Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 237 (2005) 238–244
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Selection, mutation and sexual reproduction in an infinite haploid population with a genome of finite length

Anna Fukshansky*

Department of Computer Science, Royal Holloway University of London, Egham, Surrey, TW20 0EX, UK

Received 27 January 2005; received in revised form 29 March 2005; accepted 14 April 2005
Available online 22 June 2005

Abstract

We present a model which describes mutation, selection and sexual reproduction in an infinite haploid population with a finite genome. Each generation is described using an approximation which assures a certain persistent form of the distribution of the number of deleterious elements. The steady state exists and is determined. In addition, we conclude that the introduction of sexual reproduction increases the mean fitness in the equilibrium.

Keywords: Population genetics; Mathematical modelling; Fitness; Mutation–selection equilibrium; Recombination; Molecular evolution

1. Introduction

One major goal of population genetics is to describe and understand the evolution of a population over generations. The classical approach to describe the influence of mutation and selection in an infinite haploid population is the well-known model of Kimura and Maruyama (1966). There are many approaches and extensions of this model, which differ according to their specific aims (e.g. Barton, 1992; Dawson, 1999; Johnson, 1999; Shpak and Kondrashov, 1999; Barton and Shpak, 2000; Bürger, 2001; Fukshansky, 2001, 2004). In a previous model (Fukshansky, 2004) we suggested a description of a haploid infinite population where the mutations can be of different degree of danger to the organism. We described the influence of selection and mutation on such a population and introduced a genome of finite length. This approach enabled us to describe repeated and in particular positive mutations. The results of this paper are the following: Using an approximation we can show that if in some generation t the distribution of the number of mutated elements is of a certain form, namely the sum of a Poisson distribution and a small addition of a certain type, then the same form is preserved in the next generation and finally in the steady state. In the present paper we extend this model by including sexual reproduction, i.e. by allowing recombination in the population. We do this by applying the model described by Dawson (1999), where recombination was introduced in a population with an infinite genome, to our model with a finite genome. Like in Dawson's paper, we consider a population of haploid individuals which mate to produce diploid zygotes which undergo meiosis to produce haploid offspring. Our model describes no linkage between genes in the genome. Our results show that when we include recombination, the form of the distribution is still preserved in the next generation and the steady state exists and is of a similar form as without recombination (i.e. in an asexual haploid population). One main result is that according to our model the mean fitness in the steady state increases when recombination is introduced.

In the previous model without recombination as well as in the present one we describe the genome as a finite

*Tel.: +44 1784 443698; fax: +44 1784 439786.
E-mail address: a.fukshansky@cs.rhul.ac.uk.

0022-5193/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2005.04.011

set of elements. A mutation can hit any element with equal probability. An element is some piece of the genome, or even some characteristic of a genome, which implicitly is assigned to some locus or more generally some location in the genome. When we considered selection and mutation only, the position of the element in the genome did not play any role, but it does now when we introduce recombination, since for a fixed locus one of the elements of the two parents is inherited by the offspring. To describe this sexual model we use the initial definition of the hypergeometric model, which has been developed and described in Barton (1992), Shpak and Kondrashov (1999) and Barton and Shpak (2000). The necessary condition for this is the equiprobability of genotypes within each prototype. This condition is satisfied, because our elements are susceptible to mutations with equal probability, so, according to the position of deleterious elements for a given number of deleterious elements (phenotype) each genotype, i.e. the choice of this number of deleterous elements in the set of all elements, is equally probable.

2. Notation of the model

We are dealing with a population which is exposed to selection, mutation and recombination of randomly mated parent gametes. The genome of each organism is represented as a finite set of M elements. Each element can belong to one of the $N+1$ classes, which have a different local fitness s_n, where in the special class 0, $s_0 = 1$, and in the other N classes which contain the *deleterious* elements, $1 > s_1 > s_2 > \cdots > s_N > 0$. The fitness of the whole genome with y_n elements in class n is equal to the product of the local fitnesses of all the elements,

$$s_1^{y_1} s_2^{y_2} \cdots s_N^{y_N} \tag{1}$$

in particular the fitness is multiplicative. Each element *mutates* with probability u and then transfers into the class n with probability γ_n, $\sum_{n=0}^{N} \gamma_n = 1$. The mean of the number of mutations per genome per generation step is $U = uM$. The expected fraction of elements in class $n, 1 \leqslant n \leqslant N$, is given in each generation by the parameter

$$\alpha_{tn} = \frac{E[Y_n(t)]}{E[X(t)]},$$

where $E[Y_n(t)]$ is the expected value of elements in class n, and $E[X(t)]$ is the expected number of all deleterious elements. We set

$$f_t = \sum_{n=1}^{N} \alpha_{tn} s_n, \tag{2}$$

this is the mean fitness of a genome with one mutation in generation t.

Each generation is described by the distribution of the number of deleterious elements, $P_t(i), 0 \leqslant i \leqslant M$ with the generating function

$$\widetilde{P}_t(s) = \sum_{i=0}^{\infty} P(i) s^i$$

and the set of parameters α_{tn}, $1 \leqslant n \leqslant N$.

We now determine the mean fitness in some generation t like in Fukshansky (2001).

Let a genome have i deleterious elements which are distributed in the classes $1, 2, \ldots, N$ according to the vector $\vec{y} = (y_1, y_2, \ldots, y_N)$, with $\sum_{n=1}^{N} y_n = i$. The fitness of such a genome is given in Eq. (1). The probability for such a genome is

$$P_t(i, \vec{y}) = P_t(\vec{y}|i) P_t(i) = \frac{i! \alpha_{t1}^{y_1} \cdots \alpha_{tN}^{y_N}}{y_1! \cdots y_N!} P_t(i).$$

Hence, the mean fitness is derived as follows:

$$\begin{aligned} w_t &= \sum_{i=0}^{\infty} \sum_{\sum y_n = i} P_t(i, \vec{y}) s_1^{y_1} \cdots s_N^{y_N} \\ &= \sum_{i=0}^{\infty} P_t(i) \sum_{\sum y_n = i} \frac{i! (\alpha_{t1} s_1)^{y_1} \cdots (\alpha_{tN} s_n)^{y_N}}{y_1! \cdots y_N!} \\ &= \sum_{i=0}^{\infty} P_t(i) f_t^i, \end{aligned}$$

where f_t is as set in Eq. (2) above. So we have

$$w_t = \widetilde{P}_t(f_t). \tag{3}$$

After selection but before mutation the distribution is denoted by $P'_t(i), 0 \leqslant i \leqslant M$, as well as all the parameters are indicated by a prime. After mutation the distribution is $P''_t(i), 0 \leqslant i \leqslant M$, and the parameters are marked with two primes.

In the following section we will describe mutation and selection. The recombination process will be introduced in Section 4 and will lead us to the next generation with distribution P_{t+1}.

3. Selection and mutation

All formulae in this section are taken over directly or adapted from Fukshansky (2004). During selection, genomes with too many deleterious elements are eliminated according to the fitness. We find that after selection the distribution and the parameters are