Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 225 (2003) 497–505
Journal of Theoretical Biology
www.elsevier.com/locate/jtbi

Supercritical branching processes and the role of fluctuations under exponential population growth

Susanna C. Manrubia*, María Arribas, Ester Lázaro

Centro de Astrobiología INTA-CSIC, Ctra. de Ajalvir km. 4, E-28850 Torrejón de Ardoz, Madrid, Spain

Received 23 April 2002; received in revised form 22 July 2003; accepted 28 July 2003

Abstract

We study some exact properties of supercritical branching processes. A proper rescaling of the relevant variable allows us to determine the distribution of population sizes after a number of generations have elapsed. Both time-continuous and discrete processes are analysed and compared. The obtained results are of relevance for the growth of populations that are not resource limited (a typical situation in some biological processes that can be modelled by laboratory experiments). Large fluctuations inherent to the process play a main role when bottlenecks occur.

Keywords: Supercritical branching process; Population fluctuations; Polymerase chain reaction

1. Introduction

The application of branching processes to the analysis of biological, demographical and physical problems started more than a century ago, with the pioneering work of Galton and Watson (1874) on the extinction of surnames. This problem is equivalent to that of the extinction of a mutant allele in a population, although this relation was noted only much later (Fisher, 1922; Haldane, 1927). Currently, the theory of branching processes (Harris, 1963; Athreya and Ney, 1972) is a basic tool in the study of population genetics (Gale, 1990). Simultaneous advances in the description of the mathematical properties and of the possible applications of the theory have led to a better understanding of a large number of problems that can be described through multiplicative, branching dynamics (Feller, 1957). A non-exhaustive list includes chain reactions (Everett and Ulam, 1948), birth and death processes (Yule, 1924) and their application to self-similarity in taxonomy (Chu and Adami, 1999), queue theory (Kendall, 1951), extinction cascades (Flyvbjerg et al., 1993), fragmentation processes (Vlad, 1991) and human genealogy (Derrida et al., 1999).

Analytical techniques developed in the general framework of branching processes permit to calculate a number of statistical quantities for this class of systems, like the probability that the descendants of a single entity become eventually extinct or the distribution of progeny sizes. Subcritical and critical situations, where, on the average, the population keeps constant or decreases with time, have been more often explored. In particular, most results in population genetics are derived under the assumption that the population is finite and at equilibrium (Gale, 1990). The supercritical case, in which the total population can diverge, is more involved (Buhler, 1971, 1972; Sawyer, 1976; Lange and Fan, 1997). Still, since many real systems include non-stationary, often exponentially growing populations, exact results for the supercritical case could have, apart from their mathematical interest, an application to demographic experiments. For example, it has been recently shown that the observed distribution of family sizes (defined as the set of individuals sharing their surname), can be recovered through a model where the total population grows exponentially in time (Manrubia and Zanette, 2002). This seems to be an essential ingredient to quantitatively recover the observations.

In this work, we study a class of supercritical branching processes and obtain exact results for the distribution of descendants after an arbitrary number of generations have elapsed. In its discrete form, this

*Corresponding author. Tel.: +34-91-520-6425; fax: +34-91-520-1074.
E-mail address: cuevasms@inta.es (S.C. Manrubia).

0022-5193/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0022-5193(03)00294-7

problem was first tackled by Harris (1948), who gave it a formal solution. In time-continuous formulation (Athreya and Ney, 1972) it was solved by Kendall (1949). Nonetheless, very few exact results have been derived. Here we show that an appropriate rescaling of the relevant variable returns fixed point equations for statistical quantities both in the discrete- and continuous-time formulations, and allows in many cases an exact solution of the problem.

Our work is also motivated by the replicative properties of certain biological systems (such as viruses or bacteria) where the excess of resources allows an exponential growth in the total population number during a relatively long interval of time. Very often, the initial individual continues replicating during the whole length of the experiment. Hence, in the models to be discussed we take this fact into account by setting to zero the death probability of our elements. We will consider both discrete and time-continuous systems, and compare their properties under similar assumptions. This comparison is important in order to disentangle the mechanisms at play in real processes and their relevance in defining the statistical properties observed. Indeed, other recent studies have performed this comparison between the two genealogies (discrete and continuous) generated by a rare allele (Rannala, 1997) and by a multi-locus, finite population (Rogers and Prügel-Bennett, 2000).

We start by analysing a supercritical branching process evolving through discrete generations. At each generation, all the individuals previously present are included, plus their progeny, which is supposed to be Poisson distributed. This situation is analogous to the growth experienced by a viral population in standard laboratory experiments. Indeed, a model for the evolution of a viral population growing in the way described and later subjected to repeated bottleneck events has been successfully applied to the description of a number of laboratory experiments (Lázaro et al., 2002). Subsequently, and with the aim of linking the discrete and time-continuous processes, we study the case where each element either replicates or survives unchanged to the next generation. A good example of this type of processes is the replication of DNA by the polymerase chain reaction (PCR). This reaction allows the DNA target to replicate for a number of cycles, which are equivalent to the number of generations of the time-discrete branching processes. As a third example, and using the approach introduced in the first sections, the case of a pure-birth process is analysed. This dynamics closely corresponds to the case of bacterial growth, where duplication of each element happens stochastically, independently of the other elements, and at randomly chosen times.

The number of individuals present in the exponentially growing regime attains a stationary distribution for a properly rescaled variable. Our main result concerns the derivation of a number of analytical properties of that distribution. We study the three different processes introduced and compare the results. Further, we apply the rescaling procedure to experimental results obtained from DNA amplification by PCR and show that indeed a stationary distribution appears during the few cycles where the molecular population grows exponentially. Our results show that the large fluctuations observed in the yield of DNA obtained in different replicates cannot be simply ascribed to the amplification of differences in the initial composition of different samples.

Finally, as an additional illustration of the role of population fluctuations, we study the discrete, Poisson-distributed offspring model, and allow for mutations to occur. Statistical distributions for the fraction of mutants after a number of generations have elapsed are numerically obtained. We show that the presence of large fluctuations is relevant when interpreting repeated laboratory experiments and, if mutations are considered, play a main role when bottlenecks occur.

2. Time-discrete processes

The most suitable formalism for the study of branching processes is that of generating functions. Given a probability distribution $p(k)$, its generating function is defined as

$$F(s) = \sum_{k=0}^{\infty} p(k) s^k \tag{1}$$

and represents a different way of encoding all the information about $p(k)$. Indeed, a given probability distribution is completely characterized if we know all of its moments $\langle k^i \rangle$, which can be obtained by knowing up to the i-th derivative of $F(s)$ (Harris, 1963).

Consider a branching process which starts with a single individual at the 0-th generation. Suppose that the average growth of the population is m, such that the expected size of the population at generation g is $\langle n(g) \rangle = m^g$. Under iteration of the branching process, a distribution $p_g(n)$ stating the probability that $n(g)$ individuals are present at generation g develops. The variable $n(g)$ can be properly rescaled to a new variable $w(g)$,

$$w(g) = \frac{n(g)}{m^g} \tag{2}$$

with average value $\langle w(g) \rangle = 1$, and which attains a limiting stationary distribution $h(w)$ in the limit $g \to \infty$. The moment generating function for $h(w)$ is defined as

$$f(s) = \lim_{g \to \infty} f(s, g) \equiv \langle e^{s w_g} \rangle, \tag{3}$$

which satisfies

$$f(s, g+1) = F[f(s/m, g)]. \tag{4}$$

The function $F(s)$ is as defined in Eq. (1). In only a few cases can $h(w)$ be calculated explicitly. Some examples can be found in Harris (1963), and Cistyakov (1957) obtained the asymptotic form of the probabilities for w close to its average value $\langle w \rangle = 1$.

The function $f(s)$ attains an asymptotically stationary shape which allows the calculation of all the moments of $h(w)$,

$$\langle w^i \rangle = \left. \frac{\mathrm{d}^i f(s)}{\mathrm{d}s^i} \right|_{s=0}. \tag{5}$$

In the following, we study different prescriptions for $p(k)$ and analyse time-discrete and time-continuous models using this methodology.

2.1. Poisson distribution of offspring

The evolution of our system starts with a single individual at generation $g = 0$. The first generation is formed by its descendants and the individual itself. Hence, the probability $p(k)$ that there are k individuals present at $g = 1$ is

$$p(0) = 0, \quad p(k) = \frac{\mathrm{e}^{-(m-1)}(m-1)^{k-1}}{(k-1)!} \quad \text{for } k \geqslant 1, \tag{6}$$

where $m > 1$ is the average number of "offspring" per individual (including itself) after one generation, and represents the average growth rate of the population. The generating function for this probability distribution can be readily calculated:

$$F(s) = \sum_{k=1}^{\infty} \frac{\mathrm{e}^{-(m-1)}(m-1)^{k-1}}{(k-1)!} s^k \tag{7}$$

$$= s\mathrm{e}^{(m-1)(s-1)}. \tag{8}$$

The solution $s^* < 1$ to the equation $F(s) = s$ returns the probability that the system goes eventually extinct (Harris, 1963). In our case, and due to the deterministic addition of all previous elements to each next generation, the only solution is $s^* = 0$ (total extinction is not possible).

We are interested in the distribution of the progeny after many generations. We have defined as $p_g(n)$ the probability that n individuals are present after g generations. If $\langle n(g) \rangle$ is the average number of individuals at generation g, then $\langle n(g+1) \rangle = m\langle n(g) \rangle$ is the average size of the population at generation $g+1$. This suggests that a rescaling of the form

$$w(g) = \frac{n(g)}{m^g}, \tag{9}$$

$$h(w) = m^g p_g(w(n)) \tag{10}$$

(Harris, 1963) would produce a function $h(w)$ with an invariant shape under increasing g, and allow an evaluation of some of its limiting properties.[1] The new variable w represents the relative size of the population of descendants of the initial sequence with respect to the expected average m^g.

Let us now consider the moment generating function $f(s, g)$ of the variable w at generation g, which according to Eq. (4) and rescaling (9) satisfies

$$f(s, g+1) = \sum_{k \geqslant 0} p(k) \left[f\left(\frac{s}{m}, g\right) \right]^k, \tag{11}$$

where $p(k)$ is as defined in Eq. (6). Hence,

$$f(s, g+1) = \sum_{k \geqslant 1} \frac{\mathrm{e}^{-(m-1)}(m-1)^{k-1}}{(k-1)!} \left[f\left(\frac{s}{m}, g\right) \right]^k$$

$$= f\left(\frac{s}{m}, g\right) \mathrm{e}^{-(m-1)[1-f(s/m,g)]}. \tag{12}$$

The function $f(s, g)$ reaches a limit shape for $g \to \infty$, where it becomes independent of g,

$$f(s) = f\left(\frac{s}{m}\right) \mathrm{e}^{-(m-1)[1-f(s/m)]}. \tag{13}$$

This type of recursion often display scaling solutions.[2] We make the ansatz that, in the limit $s \to -\infty$ -that is for small w- the distribution $h(w)$ is a power law with an exponent β_1, such that the regular part of the moment generating function can be written in general as $[f(s) - s^*] \sim |s|^{-\beta_1 - 1}$. Due to the fact that the progenitor individual always survives to the next generation, the probability of eventual extinction of the population is, by definition, $s^* = 0$ in the current case. Substituting in Eq. (13) and expanding around small $f(s)$, we get $1 = m^{\beta_1 + 1}\mathrm{e}^{1-m}$, which predicts that $h(w) \simeq w^{\beta_1}$, with

$$\beta_1 = \frac{m-1}{\ln m} - 1. \tag{14}$$

The distribution $h(w)$ increases up to $w \simeq 1$, and then experiences a fast decrease (indeed faster than any exponential function, see Derrida et al., 2000).

In order to obtain the moments of the distribution $h(w)$, one can always try to solve Eq. (13) in powers of s and get the expected values $\langle w^i \rangle$. We obtain for the

[1] The random variables $w(g)$ are a martingale and, since their expected value is $\langle w \rangle = 1$ by construction, they converge to a random variable whether or not $\langle w^2 \rangle$ is finite and whether or not $m > 1$ (Harris, 1963).

[2] A demographic problem where a similar recursion equation appears is that of the distribution of ancestors of a present individual in a closed population (Derrida et al., 1999). Recently, Chang (1999) has proven that the solution to that problem for $w \to 0$ is indeed of the power-law type.

first orders in s,

$$f(s) = 1 + s + \frac{1+m}{2m}s^2 + \frac{3+4m+4m^2+m^3}{6m^2(1+m)}s^3 + \frac{15+13m+22m^2+14m^3+7m^4+m^5}{24m^3(1+m+m^2)} \times s^4 + O(s^5), \quad (15)$$

where the coefficient of the first order in s results from the normalization condition imposed in the rescaling, $\langle w \rangle = 1$.

2.2. Single offspring per generation

Let us repeat the analysis of the previous subsection considering that, at each generation, individuals in the population either split in two with probability $p(2) = \mu$ or just survive with probability $p(1) = 1 - \mu$. Other possibilities are zero ($p(0) = p(k) = 0,\ k > 2$). The fundamental equation, which is the recursion for the moment generating function of the rescaled variable $w = n/(1+\mu)^g$ in the limit $g \to \infty$ reads now

$$f(s) = \left[(1-\mu) + \mu f\left(\frac{s}{1+\mu}\right)\right] f\left(\frac{s}{1+\mu}\right). \quad (16)$$

Again, the system has zero probability of extinction, such that the singular part of $f(s) = s^* = 0$. Using the same scaling ansatz as previously, we get that for small w the distribution $h(w) \simeq w^{\beta_2}$, with

$$\beta_2 = -\frac{\ln(1-\mu)}{\ln(1+\mu)} - 1. \quad (17)$$

A solution in powers of s for Eq. (16) returns

$$f(s) = 1 + s + \frac{1}{1+\mu}s^2 + \frac{2}{(1+\mu)^2(2+\mu)}s^3 + \frac{6+\mu}{(1+\mu)^3(2+\mu)(3+3\mu+\mu^2)}s^4 + O(s^5). \quad (18)$$

It is possible to extract some information about the tail of $h(w)$, which corresponds to the regime $s \gg 1$, when the set of probabilities $p(k)$ is zero for $k > k_{max}$. Then, the moment generating function is a polynomial in s where only the last term contributes. For very large values of g and s, we can approximate Eq. (16) as

$$f(s) \sim \mu \left[f\left(\frac{s}{1+\mu}\right)\right]^{k_{max}}, \quad (19)$$

where, in our case, $k_{max} = 2$. If we define $g(s) \equiv \ln f(s)$ and take logarithms in the previous expression, we get

$$g(s) \sim 2g\left(\frac{s}{1+\mu}\right) \quad (20)$$

which has a scaling solution of the form $g(s) \propto s^{\alpha'}$, with

$$\alpha' = \frac{\ln 2}{\ln(1+\mu)}. \quad (21)$$

Consequently, $f(s)$ becomes

$$f(s) \propto \exp\{s^{\ln 2/\ln(1+\mu)}\}. \quad (22)$$

Once more, for very large values of s, the integral which relates $f(s)$ to $h(w)$ can be calculated through the Laplace method, since only the value of w which makes the integrand maximal (w^*) contributes,

$$f(s) = \int_0^\infty e^{sw} h(w)\, dw \simeq \max_w [e^{sw} h(w)]. \quad (23)$$

If we now assume that

$$h(w) \propto \exp\{-w^\alpha\}, \quad (24)$$

then $\ln f(s) \propto s^{\alpha'} \propto \max_w [sw - w^\alpha]$. The maximal value is attained for $w^* = (s/\alpha)^{1/(\alpha-1)}$, and hence

$$s^{\alpha'} \simeq s\left(\frac{s}{\alpha}\right)^{1/(\alpha-1)} - \left(\frac{s}{\alpha}\right)^{\alpha/(\alpha-1)}, \quad (25)$$

from where

$$\alpha = \frac{\ln 2}{\ln[2/(1+\mu)]}. \quad (26)$$

In general, for arbitrary k_{max}, the exponent above turns out to be

$$\alpha = \frac{\ln(k_{max})}{\ln(k_{max}/(1+\mu))}. \quad (27)$$

For the system to be nontrivial, $k_{max} \geqslant 2$, and $\mu < 1$. With these bounds to the parameters, one can see that $\alpha > 1$, and thus the tail of the distribution $h(w)$ decays in all cases faster than exponentially.

We can try to approximate the function $h(w)$ in the whole range of w through a composition of the obtained initial power law plus the stretched exponential tail. After normalising the function to unity, we get

$$h(w) = \frac{\alpha}{\Gamma((\beta_2+1)/\alpha)} w^{\beta_2} e^{-w^\alpha} \quad (28)$$

with the exponents β_2 and α given by Eqs. (17) and (26), respectively. Thanks to $h(w)$ being peaked around $w = 1$, the two asymptotic results for $w \to 0$ and $w \to \infty$ approximate well the whole function. In Fig. 1 we compare numerical simulations of the process with the analytical result (28).

3. Time-continuous process

The continuous-time representation of the replication process can be obtained by defining a replication rate per unit time μ, such that, in a small time-interval Δ, the probability of replication is $p(2) = \mu\Delta$, and that of remaining unchanged is $p(1) = 1 - \mu\Delta$. The population would grow exponentially at the rate μ, such that after a time t, $\langle n(t) \rangle = e^{\mu t}$. Hence, the relevant rescaled variable is now $w = n/e^{\mu t}$. For a small time interval $\Delta \to 0$ we can take a first-order approximation for the population growth, $\langle n(t+\Delta) \rangle = \langle n(t) \rangle (1 + \mu\Delta)$, such

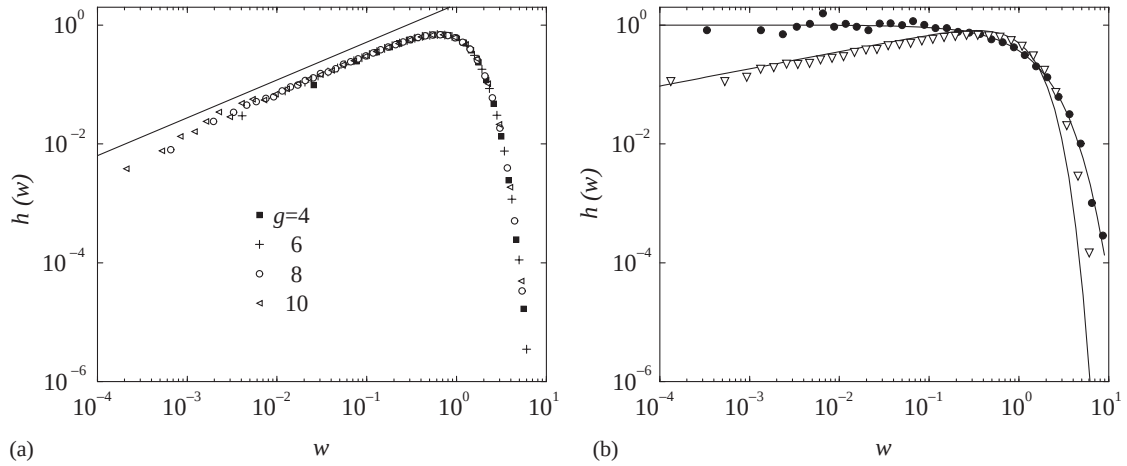

Fig. 1. Distribution functions of the rescaled variable w for the three models studied in this work. (a) Distributions obtained in the discrete-time process with a Poisson distribution of offspring after 4, 6, 8, and 10 generations. We observe that, once the rescaling has been applied, these functions collapse on a single curve $h(\ln(w))$. The solid line shows the analytic result $h(w) \simeq w^{\beta}$, which holds for small w. In this case, $m = 2.5$. (b) Distributions obtained in the discrete-time process with simple replication per generation (with $\mu = 0.25$, triangles) and in the continuous-time process ($\mu = 0.2$, filled circles). Symbols stand for numerical simulations, solid lines for the analytical results (28) and (33).

that the recursion equation for the moment generating function becomes

$$f(s, t + \Delta) = (1 - \mu\Delta) f\left(\frac{s}{1 + \mu\Delta}, t\right) + \mu\Delta \left[f\left(\frac{s}{1 + \mu\Delta}, t\right)\right]^2. \quad (29)$$

The continuous-time picture amounts to performing the limit $\Delta \to 0$ in the previous expression. Developing to first order in Δ we obtain a partial differential equation for $f(s, t)$:

$$\frac{\partial f(s,t)}{\partial t} = \mu f^2(s,t) - \mu f(s,t) - s\mu \frac{\partial f(s,t)}{\partial s}. \quad (30)$$

Once more, we can take advantage of the fact that the solution to our problem is the time-independent function $f(s)$ obtained from the fixed point $\partial f(s, t)/\partial t = 0$,

$$s \frac{\mathrm{d} f(s)}{\mathrm{d} s} = f(s)[f(s) - 1] \quad (31)$$

which has the exact solution

$$f(s) = \frac{1}{1 - s}, \quad (32)$$

and where we have already included the normalization condition $f'(s = 0) = \langle w \rangle = 1$. Performing an inverse Laplace transform on Eq. (32) we immediately obtain the distribution $h(w)$ for the variable w,

$$h(w) = \mathrm{e}^{-w}. \quad (33)$$

A different derivation of this result was already carried out by Kendall (1949).

4. Comparison between the models

The three models studied can be easily simulated in a computer. Both the analytical and computational results reveal that they are very similar qualitatively. All the obtained distributions have an initial power-law shape before reaching a maximum close to the value $w = 1$, after which a fast decaying tail appears. In the time-continuous case, the tail is a pure exponential function; in any discrete case where $p(k > k_{max}) = 0$ it is a stretched exponential with exponent larger than unity; if the distribution $p(k)$ is positive for all k, the tail decays slower than a stretched exponential of the previous kind but faster than any exponential function. Note that the distribution in the continuous-time representation could have been obtained by performing the limit $\mu \to 0$ in Eq. (28), which implies $\beta \to 0$, $\alpha \to 1$ and recovers Eq. (33). It is interesting that the time-continuous process has a distribution $h(w)$ independent of the replication rate μ.

The non-stationary distributions $p_g(n)$ can be readily recovered by inverting, in each case, the rescaling performed. This does not add further information on the process to the results obtained by studying $h(w)$. In the time-continuous case, the dependence on μ (which in this case only affects the growth of the population and not the moments of the distribution), enters through a trivial multiplicative factor to $h(w)$.

Although these supercritical processes have a well-defined average at any moment in time, there are finite differences between realizations also in the limit $g, t \to \infty$, as is pointed out by the fact that the variance $\sigma^2 = \langle w^2 \rangle - \langle w \rangle^2$ remains positive,

$$\sigma_1^2 = \frac{1}{m}, \quad \sigma_2^2 = \frac{1 - \mu}{1 + \mu}, \quad \sigma_3^2 = 1 \quad (34)$$

(the subindexes stand for the three models according to the order of introduction in the text) and comparable to the average of the process, which is unity in the rescaled variable.

This has important implications when working with multiplicative processes of this type, as for instance in laboratory populations, which often experience an exponential growth. The final output of an experiment depends strongly on population fluctuations (particularly relevant in the initial time steps), and repetition of the same procedure might lead to essentially different results. The fluctuations of this type of processes are non-Gaussian and the errors obtained with few realizations can be "unexpectedly" large.

However, we have addressed up to now the "worse behaved" situation, since starting with a single individual at each generation maximizes population fluctuations and variance (34) of the process. An initial condition with a larger number I of individuals softens the effect of fluctuations: the variable $n(g)$ (number of individuals at generation g) would now be the sum of I independent, identically distributed variables. Hence, in the limit of a very large initial population the final distribution approaches a Gaussian curve.

5. Population fluctuations in PCR

Polymerase chain-reaction produces high yields of DNA starting with a low initial number of DNA molecules (Mullis and Faloona, 1987). The reaction is carried out through a discrete amplification process that involves the replication of DNA for a number of cycles (usually less than 40). Each cycle consists of the following steps: (i) heating at 94° to get the separation of the two single strands of the DNA double helix, (ii) hybridization of the oligonucleotides used as primers for DNA synthesis, and (iii) elongation of the primers to get a copy of the single DNA strands. Simultaneous with DNA synthesis, double strands unable to replicate (until the process of denaturization takes place again in the next cycle) are being generated. Therefore, we can think of each PCR cycle as corresponding to one generation of a discrete-time branching process.

Recently, procedures allowing to follow the yield of the PCR reaction on a cycle-by-cycle basis have been developed (Wittwer et al., 1997). One of these, to be used here, is based in the detection of the increase of fluorescence that takes place when a double strand DNA specific dye (SYBR Green I) is included in the assay.

At each cycle, the DNA content is ideally doubled (this means 100% efficiency, $\mu = 1$). In ideal conditions, differences in the final yield can only be ascribed to differences in the initial concentration. In order to observe the role played by population fluctuations (as described in the theoretical models here presented), it is convenient to carry out the PCR experiment under conditions involving the replication of a fraction of the total number of molecules ($\mu < 1$). The lower the μ, the higher the difference between two independent realizations of the process. Some conditions that may decrease the efficiency of the reaction include deviations from the optimal concentration of substrates determined empirically (e.g. deoxyribonucleotide triphosphates and magnesium ions). In the experiment carried out in this work, we have amplified a 217 bp fragment (corresponding to the sequence from nucleotide 1002 to nucleotide 1218 of the cDNA of foot-and-mouth disease virus) using a reaction mixture containing 1 mM of $MgCl_2$, which renders a reaction efficiency of 60% (see Fig. 2) and an average C_t value of 22.7 ± 0.5.[3] In general, and for a fixed value of the reaction efficiency, the lower the

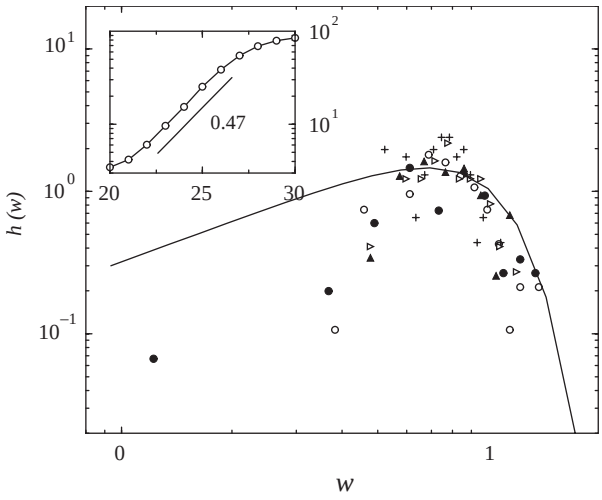

Fig. 2. Distribution of the background subtracted relative fluorescence obtained in the exponentially growing phase (cycles 22–26) of the real-time amplification of the 217 bp DNA fragment. The 96 replicates of the reaction had an initial amount of DNA target of around 10^5 molecules in 50 μl of a buffer that contains 100 μM of each deoxyribonucleotide triphosphates; 10 mM Tris-HCl, pH 8.3; 50 mM KCl, 1 mM $MgCl_2$, 5 units of Taq Gold Polymerase (Perkin-Elmer), SYBR Green I (Molecular Probes, 1:75,000 dilution of the $10,000\times$ stock solution), and 12 pmol of the two flanking primers. Cycling conditions were 94° for 20 s, 64° for 20 s, and 68° for 1 min, after an initial step of enzyme activation of 5 min at 95°. Increase of fluorescence was followed using a real time PCR detection system (iCycler Thermal Cycler from Bio-Rad). Symbols represent experimental results corresponding to the DNA yields obtained in five consecutive amplification cycles. Experimental results are compared with the solution of the time-discrete model with duplication at a rate $\mu = 0.6$ per generation. The inset displays the measured average fluorescence for cycles 20–30. The slope of the line gives the population growth rate in the exponential regime, $0.47 = \ln(1 + \mu)$, and from it we can derive the rescaling coefficient $1 + \mu = 1.6$, and the efficiency of the process, which in this case is around 60%.

[3] The threshold cycle C_t corresponds to the cycle at which fluorescence starts to be detectable.

amount of DNA, the higher the C_t value. Too high C_t values (corresponding to a low number of initial molecules) may return unreliable results, due to the generation of unspecific products and primer dimers contributing to fluorescence. To circumvent this problem, the starting number of DNA molecules in our reactions was of order $I = 10^5$. In these conditions, the increase of fluorescence in the exponential phase of the reaction corresponds to the replication of the specific product. Therefore, PCR corresponds to those cases in which I cannot be lowered arbitrarily, and thus variances well below those of the $I = 1$ process are expected.

The obtained results agree with our previous expectations. We have analysed the population distribution of 96 replicates from cycles 22 to 26, where the population was growing at an exponential rate (see inset of Fig. 2). The average increase of fluorescence over the independent realizations fixes the value $\mu \approx 0.6$ that we will use in the rescaling of the distributions at subsequent cycles. The data collapse of the five consecutive cycles cited is displayed in Fig. 2. We observe that, indeed, the distribution of scaled population sizes attains a stationary shape which, in this case, is broader than a Gaussian but significatively narrower than the theoretical curve corresponding to the case $I = 1$ (solid line in Fig. 2).

A very interesting kinetic model for PCR was developed and analysed by Stolovitzky and Cecchi (1996). In a certain regime, they obtained a model equivalent to the one here discussed. Their detailed analysis of the role played by the size of the initial population nicely showed that a transition from a broad distribution with a scaling part to a Gaussian one takes place as the number of starting molecules increases.

6. Role of mutations

Let us finish by analysing the spreading of neutral mutations in a population developed from a single individual. Some previous studies have analysed the distribution of point mutations in models mimicking PCR dynamics (Sun, 1995; Wang et al., 2000). Here we will consider the discrete-time process with a Poisson distribution of offspring described in Section 2.1 which represents the growth of a viral population. The simplest way to include mutations in its evolution consists in using an infinite genome approximation and suppose that mutations in a nucleotide occur with probability p and are neutral. This simply means that we attach a label to each of the individuals saying how many mutations it carries (not affecting its replication rate, that is its fitness). We are interested in the fraction r_i of individuals with i mutations after g generations. These fractions are defined as the quotient between the total number of individuals carrying i mutations and the total population in the generation considered.

In this case, we are dealing with a multitype branching process. Each of the lineages produced by a new mutant is a Poisson branching process independent of the evolution of the rest of the population. This fact permits to obtain a number of analytic properties (Kingman, 1993; Lange and Fan, 1997). Here, we are interested in the distributions of the fraction of mutants, $q(r_i)$ for each type. These quantities are relevant, for instance, when a population experiences a bottleneck. Its subsequent evolution depends critically on the subpopulation selected. Thus it is important to know what are the statistical properties of small subsets of individuals generated by a realization of the process. We perform the evolution on a flat fitness landscape (the average number of offspring is not affected by the mutations).

We have studied the previous system numerically and present results for a representative case with parameters $p = 0.01$ (mutation probability) and $m = 4$ at generation $g = 8$ (see Fig. 2). With the parameters chosen, the final population is formed by a large fraction of individuals identical to the seed (as evidenced by the large weight of $q(r_1)$ close to $r_1 = 1$) and by a varying, typically smaller, fraction of mutants. The probability that around 15% of the individuals in the last generation are mutants is about 0.13. Although, after so few generations, mutants are comparably less represented than forms identical to the initial individual, we observe that the distribution of their relative presence when the experiment finishes is extremely broad, and varies along several orders of magnitude. This evidences that repetition of the experiment under the same conditions can lead to very different compositions of the final population.

7. Conclusions

We have discussed three different supercritical branching models with the aim of studying the effect of population fluctuations under exponential growth of the number of individuals. The distribution for the expected number of individuals when the experiment finishes has been analytically calculated. We observe that this distribution is very broad, such that two independent realizations typically differ in an amount comparable to the expected mean if, initially, a single individual started the process. Our results are of relevance when studying different populations in their exponentially growing phase. When neutral mutations are added to the process, the composition of the final population suffers again of large fluctuations in the number of each mutant type present (Fig. 3).

The models here discussed are too simple to quantitatively characterize real processes. We have seen an experimental example where violation of the strict

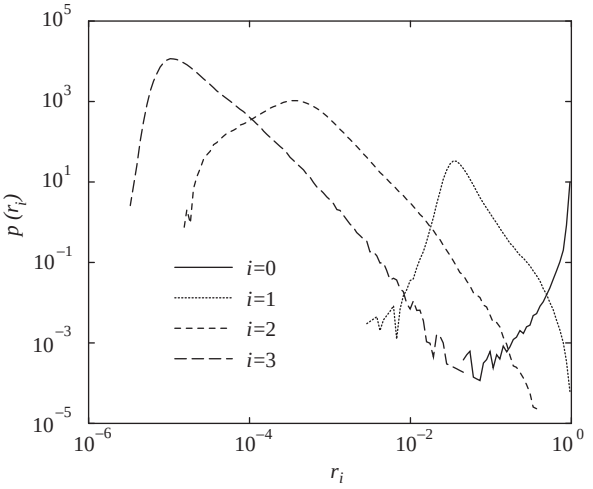

Fig. 3. Distributions of the fraction of each type of mutant. See main text.

condition $I = 1$ results in a reduction of the relative fluctuations. Due to the critical contribution of the initial population size to the fluctuations observed at the end of the branching process, it would be desirable to have an experimental system showing exponential growth and starting with $I = 1$. The development of a bacterial colony from an individual could be one of these systems. Experiments are being carried out to assess whether fluctuations in the colony areas (assumed to be proportional to the total number of cells) can be explained with the models developed in this work. In the case of viral growth, it is indeed possible to start the experiment with a single infecting particle (Escarmís et al., 2002). It would be interesting to know if the distribution of population sizes in that case agrees with that for the model here studied. In lytic viruses, this distribution could be easily estimated through the calculation of the areas of lytic plaques at a given time.

The introduction of correlated fitness landscapes (where natural selection would act to favor fitter variants) would be also interesting to analyse from the theoretical viewpoint, in order to get closer to real processes. In addition, neither discrete generation representations nor homogeneous, continuous-time processes are expected to accurately represent real replication dynamics. Nevertheless, we have derived a number of qualitative features which are expected to hold also in more realistic models (including for instance age-dependent replication).

The models here studied belong to a class known as *non-self-averaging* processes. That is, each realization of the system keeps memory of the fluctuations it has experienced along its history, and time averages do not compensate for them. This fact was already pointed out by Harris (1963), and implies that a whole statistical characterization of the process requires averaging over independent realizations. This effect has been also studied in relation to the genetic variability expected in natural populations (Derrida and Peliti, 1991; Higgs, 1995).

Some laboratory experiments subject populations to repeated bottlenecks in order to "guide" its evolution or to study how it would react in a similar natural environment (Escarmís et al., 1996, 2002; Lázaro et al., 2002). It is to be expected that evolution under repeated bottleneck passages reflects the non-self-averaging nature of the process and produces large deviations from the characteristics of the initial element, and time histories macroscopically different from realization to realization.

Acknowledgements

The authors acknowledge B. Derrida for discussions, and J. Pérez-Mercader for promoting interdisciplinary research at CAB. Financial support from INTA, CAM, and the Spanish Ministerio de Ciencia y Tecnología (SCM benefits from a Ramón y Cajal contract) is acknowledged.

References

Athreya, K.B., Ney, P.E., 1972. Branching Processes. Springer, New York.

Buhler, W., 1971. Generations and the degree of relationship in a supercritical Markov branching process. Z. Wahrschein. Verw. Gebiete 18, 141–152.

Buhler, W., 1972. The distribution of generations and other aspects of the family structure of branching processes. Proceedings of the Sixth Berkeley Symposium on Mathematics, Statistics and Probability, Vol. 3, Berkeley, CA, pp. 463–480.

Chang, J.T., 1999. Recent common ancestors of all present-day individuals. Adv. Appl. Prob. 31, 1002–1026.

Chu, J., Adami, C., 1999. A simple explanation for taxon abundance patterns. Proc. Natl. Acad. Sci. USA 96, 15,017–15,019.

Cistyakov, V.P., 1957. Local limit theorems of the theory of branching random processes. Theory of Prob. Appl. 2, 360–374.

Derrida, B., Peliti, L., 1991. Evolution in a flat fitness landscape. Bull. Math. Biol. 53, 355–382.

Derrida, B., Manrubia, S.C., Zanette, D.H., 1999. Statistical properties of genealogical trees. Phys. Rev. Lett. 82, 1987–1990.

Derrida, B., Manrubia, S.C., Zanette, D.H., 2000. Distribution of repetitions of ancestors in genealogical trees. Physica A 281, 1–16.

Escarmís, C., Dávila, M., Charpentier, N., Bracho, A., Moya, A., Domingo, E., 1996. Genetic lesions associated with Muller's ratchet in an RNA virus. J. Mol. Biol. 264, 255–267.

Escarmís, C., Gómez-Mariano, G., Dávila, M., Lázaro, E., Domingo, E., 2002. Resistance to extinction of low fitness virus subjected to plaque-to-plaque transfers: diversification by mutation clustering. J. Mol. Biol. 315, 647–661.

Everett, C.J., Ulam, S., 1948. Multiplicative systems, I. Proc. Natl. Acad. Sci. USA 34, 403–405.

Feller, W., 1957. An Introduction to Probability Theory and its Applications. Wiley, New York.

Fisher, R.A., 1922. On the dominance ratio. Proc. R. Soc. Edinburgh 42, 321–341.

Flyvbjerg, H., Sneppen, K., Bak, P., 1993. Mean field theory for a simple model of evolution. Phys. Rev. Lett. 71, 4087–4090.

Gale, J.S., 1990. Theoretical Population Genetics. Unwin Hyman, London.

Galton, F., Watson, H.W., 1874. On the probability of the extinction of families. J. R. Anthropol. Inst. 4, 138–144.

Haldane, J.B.S., 1927. A mathematical theory of natural and artificial selection. Part V: Selection and mutation. Proc. Cambridge Philos. Soc. 26, 838–844.

Harris, Th.E., 1948. Branching processes. The Annal of Math. Stat. 19, 474–494.

Harris, Th.E., 1963. The Theory of Branching Processes. Springer, Berlin.

Higgs, P.G., 1995. Frequency distributions in population genetics parallel those in statistical physics. Phys. Rev. E 51, 95–101.

Kendall, D.G., 1949. Stochastic processes and population growth. J. Roy. Stat. Soc. B 11, 230–264.

Kendall, D.G., 1951. Some problems in the theory of queues. J. Roy. Stat. Soc. B 13, 151–185.

Kingman, J.F.C., 1993. Poisson Processes. Oxford Science Publications, Oxford.

Lange, K., Fan, R.Z., 1997. Branching process models for mutant genes in nonstationary populations. Theor. Popul. Biol. 51, 118–133.

Lázaro, E., Escarmís, C., Domingo, E., Manrubia, S.C., 2002. Modeling viral genome fitness evolution associated with serial bottleneck events: evidence of stationary states of fitness. J. Virol. 76, 8675–8681.

Manrubia, S.C., Zanette, D.H., 2002. At the boundary between biological and cultural evolution: The origin of surname distributions. J. Theor. Biol. 216, 461–477.

Mullis, K.B., Faloona, F.A., 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155, 335–350.

Rannala, B., 1997. On the genealogy of a rare allele. Theor. Popul. Biol. 52, 216–223.

Rogers, A., Prügel-Bennett, A., 2000. Evolving populations with overlapping generations. Theor. Popul. Biol. 57, 121–129.

Sawyer, S.A., 1976. Branching diffusion processes in population genetics. Adv. Appl. Prob. 8, 659–689.

Stolovitzky, G., Cecchi, G., 1996. Efficiency of DNA replication in the polymerase chain reaction. Proc. Natl. Acad. Sci. USA 93, 12947–12952.

Sun, F., 1995. The polymerase chain reaction and branching processes. J. Comput. Biol. 2, 63–86.

Vlad, M.O., 1991. A stochastic renormalization approach to multifragmentation. Phys. Lett. A 160, 523–527.

Wang, D., Zhao, C., Cheng, R., Sun, F., 2000. Estimation of the mutation rate during error-prone polymerase chain reaction. J. Comput. Biol. 7, 143–158.

Wittwer, C.T., Hermann, M.G., Moss, A.A., Rasmussen, R.P., 1997. Continuous fluorescence monitoring of rapid cycle DNA amplification. Biotechniques 22, 130–131.

Yule, G.U., 1924. A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, F.R.S. Philos. Trans. R. Soc. B 213, 21–87.