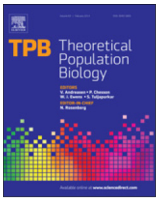
Theoretical Population Biology xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb



Mutation in populations governed by a Galton–Watson branching process

Conrad J. Burden[a,b,*], Yi Wei[a]

[a] Mathematical Sciences Institute, Australian National University, Canberra, Australia
[b] Research School of Biology, Australian National University, Canberra, Australia

ARTICLE INFO

Article history:
Received 17 July 2017
Available online xxxx

Keywords:
Mutation
Galton–Watson
Multitype branching process

ABSTRACT

A population genetics model based on a multitype branching process, or equivalently a Galton–Watson branching process for multiple alleles, is presented. The diffusion limit forward Kolmogorov equation is derived for the case of neutral mutations. The asymptotic stationary solution is obtained and has the property that the extant population partitions into subpopulations whose relative sizes are determined by mutation rates. An approximate time-dependent solution is obtained in the limit of low mutation rates. This solution has the property that the system undergoes a rapid transition from a drift-dominated phase to a mutation-dominated phase in which the distribution collapses onto the asymptotic stationary distribution. The changeover point of the transition is determined by the per-generation growth factor and mutation rate. The approximate solution is confirmed using numerical simulations.

1. Introduction

Since their introduction to the field by Haldane (1927), Galton–Watson (GW) branching processes have been an important part of the population genetics landscape (Patwa and Wahl, 2008). For example, probabilities of non-extinction derived through branching process approximations play an indispensable role in many complex population models (e.g. Desai and Fisher, 2007). However, as argued by Mode et al. (2013), the influence of models based on GW branching processes has in general been overshadowed, at least within the text book literature, by that of Wright–Fisher (WF) based models. Much of the WF model's dominance can be attributed to the intuitive appeal of the coalescent (Kingman, 1982), which is a natural consequence of WF models but mathematically formidable for a GW process (Lambert et al., 2013), and to the WF model's well-known diffusion limit via the forward Kolmogorov equation, as championed by Kimura (1955a, b, 1964).

Somewhat lesser known than the work of Kimura, and predating it by four years, is a solution to the diffusion limit of a GW branching process published by Feller (1951a). It is surprising that, although Feller's solution was presented in the context of genetics, the vast majority of applications of Feller's solution have been to areas other than genetics (see Gan and Waxman, 2015 and references therein). It is equally surprising that when population genetics per se is modelled as a branching process, it is generally as a discrete state space simulation (Mode and Sleeman, 2012; Cyran and Kimmel, 2010) or a continuous birth–death process (Stadler et al., 2015), without reference to Feller's diffusion limit.

This paper follows on from an earlier work (Burden and Simon, 2016) in which Feller's diffusion limit is exploited to study genetic drift in haploid populations governed by a GW branching process. In that work it was shown that, in the absence of mutations and selection, expected fixation times and probabilities of fixation for a critical branching process match those of the WF model. However, for a supercritical branching process there is a finite probability that an allele will never fix. The dynamics of the branching process enabled an estimate to be made of the time since the most recent common ancestor of an extant population, for instance, mitochondrial Eve.

The current paper extends the branching model to a multi-allelic population with mutations, and is equivalent to a multitype branching process (Mode, 1971; Haccou et al., 2005). Multitype branching processes have been applied in population science to modelling cancers (Durrett and Moseley, 2010; Iwasa et al., 2003), modelling bacterial cultures (Wahl and Dai Zhu, 2015), and in ecological modelling (Antia et al., 2003; Caswell, 2001, Chapter 15).

Our model is set out in detail in Section 2, and the diffusion limit forward Kolmogorov equation is derived in Section 3. Our choice of diffusion limit is such that continuum time is scaled by the log of the per-generation growth factor λ, and the population size is scaled by the mean exponential growth. This leads to a slightly more elegant forward Kolmogorov equation than Feller's original, but with the same physical interpretation (see Eq. (27)).

---

* Corresponding author at: Mathematical Sciences Institute, Australian National University, Canberra, Australia.
E-mail addresses: conrad.burden@anu.edu.au (C.J. Burden), u5166590@anu.edu.au (Y. Wei).

https://doi.org/10.1016/j.tpb.2017.12.001
0040-5809/

Our scaling has the disadvantage that it is not suitable for critical growth, $\lambda = 1$, thus limiting our analysis to the supercritical case. On the other hand it has the advantage that the solution is classified in terms of a 1-parameter family of density functions (see Eq. (31) and (32)).

In Section 4 Feller's method of solution via a Laplace transform for the 1-allele case is briefly summarised in order to facilitate analysis of the case of non-zero mutations in Sections 5 and 6. Although we are unable to find a complete analytic solution, we are able to obtain the asymptotic stationary solution for the case of 2 alleles, and also an approximate solution for all times in the biologically realistic limit of low mutation rates. An interesting result is that the solution undergoes a rapid changeover in behaviour from a perturbation on the zero-mutation solution to an asymptotic collapse onto a state in which any extant population partitions into subpopulations in proportions determined by mutation rates. Section 7 is devoted to numerical simulations to confirm our analytical results, and to confirm consistency of the model with mitochondrial genomic data. Section 8 is devoted to a discussion and conclusions.

2. The model

We consider a population of $M(t)$ haploid individuals which are assumed to reproduce in discrete, non-overlapping generations $t = 0, 1, 2, \ldots$. The population is subdivided at any generation into K allele types, and the number of copies of type i within the population is $Y_i(t)$. Thus

$$\sum_{i=1}^{K} Y_i(t) = M(t). \tag{1}$$

The individuals are assumed to reproduce according to a GW process whereby the number of offspring per individual of allele type i is given by a set of identically and independently distributed (i.i.d.) random variables $S_\alpha^{(i)}$, $\alpha = 1, \ldots, Y_i(t)$, whose common distribution is denoted by a generic non-negative integer valued random variable $S^{(i)}$ with mean and variance

$$E(S^{(i)}) = \lambda_i, \qquad \mathrm{Var}\,(S^{(i)}) = \sigma_i^2, \qquad i = 1, \ldots, K. \tag{2}$$

Furthermore the alleles are assumed to undergo random mutations from type i to type j at a rate u_{ij} per individual per generation, where

$$u_{ij} \geq 0, \qquad \sum_{j=1}^{K} u_{ij} = 1. \tag{3}$$

A single time step is illustrated in Fig. 1.

Define the number offspring born to parents of allele type i in generation t to be

$$W_i = \sum_{\alpha=1}^{Y_i(t)} S_\alpha^{(i)}. \tag{4}$$

During its lifetime the new generation undergoes mutations, culminating in a new mature generation in which the number of individuals of type i is expressible as a sum of random variables[1]

$$Y_i(t+1)|\mathbf{W} = V_{1i} + V_{2i} + \cdots V_{Ki}, \tag{5}$$

where V_{ji} is the number of individuals who begin life as allele type j and mature to become allele type i. For fixed parental type j the V_{ji} have a multinomial distribution:

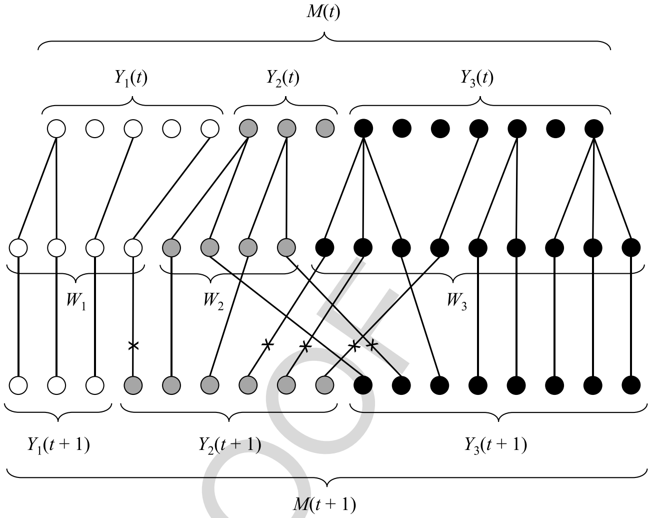


Fig. 1. One time step of the GW model with mutations: At time step t a population of $M(t)$ individuals is partitioned into subsets containing $Y_i(t)$ individuals of allele type i. Each individual generates a random number of offspring of the same allele type as its parent, and the number of offspring initially of type i is defined as W_i. Individuals may mutate during their lifetime to create the new generation containing $Y_i(t+1)$ individuals of allele type i. Timelines of individuals who have changed their identity during maturation are marked with a ×.

$$(V_{j1}, \ldots, V_{jK}) \sim \mathrm{Multinom}(W_j, (u_{j1}, \ldots, u_{jK})). \tag{6}$$

Note also that for fixed i and conditional on $\mathbf{Y}(t)$, the V_{ji} are independent.

In the following we make use of the convention that, given two random variables X_1 and X_2, $E(X_1|X_2)$ and $\mathrm{Var}\,(X_1|X_2)$ represent the random variables $g(X_2)$ and $h(X_2)$, respectively, where $g(x) = E(X_1|X_2 = x)$ and $h(x) = \mathrm{Var}\,(X_1|X_2 = x)$ (see Grimmett and Stirzaker, 2001, Def. 3.7.3). From Eqs. (2) and (4) and the independence of the $W_i|\mathbf{Y}(t)$ we have that

$$\begin{aligned} E(W_i|\mathbf{Y}(t)) &= \lambda_i Y_i(t), \\ \mathrm{Var}\,(W_i|\mathbf{Y}(t)) &= \sigma_i^2 Y_i(t), \\ \mathrm{Cov}\,(W_i, W_j|\mathbf{Y}(t)) &= 0, \qquad \text{for } i \neq j, \end{aligned} \tag{7}$$

while from Eqs. (5) and (6) we have that

$$\begin{aligned} E(Y_i(t+1)|\mathbf{W}) &= \sum_{j=1}^{K} u_{ji} W_j, \\ \mathrm{Var}\,(Y_i(t+1)|\mathbf{W}) &= \sum_{j=1}^{K} u_{ji}(1 - u_{ji}) W_j, \\ \mathrm{Cov}\,(Y_i(t+1), Y_j(t+1)|\mathbf{W}) &= -\sum_{k=1}^{K} u_{ki} u_{kj} W_k, \qquad \text{for } i \neq j. \end{aligned} \tag{8}$$

Recall the laws of total expectation, total variance and total covariance which state that for any random variables A, B and C,

$$\begin{aligned} E(A) &= E(E(A|B)), \\ \mathrm{Var}\,(A) &= E(\mathrm{Var}\,(A|B)) + \mathrm{Var}\,(E(A|B)), \\ \mathrm{Cov}\,(A|B) &= E(\mathrm{Cov}\,(A, B|C)) + \mathrm{Cov}\,(E(A|C), E(B|C)). \end{aligned} \tag{9}$$

Applying these laws to Eqs. (7) and (8) one obtains

---

[1] Throughout the paper, a vector of length K will be denoted in bold type, e.g. $\mathbf{W} = (W_1, \ldots, W_K)$.

Please cite this article in press as: Burden C.J., Wei Y., Mutation in populations governed by a Galton–Watson branching process. Theoretical Population Biology (2017), https://doi.org/10.1016/j.tpb.2017.12.001.