see an important difference between their population modelling with differential equations and our "discrete" model. Differential equations are deterministic because they use the mean behavior for all individuals in the population. Extinction occurs in their model because of decrease in population to less than one individual. On the other hand, our model accounts for statistical fluctuations. If $p = 1/2$ and $2 < \overline{N} < 1/p$, population will increase in average, but extinction will still happen with positive probability due to statistical fluctuations. In other words, modelling populations with differential equations is analogous to mean-field theories in Statistical Mechanics, which we know can lead to wrong results.

We finally note that as Monte Carlo simulations are being increasingly used in Genetics, it is important to use the methods from the Physics of critical phenomena to better understand results derived from these simulations.

References

[1] T. E. Harris, The Theory of Branching Processes, Dover, New York, 1989.

[2] J. B. S. Haldane, A mathematical theory of natural and artificial selection. Part V: selection and mutation, Proc. Camb. Phil. Soc. 26 (1927) 838.

[3] J. C. Avise, J. E. Neigel, J. Arnold, Demographic influences on mitochondrial DNA lineage survivorship in animal populations, J. Molec. Evol. 20 (1984) 99.

[4] S. M. de Oliveira, G. A. de Medeiros, P. M. C. de Oliveira, D. Stauffer, Studying the number of lineages through Monte Carlo simulations of biological ageing, Int. J. Mod. Phys. C 9 (1998) 809.

[5] P. M. C. de Oliveira, S. M. de Oliveira, K. P. Radomski, Simulating the mitochondrial DNA inheritance, Theory Biosci. 120 (2001) 77.

[6] P. M. C. de Oliveira, Evolutionary computer simulations, Physica A 306 (2002) 351.

[7] M. Sitarz, A. Maksymowicz, Divergent evolution paths of different genetic families in the Penna model, to appear in Int. J. Mod. Phys. C.

[8] J. T. Chang, Recent common ancestors of all present-day individuals, Adv. Appl. Probab. 31 (1999) 1002.

[9] B. Derrida, S. C. Manrubia, D. H. Zanette, Statistical properties of genealogical trees, Phys. Rev. Lett. 82 (1999) 1987.

[10] B. Derrida, S. C. Manrubia, D. H. Zanette, Distribution of repetitions of ancestors in genealogical trees, Physica A 281 (2000) 1.

[11] B. Derrida, S. C. Manrubia, D. H. Zanette, On the genealogy of a population of biparental individuals, J. theor. Biol. 203 (2000) 303.

[12] R. L. Devaney, A First Course in Chaotic Dynamical Systems, Addison Wesley, Reading, 1992.

[13] J. C. Avise, Mitochondrial DNA and the evolutionary genetics of higher animals, Phil. Trans. R. Soc. Lond. B 312 (1986) 325.

[14] R. L. Cann, M. Stoneking, A. C. Wilson, Mitochondrial DNA and human evolution, Nature 325 (1987) 31.

[15] A. C. Wilson, R. L. Cann, The recent african genesis of humans, Sci. Am. 266 (4) (1992) 68.

[16] A. G. Thorne, M. H. Wolpoff, The multiregional evolution of humans, Sci. Am. 266 (4) (1992) 76.

[17] C. Duarte, J. Mauricio, P. B. Pettitt, P. Souto, E. Trinkaus, H. van der Plicht, J. Zilhao, The early upper paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia, Proc. Natl. Acad. Sci USA 96 (1999) 7604.

[18] W. M. Brown, Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis, Proc. Natl. Acad. Sci USA 77 (1980) 3605.

[19] C. Newell, Methods and models in Demography, John Wiley and Sons, Chichester, 1988.

[20] G. Grimmett, Percolation, 2nd Edition, Springer-Verlag, Berlin, 1999.

[21] N. Takahata, Allelic genealogy and human evolution, Mol. Biol. Evol. 10 (1) (1993) 2.

[22] C. Stringer, Out of Ethiopia, Nature 423 (2003) 692.

[23] P. A. Underhill, L. Jin, A. A. Lin, S. Q. Mehdi, T. Jenkins, D. Vollrath, R. W. Davis, L. L. Cavalli-Sforza, P. J. Oefner, Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography, Genome Research 7 (1997) 996.

[24] D. Miller, J. Summers, S. Silber, Environmental versus genetic sex determination: a possible factor in dinosaur extinction?, Fertility and Sterility 81 (4) (2004) 954.